Biodegradation of pentachlorophenol by *Bacillus tropicus* isolated from activated sludge of a wastewater treatment plant in Durban South Africa

BY

OLADIPUPO ABIODUN AREGBESOLA

Submitted in fulfilment of the academic requirement for the degree of Doctor of Philosophy (PhD) in the Discipline of Microbiology; School of Life Sciences; College of Agriculture, Engineering and Science at the University of KwaZulu-Natal, Westville campus, Durban, South Africa.

As the candidate's supervisor, we approve this Thesis for submission.

Supervisor: Prof Ademola O. Olaniran

Signed: [redacted]

Date: 22 April 2021

Co-Supervisor: Dr. Mduduzi P. Mokoena

Signed: [redacted]

Date: 22 April 2021

PREFACE

The experimental work described in this thesis was carried out in the Discipline of Microbiology, School of Life Sciences, University of KwaZulu-Natal (Westville Campus), Durban, South Africa from August 2016 to November 2019, under the supervisions of Professor Ademola Olaniran and Dr. Mduduzi Paul Mokoena.

These studies represent the original work by the author and have not otherwise been submitted in any form for any degree or diploma to any tertiary institution. Where use has been made of the work of others, it is duly acknowledged in the text.

DECLARATION 1- PLAGIARISM

I, Oladipupo Abiodun **Aregbesola** declare that,

(1). The research reported in this thesis except where otherwise indicated, is my original research.

(2). This thesis has not been submitted for any degree or examination at any other University.

(3). This thesis does not contain other person's data, pictures, graphs, or other information, unless specifically acknowledged as being sourced from other persons.

(4). This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

(a). their words have been re-written, but the general information attributed to them has been referenced.

(b). where their exact words have been used, then their writings have been placed in italics and inside quotation marks and referenced.

(5). This thesis does not contain texts, graphics or tables copied and pasted from the internet unless specifically acknowledged, and the source being detailed in the thesis and in the references sections.

Signed: [redacted]

Name: Oladipupo Abiodun Aregbesola

Date: 20 April 2021

DECLARATION 2- PUBLICATIONS

Details of contributions to publications that form part and/or include research presented in this thesis (include publications in preparations, submitted, in press and published and give details of the contributions of each author to the experimental work and writing of each publication).

Publication 1

Title: Biotransformation of pentachlorophenol by an indigenous *Bacillus cereus* AOA-CPS1 isolated from wastewater effluent in Durban, South Africa.

Journal/Status: ***Published in the Journal: Biodegradation, Springer Nature***

Authors Contribution:

O. A. Aregbesola – Conceptualization, methodology, laboratory work, data curation, formal data analysis, software, validation, visualization, writing (original draft), review and editing of manuscript.

P. O. Mokoena – Conceptualization, funding acquisition, project administration, resources, review and editing of manuscript.

A. O. Olaniran – Conceptualization, funding acquisition, project administration, resources, supervision, review and editing of manuscript.

Publications 2

Title: Optimization of pentachlorophenol degradation by a newly isolated *Bacillus tropicus* strain AOA-CPS1 using response surface methodology and elucidation of the biodegradation kinetic parameters

Journal/ Status: ***Bioremediation/With Editor***

Author's Contribution:

O. A. Aregbesola – Conceptualization, methodology, laboratory work, data curation, formal analysis, software, validation, visualization, writing (original draft), review and editing of manuscript.

P. O. Mokoena – Conceptualization, funding acquisition, project administration, resources, review and editing of manuscript.

A. O. Olaniran – Conceptualization, funding acquisition, project administration, resources, supervision, review and editing of manuscript.

Publications 3

Title: Classic Pentachlorophenol Hydroxylating Phenylalanine 4-Monooxygenase from Indigenous *Bacillus tropicus* strain AOA-CPS1: Cloning, Overexpression, Purification, Characterization and Structural Homology Modelling.

Journal/Status: ***Being finalized for submission***

Author's Contribution:

O. A. Aregbesola – Conceptualization, methodology, laboratory work, data curation, formal analysis, software, validation, visualization, writing (original draft), review and editing of manuscript.

A. A. Kumar – Conceptualization, methodology, formal analysis, software, supervision review and editing of manuscript.

P. O. Mokoena – Conceptualization, funding acquisition, project administration, resources, review and editing of manuscript.

A. O. Olaniran – Conceptualization, funding acquisition, project administration, resources, supervision, review and editing of manuscript.

Publications 4

Title: Biochemical and Structural Characterization of a Pentachlorophenol Hydroxylating Cytochrome P450 Monooxygenase from *Bacillus tropicus* strain AOA-CPS1.

Journal/Status: ***International Journal of Biological Micromolecules/Under Review***

Authors Contributions:

O. A. Aregbesola – Conceptualization, methodology, laboratory work, data curation, formal analysis, software, validation, visualization, writing (original draft), review and editing of manuscript.

A. A. Kumar – Conceptualization, methodology, formal analysis, software, supervision review and editing of manuscript.

M. P. Mokoena – Conceptualization, funding acquisition, project administration, resources, review and editing of manuscript.

A. O. Olaniran – Conceptualization, funding acquisition, project administration, resources, supervision, review and editing of manuscript.

Publications 5

Title: The role of benzoquinone reductase, a homolog of pterin 4a-hydroxytetrahydrobiopterin dehydratase in *Bacillus cereus* strain AOA-CPS1 phenylalanine hydroxylation system and pentachlorophenol degradation pathway: Cloning, overexpression, purification, characterization, and structural homology modelling.

Journal/Status: ***Published in International Journal of Biological Macromolecules***

Authors Contributions:

O. A. Aregbesola – Conceptualization, methodology, laboratory work, data curation, formal analysis, software, validation, visualization, writing (original draft), review and editing of manuscript.

A. A. Kumar – Conceptualization, methodology, formal analysis, software, supervision review and editing of manuscript.

M. P. Mokoena – Conceptualization, funding acquisition, project administration, resources, review and editing of manuscript.

A. O. Olaniran – Conceptualization, funding acquisition, project administration, resources, supervision, review and editing of manuscript.

Publications 6

Title: Cloning, overexpression, purification, characterization, and structural modelling of a metabolically active Fe^{2+} dependent 2,6-dichloro-*p*-hydroquinone 1,2-dioxygenase (CpsA) from *Bacillus cereus* strain AOA-CPS1

Journal/Status: ***Published in International Journal of Biological Macromolecules***

Authors Contributions:

O. A. Aregbesola – Conceptualization, methodology, laboratory work, data curation, formal analysis, software, validation, visualization, writing (original draft), review and editing of manuscript.

A. A. Kumar – Conceptualization, methodology, formal analysis, software, supervision review and editing of manuscript.

P. O. Mokoena – Conceptualization, funding acquisition, project administration, resources, review and editing of manuscript.

A. O. Olaniran – Conceptualization, funding acquisition, project administration, resources, supervision, review and editing of manuscript.

Publications 7

Title: Whole-genome sequencing, genome mining, metabolic reconstruction and evolution of pentachlorophenol and other xenobiotics degradation pathways in *Bacillus tropicus* strain AOA-CPS1

Journal/Status: ***Functional and Integrated Genomics /With Editor***

Authors Contributions:

O. A. Aregbesola – Conceptualization, methodology, laboratory work, data curation, formal analysis, software, validation, visualization, writing (original draft), review and editing of manuscript.

A. A. Kumar – Conceptualization, methodology, formal analysis, software, supervision review and editing of manuscript.

M. P. Mokoena – Conceptualization, funding acquisition, project administration, resources, review and editing of manuscript.

A. O. Olaniran – Conceptualization, funding acquisition, project administration, resources, supervision, review and editing of manuscript.

Signed: [redacted]

Name: Oladipupo Abiodun Aregbesola

22

Date: 08 October 2020

Abstract

Pentachlorophenol (PCP) is a persistent organic compound that bio-accumulates in the environment due to its recalcitrant nature and has been listed as a priority pollutant due to its toxicological properties. The recent incidences of xenobiotic at different sites and provinces in South Africa, other African Countries and Europe is worrisome and required a proactive measure. Biodegradation has been projected as one of the best ways to ameliorate recalcitrant impacted sites. This study thus aims to isolate and characterize PCP-degrading microorganism from the environment; degrade PCP and other compounds with the isolate in batch culture; optimize biotransformation processes for effective and efficient transformation; determine biodegradation kinetic parameters; profile metabolites produced; detect and amplify PCP-degrading genes from the selected isolate; map the degradation pathway; clone and overexpress the catabolic genes heterologous; purified and characterized the protein both biologically and structurally and sequence the whole-genome of the isolate with the view to determine the evolution and arrangements of PCP-catabolic genes in the genome of the isolate as well as exploring other potentials of the isolates. A PCP-degrading strain was isolated and characterized using a PCR amplification and analysis of the 16S rRNA. Biodegradation process parameters were optimized using response surface methodology. Degradation kinetics were determined via substrate inhibition models, while PCP transformation was evaluated by spectrophotometric and GC-MS analysis. Catabolic genes were detected and amplified via PCR. Genes were cloned via heat-shock technique using chemically competent cells. Proteins purification, digestion and sequencing were done using affinity chromatography, tryptic digestion, and Liquid Chromatography–Mass Spectrometry (LC-MS) techniques respectively. Pacific Biosciences RS II sequencer with the Single Molecule, Real-Time (SMRT) Link was used to sequence the whole genome of the isolate. Coting's were assembled and analysed using the HGAP4-*de-novo* assembly application. Genes were annotated on the Rapid Annotation using Subsystem Technology tool kit (RASTtk) and *ab initio* prediction (PROKKA) using the prokaryotic genome annotation pipelines. Metabolic model pathways of the bacterium was reconstructed using the RAST SEED Viewer. Primarily, the isolate was identified as *Bacillus cereus* strain AOA-CPS1 (*Bc*AOA) based on the 16S rDNA sequence analysis. However, a quality control test by NCBI for the submitted whole genome sequence of the strain, using an average nucleotide identity

(which compares the submitted genome sequence against the whole genomes of the type strains that are already in GenBank) resulted in the renaming of *Bc*AOA as *Bacillus tropicus* strain AOA-CPS1 (*Bt*AOA). *Bc*AOA was renamed as *Bt*AOA (based on the whole genome data submitted at NCBI under accession number CP049019). The bacterium degraded 74% of PCP (within 9 days at initial PCP concentration of 350 mg l^{-1} in a batch culture) and other chlorophenolic compounds in co-metabolism. The degradation followed first and zero-order kinetics at low and high PCP concentration, respectively with biokinetic constants: maximum degradation rate (0.0996 mg l^{-1} h^{-1}); substrate inhibition constant (723.75 mg l-1) and a half-saturation constant (171.198 mg l-1) and R^2 (0.98). The genes (pcp*ABCDE,* cytochrome P450) encoding the enzymes involved in the biodegradation of PCP were amplified from the genomic DNA of *Bc*AOA. Further, depending upon the genes amplified and identified metabolites using GC-MS, there are two different PCP transformation pathways were proposed in this study.

At optimized conditions, *Bt*AOA transformed 98.2% of 500 mg L^{-1} of PCP in 6 days which represent a significant 59.2% increase in PCP transformation compared to the unoptimized conditions. The kinetic parameters for PCP transformation at optimized conditions were: 1.064 ± 0.114 mg l^{-1} h^{-1} (maximum biodegradation rate); 229 ± 19.5 mg l^{-1} (half-saturation constant); 535 mg l^{-1} (inhibition constant); and R^2 = 0.96. Each of the catabolic genes shared >99% sequences homologies with the corresponding genes in the genomes of their ancestors, however, their biological functions remain putative to date**.** The optimum temperature and pH of CpsB were 30°C and 7.0. CpsB showed functional stability between pH 6.0-7.5 and temperature 25°C-30°C. CpsB activity was stimulated by Fe^{2+}, Ca^{2+}, EDTA (0.5-1.5 mM) and Dithiothreitol (0.5-1.0 mM) but inhibited by sodium azide and sodium dodecyl sulphate (>0.5 mM). CpsB enzyme substrate reaction kinetics studies showed allosteric nature of the enzyme and followed pre-steady state using NADH as a co-substrate with apparent v_{max}, K_m, k_{cat} and k_{cat}/K_m values of 0.465 µmol·s^{-1}, 140 µmol, 0.099 s^{-1} and 7.07 x 10^{-4} µmol^{-1}·s^{-1}, respectively, for the substrate PCP and 0.259 µmol·s^{-1}, 224 µmol, 0.055 s^{-1} and 2.47 x 10^{-4} µmol^{-1}·s^{-1}, respectively, for co-substrate NADH. The Hill plots and M-W-C model reveal CpsB allosteric nature and belong to K-System. CpsB shared 100% sequence homology with aromatic amino acid hydroxylase and is classified as aromatic amino acid hydroxylase superfamily with multiple putative conserved domains and metal ion binding sites confirming its allosteric nature. *Bacillus tropicus* AOA-CPS1 Cytochrome P450 monooxygenase (P450CPS1) exhibited optimum activity at 40°C and pH 7.5.

The P450CPS1 was stable between 25ºC-30ºC retaining 100% of its residual activity after 240 min of incubation. The activity of P450CPS1 was stimulated by Mn^{2+}, Fe^{2+}, and Fe^{3+} typical of an oxidoreductase but inhibited by 2.0 mM piperonyl butoxide and sodium dodecyl sulphate. The reaction kinetics studies showed allosteric nature of P450CPS1 showing apparent v_{max}, K_m, k_{cat} and k_{cat}/K_m values of 0.069 $\mu mol \cdot s^{-1}$, 200 µmol, 0.011 s^{-1} and 5.42×10^{-5} $\mu mol^{-1} \cdot s^{-1}$, respectively, for the substrate PCP and 0.385 $\mu mol \cdot s^{-1}$, 56.46 µmol, 0.06 s^{-1} and 1.77×10^{-3} $\mu mol^{-1} \cdot s^{-1}$, respectively, for co-substrate NADH. CpsD showed optimal activity at pH 7.5 and temperature between 30ºC–40ºC. The enzyme was stable between pH 7.0 – 7.5 and temperature between 30ºC and 35ºC. CpsD activity was enhanced by Fe^{2+} ion and inhibited by sodium azide and SDS. CpsD followed Michaelis-Menten kinetic exhibiting an apparent v_{max}, K_m, k_{cat} and k_{cat}/K_m values of 0.071 µmol s^{-1}, 94 µmol, 0.029 s^{-1} and $3.13 \times 10^{-4} s^{-1} \mu mol^{-1}$, respectively, for substrate tetrachloro-1,4-benzoquinone. CpsD belongs to the pterin-4α-carbinolamine dehydratase (PCD)/dimerization cofactor of HNF-1 (DCoH) superfamily, with specific conserved protein domains of pterin-4α dehydratase (PCD), validated Pterin-4α-carbinolamine dehydratase (DCoH), and coenzyme transport and metabolism proteins. CpsA showed optimum activity at 30ºC and pH 9.0. CpsA was stable between 20ºC-40ºC, and also retained about 90% of its activity at 60ºC. The enzyme retained about 90% activity between pH 9.0 and 11.5 and 60% activity at pH 13.0. CpsA was found to be Fe^{2+} dependent as about 90% increased activity was observed in the presence of FeSO4. CpsA showed apparent V_{max}, K_m, K_{cat} and K_{cat}/K_m of 27.77±0.9 µM s^{-1}, 0.990±0.03 mM, 4.20±0.04 s^{-1} and 4.24±0.03 s^{-1} mM^{-1}, respectively at pH 9.0. CpsA 3D structure revealed a conserved 2-His-1-carboxylate facial triad motif (His 9, His 244 and Thr 11), with Fe^{3+} at the centre. The whole genome of the isolate comprises one chromosome and one plasmid. The metabolic reconstruction for *Bacillus tropicus* strain AOA-CPS1 showed that the organism has been exposed to various chlorophenolic compounds including 1,1,1-trichloro-2,2-bis(4-chlorophenyl)ethane, 1,2-dichloroethane, 1,4-dichlorobenzene, 2,4-dichlorobenzene, atrazine, and other xenobiotics previously and it has recruited enzymes for their degradation. PCP degradation by the isolate was independent of substrate concentration but highly dependent on the maximum specific growth and degradation rates. The low-affinity coefficient and high inhibition constant obtained in this study showed that the bacterium has a high affinity and tolerance to PCP, which could be explored for bulk remediation of PCP. The combination of the recombinant plasmid's vector harbouring the PCP

catabolic genes can be used for direct bioremediation of PCP in a bioreactor optimized for the growth of the hosts for overexpression of the proteins. Alternatively, concoction of the enzymes can be produced and immobilised for direct enzymatic bioremediation of PCP and other related recalcitrant xenobiotics. The study proposed that CpsD catalysed the reduction of tetrachlorobenzoquinone to tetrachloro-*p*-hydroquinone and released the products found in phenylalanine-hydroxylation system (PheOHS) via a Ping-Pong or atypical ternary mechanism; and regulate expression of phenylalanine 4-monooxygenase by blocking reverse flux in *Bacillus tropicus* AOA-CPS1 PheOHS using a probable Yin-Yang mechanism. CpsD may play a catalytic and regulatory role in *Bacillus cereus* PheOHS and PCP degradation pathway. Findings from this study provide new insights into the biological role of CpsA in PCP degradation and suggest alternate possible mechanism of ring-cleavage by dioxygenases. The study also provides the first experimental evidence of the involvement of a putative cytochrome p450 from *Bacillus tropicus* group in PCP transformation. Sequence mining and comparative analysis of the metabolic reconstruction of *Bt*AOA with the closest strain and other closely related strains suggests that the operon encoding the first two enzymes in the PCP pathway were acquired from a pre-existing pterin-carbinolamine dehydratase subsystem. The next two enzymes were recruited (via horizontal gene transfer) from the pool of hypothetical proteins with no previous specific function while the last enzyme was recruited from pre-existing enzymes from the tricarboxylic acid cycle or serine-glyoxalase cycle via horizontal gene transfer events.

ACKNOWLEDGMENTS

I wish to express my gratitude to God Almighty, the corner stone, my all in all, for his Grace upon my life and giving me the wisdom, knowledge, and strength to complete this study. My sincere appreciation go to my supervisor, Prof. Ademola Olaniran and Co-supervisors Dr Paul Mokoena for their mentorship, assistance, and guidance throughout the course of this study. I appreciate my mentors Prof. Anthonia Olufunke Oluduro, Prof. Bridget Okiemute Omofuvbe and Prof. David A. Akinpelu for their advice, guidance, and prayers during the course of this study.

I am immensely grateful to my beautiful, lovely, friendly, confidant, adviser, compassionate and adorable wife Mrs. Modupe Victoria Aregbesola (Nee: Adegboyegun) and my children Emmanuel Taiwo Aregbesola and Daniel Kehinde Aregbesola for their support, understanding, patient, endurance, prayers, and love during the course of this study. I am thankful to my siblings Mrs Florence Iyabode Olugbenro, Mrs Omowumi Ojo, Mr Michael Ayodele Aregbesola, Mr Temitope Aregbesola, Mrs Bukola Odunlami (Nee: Aregbesola), Mrs Oluwakemi Akande (Nee: Aregbesola), Mrs Iwalola Olakiitan Badeji (Nee: Aregbesola), Miss Oluwaseyi Eunice Aregbesola, Mr Kolawole Rapheal Aregbesola and Akinboboye Aregbesola and nieces Oluwadamilare Stephen Olugbenro, Oluwabamise Peter Olugbenro, Taye Deborah Olugbenro, Kehinde Sarah Olugbenro, Oluwadamilola Ojo, Oluwatoyin Elizabeth Ojo, Daniel Gbenga Ojo, Temitope Kayode, Esther Akande and Inumidun Akande for their support, prayers and love during this study.

I am grateful to the National Research Foundation of South Africa and the University of KwaZulu-Natal for the research grant. I am also grateful to the Tertiary Education Fund (TETFund) Nigeria and the management of the Obafemi Awolowo University, Nigeria for giving me financial aid in the form of scholarship and the Study Leave with Pay.

Special acknowledgement to my senior colleague's Dr Ajit Kumar, Dr Sibusiso Maseko, Dr Leanne Pillay, Dr Ejovwokoghene Collins Odjadjare, Dr Tosin Olasehinde for their advice and contributions to this study. I also wish to express my gratitude to my laboratory colleagues: Olorunfemi S. Egbewale, Kerisha Ramessar, Adu Folashade, Rajnitha Magraj, Leah Rose Pillay

and the staff and postgraduate students of the Discipline of Microbiology (Westville campus) for their kind support.

My sincere appreciation goes to Dr Nneka Augustina Ukattah (Nee: Akwu) and her husband Barr. Maurice Ukattah, Dr Temitope Adu and Mrs Oluwatosin Adu, Dr Olumide Owolabi Omoboye Dr Mrs Omoboye, Dr Ayodeji Andrew Ajayi, Olorunfemi Egbewale, Abubaka Falaki, Ms Bethany Rebekah Nathaniel, Dr Ayotunde Babalola and Dr Ayomide Labulo for their moral and financial supports during the course of this study. I am particularly grateful to Dr Nneka Augustina Ukattah (Nee: Akwu) and her husband Barr. Maurice Ukattah for being a true friend in the days of nothing.

My appreciation goes to my pastor and mentor, pastor Gabriel Adejimi (Redeemed Christian Church of God, chapel of praise provincial pastor Durban, South Africa) and his family for their prayers and counsel during the course of this study. I also appreciate all members of the RCCG Chapel of Praise, Durban, South Africa for their supports and prayers during the course of this study.

Special thanks to everyone who in one way or the other contributed to the success of this study whose names were not mentioned.

DEDICATION

This Thesis is dedicated to my parents Chief David Jimoh Aregbesola and Mrs Beatrice Aina Aregbesola of blessed memories.

CONTENTS **page**

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER THREE

CHAPTER FOUR

CHAPTER FIVE

CHAPTER SIX

CHAPTER SEVEN

CHAPTER EIGHT

LIST OF SCHEMES

CHAPTER ONE

1.0 GENERAL INTRODUCTION AND SCOPE OF THE STUDY

Agriculture is the major source of food for humans and animals, as it provides different kinds of crops such as beans, corn, rice, wheat, millet, vegetables, and fruits (Saud AL-Ahmadi, 2019). As of February 2020, the world population was put at 7.8 billion (Worldometer) and it is expected to rise to about 9.7 billion come 2050 and might jumped to about 11.2 billion come 2100 (DESA, 2015). Maximizing crop production becomes a core factor affecting the availability and affordability of foods and foods products, to meet the growing population of mankind. This quest for food security has led to the advents and deployments of sophisticated technologies such as merchandized, hydroponic, and other climate-smart technologies to boost farm produce (Poore & Nemecek, 2018).

The use of smart technology to increase crop production was greeted with concomitant spoilages from traceable (microbial, animals and/or environmental) and untraceable sources (Tsôeu et al., 2016). Plant diseases (such as viral, bacterial, and fungal) are major factors mitigating crop production and infection of plants by pathogens can have benign to chronic effects on plant health, and in turn impact human health (Balique et al., 2015). However, their etiological agents do not usually cause infections in humans (Balique et al., 2015). The out turn of plant diseases is decreased in farm produce, thereby lessen the availability of food which may lead to starvation in the affected areas (Saud AL-Ahmadi, 2019).

In response to this menace, synthetic biocides such as pesticides, herbicides, fungicides, bactericides etc. were introduced to the agricultural and other systems that were affected, mostly by microbial deterioration (Luz et al., 2018). Organochlorine pesticides (OCPs) such as pentachlorophenol (PCP) and 2,4,6-trichlorophenol are some of the major pesticides that were used to overcome the detrimental effects of microbes to farm produce (Jayaraj et al., 2016; Quinn *et al*., 2011). These agro-chemicals are used in several ways; to ensure hygienic and quality of food we eat and to also control insects and pests in our homes and farms (Quinn et al., 2011).

The importance of these chemicals cannot be underestimated, but their impacts on the environment and health of humans cannot be underrated. Moreover, indiscriminate use of

chlorophenols has impacted surrounding soils and water bodies where the compounds were used (Bosso & Cristinzio, 2014; Luz et al., 2018). Due to its recalcitrant and capacity for long-distance travel, chlorophenol fragments are still present in sites where its applications has been discontinued (Hung *et al*., 2007). Most of these chlorophenol fragments are toxic to all forms of lives (Darko *et al*., 2008).

The use of PCP and most of its derivatives have been restricted to the eradication of disease vectors (ATSDR, 2017; Centers for Disease Control, 2018). The persistence and sorption of PCP to soil and its association with non-aqueous phase liquids has led to the detection of high concentrations of PCP in soils and groundwater's (Caliz et al., 2011). PCP detected at various contaminated sites varied between 100 – 500 mg kg^{-1} in the soils and between 10 – 1000 mg l^{-1} in water (Lumar and Glaser, 1994), as against the regulated maximum 0.22 mg l^{-1} of PCP in drinking water (U. S. EPA, 1992).

Pentachlorophenol is not a natural product and it is highly resistant to microbial degradation (Copley, 2010). Nevertheless, **s**ome microorganisms have evolved strategies to circumvent the toxicity of the compound by recruiting metabolic enzymes that can perform similar functions as their original function into their degradation pathways (Kaminski et al., 2018; Kumar & Gopal, 2015). Moreover, optimization of the process parameters can enhance the degradation efficacy of the organism (Patel & Kumar, 2016).

Other strategies to improve biodegradation of PCP include but not limited to bio-stimulation, bio-augmentation, surfactant-assisted biodegradation, co-metabolism, degradation with mixed microbial consortia, addition of secondary carbon sources, enzymatic degradation etc. Enzymatic degradation is more efficient than the microbial process, as the degrading enzymes are often resistant to abnormal environmental variables than the microbial cells that produced the enzyme, most especially at low substrate concentrations (Huang et al., 2018).

Technological advancement in the accuracy of mass spectrometry and data analysis coupled with the global metabolomics profiling strategies allow simultaneous and rapid screening of numerous metabolites from varieties of groups of chemicals, making them an attractive tool for the

discovery of novel enzymatic activities and metabolic pathways (Prosser et al., 2014; Ravikrishnan et al., 2018).

Cloning and overexpression of degrading enzymes in competent cells have given insights on their properties and structures, thereby giving room to maximise their functions. Pentachlorophenol biodegradation and genetic aspect of the degradation has been studied only in Sphingomonads. The first step in the biodegradation of PCP involves hydroxylation or monooxygenation of the compound. This step is a rate limiting step catalysed by either PCP-4-monooxygenase from *S. chlorophenolicum* (Hlouchova et al., 2012) or cytochrome P450 (Ning & Wang, 2012).

PCP-4-monooxygenase is an inefficient and poorly functional monooxygenase that often resulted into a futile cycle (Hlouchova et al., 2012) while PCP hydroxylation by cytochrome p450 has only been reported in fungus (Ning & Wang, 2012). Also, 2,6-dichloro-*p*-hydroquinone 1,2-dioxygenase (PcpA) from *S. chlorophenolicum* is the only reported enzyme capable of opening the aromatic ring of 2,6-dichloro-*p*-hydroquinone (Hayes et al., 2013). These has limited the success of the development of an improved technology for biodegradation of PCP.

The republic South Africa is a country with limited water resources, with its available water resources being utilised to their maximum capacity (Quinn et al., 2011). A diverse array of pesticides have been found in South African waters (Bollmohr et al., 2007; Du Preez et al., 2004; Thiere & Schultz, 2004) and foods (DOH, 2005). The pesticides were detected during run-off, and dry spell, one-off sampling, and water monitoring agencies during the planting seasons (Quinn et al., 2011). Moreover, agrochemicals have been found in birds (Bouwman et al., 2008), fishes (Barnhoorn et al., 2009), which serves as an index for pesticide contamination within various habitats.

The re-emergence and distribution of organochlorine pesticides in different environments and provinces of South Africa is alarming and worrisome. Organochlorine pesticides and their metabolites were detected in canals, rivers, and estuaries in the eThekwini area of kwaZulu-Natal, at concentrations beyond detection limit (WRC, 2015). Also, chlorophenols were recently found in uMgeni river, kwaZulu-Natal (Gakuba et al., 2018), Buffalo River of Eastern Cape, South Africa (Yahaya et al., 2019).

The incidence of polychlorinated biphenyls (PCBs) analogues and dichlorodiphenyltrichloroethane (DDT) in the leafs and roots of fresh vegetables in Durban, South Africa (Olatunji, 2019), PCP congeners and polycyclic aromatic hydrocarbons (PAHs) in Nandoni Dam in Limpopo Province of South Africa (Nthunya et al., 2019), showed a wide distribution of organochlorine pesticides contamination in South African environment. There is an urgent need for an environmentally friendly strategies such as bioremediation, to ameliorate the contaminated environment. Thus, this research aims to bio-prospect for indigenous bacterial strain(s) in activated sludge of domestic wastewater treatment plant in KwaZulu-Natal province of South Africa and to evaluate their potential for bulk remediation of PCP and its congeners.

1.1 Hypotheses

It is hypothesized that activated sludges of domestic wastewater effluents from various wastewater treatment plants in Durban, KwaZulu-Natal Province of South Africa contain indigenous microbes with potential for bulk bioremediation of PCP and its congeners. It is also hypothesized that these microbes would have evolved pathways and degrading-enzymes capable of utilizing chlorophenols as their sole carbon sources, while simultaneously degrading these persistent biocides.

1.2 Aims

The aims of this study were to:

- Isolate and identify autochthonous bacteria capable of degrading pentachlorophenol and its congeners from an activated sludge of a wastewater treatment plant in Durban, South Africa.
- Optimize the biodegradation process parameters for effective and efficient transformation of PCP and other chlorophenolic compounds by the isolates using the response surface methodology (RSM).
- Degrade PCP using RSM optimized conditions and to profile the degradation kinetic parameters of the PCP degradation by the selected isolates.

- ❖ Map the PCP degradation pathways in the selected isolate via PCR detection of the catabolic genes associated with PCP degradation as well as profiling the metabolites formed.
- ❖ Clone, express, purify and characterize the various enzymes that are involved in PCP degradation pathways by the isolate.
- ❖ Sequence, annotate and analyse the whole genome of selected PCP degrading isolate to fully understand and map PCP catabolic pathways in the isolate.

1.3 Objectives

The following objectives were pursued in order to realize the aims of the study:

- Culture enrichment, isolation, purification, and identification of indigenous bacteria capable of degrading PCP and its congeners from an activated sludge of a wastewater treatment plant in Durban, KwaZulu-Natal Province of South Africa.
- Optimization of the biodegradation process parameters (pH, substrate concentrations, incubation time, secondary carbon and nitrogen sources) using RSM.
- Degradation assay for PCP and other congeners, both individually and in co-metabolism using the pure bacterial isolate at the optimized RSM conditions.
- Evaluation of the consequence of different initial concentration of PCP on growth and degradation capability of the isolated organism.
- Computation of the PCP biodegradation kinetic parameters for the isolate.
- Determination of the substrate range of the organism.
- Detect and identify the metabolites produced during PCP biotransformation and mapping of the pathways followed by the isolate to degrade PCP.
- Detection of PCP-degrading genes from the genome of the bacterium via PCR.
- Cloning and overexpression of the catabolic genes in heterologous hosts.
- Purification of the overexpressed proteins to homogeneity and characterization.
- Sequence and annotate the Whole genome of the selected isolate and;
- Determination of the evolution and arrangement of PCP-degrading genes in the genome of the isolate as well as metabolic reconstruction of the isolate for exploration in other biotechnological processes.

1.4 Layout of thesis

The Thesis is divided into eleven chapters, chapter one described the general overview and scope of the research. This chapter also highlights the hypotheses, aims and objectives of the research as well as the layout of the thesis. Critical review of literature on the subject matter and description of the various methods that have been explored in the biodegradation of pentachlorophenol with their associated challenges and highlighted possible improvement strategies is presented in chapter two. This chapter also described the genetic aspect of PCP degradation and the metabolic pathways proposed previously.

Chapter three dealt with the culture enrichment, isolation, and profiling of PCP-degrading bacteria from an activated sludge in Durban, South Africa. This chapter also described PCP biotransformation using a monoculture and in co-metabolism. The chapter also reported on the substrate range determination, PCP degradation kinetics parameters, catabolic genes detection, PCP degradation metabolites profiling and mapping of PCP degradation pathways in the isolate.
Chapter four described the optimization of PCP biodegradation process parameter for effective and efficient degradation of PCP using response surface methodology. It also deals with PCP degradation and kinetic studies at optimized conditions and compared PCP degradation at optimized conditions to degradation at unoptimized conditions.

Chapter five reports on characterization of classic pentachlorophenol hydroxylating phenylalanine 4-monooxygenase (CpsB) from *Bacillus cereus* strain AOA-CPS1. This chapter described, cloning, overexpression, biochemical characterization, and structural analysis of CpsB as well as the protein interaction and homology modelling of the enzyme. The chapter looks into the enzyme activity, optimum pH and temperature, pH and temperature stability, kinetic parameters, effects of inhibitors and metal ions on CpsB. The chapter also dealt with the in-gel trypsin digestion and identification of the purified CpsB in ES-MS, template-based structure prediction and homology modelling as well as evolutionary relationships of CpsB with other monooxygenases. Chapter five also described a new catalytic role of bacteria phenylalanine-4-monooxygenase in xenobiotic degradation and provided the first experimental evidence on the involvements of bacteria phenylalanine-4-monooxygenase in chlorophenol enzymatic monooxygenation and described the stoichiometry of the PCP hydroxylation reaction.

Chapter six highlights the role of benzoquinone reductase (CpsD), a homolog of pterin 4a-hydroxytetrahydrobiopterin dehydratase in *Bacillus tropicus* strain AOA-CPS1 phenylalanine hydroxylation system and pentachlorophenol degradation pathway. The chapter described cloning, overexpression, biochemical characterization, and structural homology modelling of pterin 4a-hydroxytetrahydrobiopterin dehydratase. Chapter six also described a new catalytic role (benzoquinone reductase) of a bacteria pterin 4a-hydroxytetrahydrobiopterin dehydratase in xenobiotic degradation and presents the first experimental evidence on the enzymatic reduction of tetrachloro-1,4-benzoquinone to tetrachloro-1,4-hydroquinone by a pterin 4a-hydroxytetrahydrobiopterin dehydratase (CpsD) from *Bacillus tropicus* strain AOA-CPS1.

Cloning, overexpression, purification, characterization, and structural modelling of a metabolically active Fe^{2+} dependent 2,6-dichloro-*p*-hydroquinone 1,2-dioxygenase (CpsA) from *Bacillus tropicus* strain AOA-CPS1 is described in chapter seven**.** Chapter seven also described the first experimental evidence of the biological role of a hypothetical protein belonging to the Bacillus multispecies glyoxalase family protein as an aromatic ring-cleaving dioxygenase; 2,6-dichloro-*p*-hydroquinone 1,2-dioxygenase that convert 2,6-dichloro-*p*-hydroquinone to 2-chloromaleylacetate.

Chapter eight described cloning, overexpression, purification, and biochemical characterization of a putative pentachlorophenol hydroxylating cytochrome p450 monooxygenase (P450CPS1) from a newly isolated *Bacillus tropicus* strain AOA-CPS1. The chapter is devoted to the enzyme activity, kinetic parameters, effects of pH, temperature, metal ions and inhibitors on P450CPS1 hydroxylating activity, homology modelling and evolutionary relationships of P450CPS1 with other PCP-4-monooxygenases

Chapter nine highlighted the features and characteristics of the whole genome of the isolate used in this study. The chapter described the evolution and arrangements of PCP degrading genes as well as the metabolic reconstruction of the isolate for better understanding of the biotechnological applications of the isolate. Chapter ten summarized the main findings of the study in relation to the set aims and objectives, highlights the limitations to the study and provides perspectives on the future development of the research.

1.5 References

ATSDR. (2017). Agency for Toxic Substances and Disease Registry. Substance priority list (candidates for toxicological profiles)., (October), 190. Retrieved from www.atsdr.cdc.gov/SPL/resources.%0Aiii

Balique, F., Lecoq, H., Raoult, D., & Viruses, P. C.-. (2015). Can plant viruses cross the kingdom border and be pathogenic to humans? *Mdpi.Com*. Retrieved from https://www.mdpi.com/1999-4915/7/4/2074

Barnhoorn, I. E. J., Bornman, M. S.. Jansen van Rensburg, C.J. & Bouwman, H. (2009). DDT residues in water, sediment, domestic and indigenous biota from a currently DDT- sprayed area. *Chemosphere*, *77*, 1236–1241.

Bollmohr, S. Day, J.A. & Schultz, R.Bollmohr, S.; Day, J.A. & Schultz, R. (2007). Temporal variability in particle-associated pesticide exposure in a temporarily open estuary, Western Cape, South Africa. *Chemosphere*, *68*, 479–488.

Bosso, L., & Cristinzio, G. (2014). A comprehensive overview of bacteria and fungi used for pentachlorophenol biodegradation. *Reviews in Environmental Science and Biotechnology*, *13*(4), 387–427. https://doi.org/10.1007/s11157-014-9342-6

Bouwman, H.; Polder, A.; Venter, B. & Skaare, J.U. (2008). Organochlorine contaminants in cormorant, darter, egret, and ibis eggs from South Africa. *Chemosphere*, *71*, 227–241.

Caliz, J., Vila, X., Martí, E., Sierra, J., Nordgren, J., Lindgren, P.-E., … Montserrat, G. (2011). The microbiota of an unpolluted calcareous soil faces up chlorophenols: Evidences of resistant strains with potential for bioremediation. *Chemosphere*, *83*(2), 104–116. https://doi.org/10.1016/j.chemosphere.2011.01.016

Centers for Disease Control. (2018). *Fourth National Report on Human Exposure to Environmental Chemicals Updated Table* (Vol. 1). Retrieved from http://www.cdc.gov/exposurereport/pdf/FourthReport_UpdatedTables_Sep2013.pdf

Copley, S. D. (2000, June 1). Evolution of a metabolic pathway for degradation of a toxic xenobiotic: The patchwork approach. *Trends in Biochemical Sciences*. Elsevier Ltd. https://doi.org/10.1016/S0968-0004(00)01562-0

Darko, G., Akoto, O., & Oppong, C. (2008). Persistent organochlorine pesticide residues in fish, sediment and water from Lake Bosomtwi, Ghana. *Chemosphere*, *72*, 21 – 24.

DESA, U. (2015). World population projected to reach 9.7 billion by 2050. Retrieved from https://scholar.google.com/scholar?hl=en&safe=off&as_q=UN DESA. World population projected to reach 9.7 billion by 2050. United Nations. Department of Economic and Social Affairs. 2015. p.3Co2-earth. Available from: https://www.co2.earth/

DOH. (2005). MRL of pesticides in foodstuff, 20.03.2011, Available from. Retrieved from http://www.doh.gov.za/department/foodcontrol/pesticides/chemical.zip%0A

Du Preez, L. H., Jansen van Rensburg, P. J., Jooste, A. M., Carr, J. A., Giesy J. P., Gross, T. S., Kendall, R. J., Smith, E. E., Van der Kraak, G. & Solomon, K. R. (2004). Seasonal exposure to triazine and other pesticides in surface waters in the western Highveld corn-production region in South Africa. *Environmental Pollution*, *135*, 131–141.

Gakuba, E., Moodley, B., Ndungu, P., & Birungi, G. (2018). Partition distribution of selected organochlorine pesticides in water, sediment pore water and surface sediment from Umngeni river, kwaZulu-Natal, South Africa. *Water SA*, *44*(2), 232–249. https://doi.org/10.4314/wsa.v44i2.09

Gong, X., Qi, S., Wang, Y., Julia, E. B. & Lv, C. (2007). Historical contamination and sources of organochlorine pesticides in sediment cores from Quanzhou Bay, Southeast China. *Marine Pollution Bulletin*, *54*, 1434–1440.

Hayes, R. P., Green, A. R., Nissen, M. S., Lewis, K. M., Xun, L., & Kang, C. (2013). Structural characterization of 2,6-dichloro-*p*-hydroquinone 1,2-dioxygenase (PcpA) from *Sphingobium chlorophenolicum*, a new type of aromatic ring-cleavage enzyme. *Mol Microbiol*, *88*(3), 523–536. https://doi.org/10.1111/mmi.12204

Hlouchova, K., Rudolph, J., Pietari, J. M. ., Behlen, L. S., & Copley, S. D. (2012).

Pentachlorophenol hydroxylase, a poorly functioning enzyme required for degradation of pentachlorophenol by *Sphingobium chlorophenolicum*. *Biochemistry*, *51*(18), 3848–3860. https://doi.org/10.1021/bi300261p

Huang, Y., Xiao, L., Li, F., Xiao, M., Lin, D., Long, X., & Wu, Z. (2018). Microbial degradation of pesticide residues and an emphasis on the degradation of cypermethrin and 3-phenoxy benzoic acid: A review. *Molecules (Basel, Switzerland)*, *23*(9). https://doi.org/10.3390/molecules23092313

Hung, C., Gong, C., Chen, H., Hsieh, H., Santschi, P.H., Wade, T.L. & Sericano, J.L. (2007). Relationships between pesticides and organic carbon fractions in sediment of the Danshui River estuary and adjacent areas of Taiwan. *Environmental Pollution*, *148*, 546 – 554.

Jayaraj, R., Megha, P., & Sreedev, P. (2016). Organochlorine pesticides, their toxic effects on living organisms and their fate in the environment. *Interdisciplinary Toxicology*, *9*(3–4), 90–100. https://doi.org/10.1515/intox-2016-0012

Kaminski, M. A., Furmanczyk, E. M., Sobczak, A., Dziembowski, A., & Lipinski, L. (2018). *Pseudomonas silesiensis* sp. nov. strain A3T isolated from a biological pesticide sewage treatment plant and analysis of the complete genome sequence. *Systematic and Applied Microbiology*, *41*(1), 13–22. https://doi.org/10.1016/j.syapm.2017.09.002

Kumar, B. L., & Gopal, D. V. R. S. (2015). Effective role of indigenous microorganisms for sustainable environment. *3 Biotech*, *5*(6), 867–876. https://doi.org/10.1007/s13205-015-0293-6

Lumar, R. T., and Glaser, J. A. (1994). *Field evaluation of the remediation of soils contaminated with wood-preserving chemicals using lignin-degrading fungi.* (O. S. Hinchee RE, Leeson A, Semprini L, Ed.) (Bioremedia). Boca Raton: Lewis Pubisher.

Luz, G. V. S., Sousa, B. A. S. M., Guedes, A. V., Barreto, C. C., & Brasil, L. M. (2018). Biocides used as additives to biodiesels and their risks to the environment and public health: A review. *Molecules*, *23*(10). https://doi.org/10.3390/molecules23102698

Ning, D., & Wang, H. (2012). Involvement of Cytochrome P450 in Pentachlorophenol Transformation in a White Rot Fungus *Phanerochaete chrysosporium*. *PLoS ONE*, *7*(9),

e45887. https://doi.org/10.1371/journal.pone.0045887

Nthunya, L. N., Khumalo, N. P., Verliefde, A. R., Mamba, B. B., & Mhlanga, S. D. (2019). Quantitative analysis of phenols and PAHs in the Nandoni Dam in Limpopo Province, South Africa: A preliminary study for dam water quality management. *Physics and Chemistry of the Earth*, *112*, 228–236. https://doi.org/10.1016/j.pce.2019.02.003

Olatunji, O. S. (2019). Evaluation of selected polychlorinated biphenyls (PCBs) congeners and dichlorodiphenyltrichloroethane (DDT) in fresh root and leafy vegetables using GC-MS. *Scientific Reports*, *9*(538), 1–10. https://doi.org/10.1038/s41598-018-36996-8

Patel, B. P., & Kumar, A. (2016). Optimization study for maximizing 2,4-dichlorophenol degradation by *Kocuria rhizophila* strain using response surface methodology and kinetic study. *Desalination and Water Treatment*, *57*(39), 18314–18325. https://doi.org/10.1080/19443994.2015.1091988

Poore, J. & Nemecek, T. (2018). Reducing food's environmental impacts through producers and consumers. *Science*, *360*(6392), 987–992. https://doi.org/10.1126/science.aaq0216

Prosser, G. A., Larrouy-Maumus, G., & de Carvalho, L. P. S. (2014). Metabolomic strategies for the identification of new enzyme functions and metabolic pathways. *EMBO Reports*, *15*(6), 657–669. https://doi.org/10.15252/embr.201338283

Quinn, L. P., de Vos, B. J., Fernandes-Whaley, M., Roos, C., Bouwman, H. & Van den Berg, J. . (2011). *Pesticide use in South Africa: One of the largest importers of pesticides in Africa*. *Pesticides in the Modern World - Pesticides Use and Management*. https://doi.org/10.5772/16995

Saud AL-Ahmadi, M. (2019). Pesticides, anthropogenic activities, and the health of our environment safety. In *Pesticides, Anthropogenic Activities and the Health of our Environment [Working Title]*. IntechOpen. https://doi.org/10.5772/intechopen.84161

Thiere, G. & Schultz, R. (2004). Runoff-related agricultural impact in relation to macroinvertebrate communities of the Lourence River, South Africa. *Water Research*, *38*, 3092–3102.

Tsôeu, L. I., Jooste, P. J., Charimba, G., & Hugo, C. J. (2016). Spoilage potential of a novel group of bacteria isolated from dairy products. *South African Journal of Science*, *Volume 112*(Number 1/2), 01–08. https://doi.org/10.17159/sajs.2016/20150227

U. S. EPA. Contaminants and remedial options at wood preserving sites (1992). Washinton, DC: U. S. EPA.

WRC. (2015). Persistent organic pollutants Threats: Threats posed by the organic pollutants to a tourist destination. Retrieved October 19, 2019, from http://www.wrc.org.za/wp-content/uploads/mdocs/PB_1977_PoPs in DBN estuaries.pdf

Yahaya, A., Okoh, O. O., Agunbiade, F. O. & Okoh, A. I. (2019). Occurrence of phenolic derivatives in Buffalo River of Eastern Cape South Africa: Exposure risk evaluation. *Ecotoxicology and Environmental Safety*, *171*(November 2018), 887–893. https://doi.org/10.1016/j.ecoenv.2019.01.037

CHAPTER TWO

Literature Review

Pentachlorophenol biodegradation: kinetics, genetics, pathways, challenges and strategies for improvement.

Oladipupo A. Aregbesola, Mduduzi P. Mokoena, Ademola O. Olaniran[1] *

Discipline of Microbiology, School of Life Sciences, College of Agriculture, Engineering and Science, University of KwaZulu-Natal (Westville Campus), Private Bag X54001, Durban, Republic of South Africa

***Corresponding Authour:**

Phone: +27 31 260 7400

Fax: +27 31 260 7809

E-mail: olanirana@ukzn.ac.za

Abstract

Pentachlorophenol (PCP) is a refractory biocide that bio-accumulates in the environment due to its recalcitrant nature and has been listed as a priority pollutant owing to its toxicological and health effects. PCP is produced either by alkaline hydrolysis of hexachlorobenzene or chlorination of phenols iron (lll) chloride or anhydrous aluminium chloride as a catalyst. PCP had been widely used as biocide; however, its use has been restricted and listed as a priority pollutant, due to its toxicity. Due to its relatively high volatility, mobility, solubility of its ionized form in water and a long-range dissemination, PCP has been found in different kinds of environments, microorganisms, eggs, mammalians milk, urine, blood, and adipose tissue. Fate of PCP includes volatilization, photolysis, leaching, adsorption, fixing, aerobic and anaerobic bioremediation. Several conventional physicochemical strategies have been used to transform PCP. These techniques are very fast, expensive, environmentally unfriendly and require further processing for complete mineralization. Bioremediation is eco-friendly and offers complete degradation of the compound. Biotransformation of PCP occurs via oxygenolysis, hydroxylation or reductive dechlorination. Microbes have evolved pathways for complete degradation of PCP, they circumvent the toxicity of PCP by transforming it into a non-toxic metabolite(s); and used the metabolites as carbon and sources. However, these pathways are still in the early stages of development and are metabolically inefficient. Several bioprocesses such as vermi-remediation, co-metabolism, pH stabilization, electron shuttling, bio-electrochemical degradation, the use of redox mediators and growth substrates, surfactant-assisted biodegradation have been explored to enhance PCP degradation. Most of these strategies ended at pilot scales. Also, limited kinetics parameters of the various microbes' profiles for PCP degradation hamper upscale of these bioprocesses. The biokinetic parameters of the participating microbes in the presence of the substrates and adjuncts used for improved bioremediation, need to be documented for an effective bioprocess to be designed for bulk remediation. Although, biodegradation has its own challenges, it remains the best option for cleaning up PCP impacted sites.

Keywords: pentachlorodibenzofuran, hexachlorodibenzofuran, Yin-Yang regulation, biodegradation, glutathione, maleylacetate, pentachlorophenol.

Abbreviations

1,4-dihydroxybenzene	(Di-OH_3B)
2,3,5,6-Tetrachlorophenol	Tet-ClP
2,3,6-trichloro-p-hydroquinone	Tri-ClHQ
2,6-dichloro-p-hydroquinone	Di-ClHQ
2-Chloromaleylacetate	2-ClMA
3,5,6-Trichloro-1,2,4-benzenetriol	Tri-$ClBOH_3$
Benzene-1,2,4-triol	Tri-OH_3B
Chloro-1,2,4-trihydroxybenzene	$ClOH_3B$
Chloro-dihydroxy-benzene	$ClOH_2B$
Dichloro-1,2,4-trihydroxybenzene	Di-$ClOH_3B$
Dichloro-dihydroxy-benzene	Di-$ClOH_2B$
Glutathione	GSH
Glutathione transferases	GSTs
Heptachlorodibenzofuran	HpCDF
Heptachlorodibenzo-para-dioxin	HpCDD
Hexachlorodibenzofuran	HxCDF
Hexachlorodibenzo-para-dioxin	(HxCDD
Maleylacetate	MA
Octachlorodibenzofuran	OCDF
Octachlorodibenzo-para-dioxin	OCDD
Pentachloroanisole	PCA
Pentachlorodibenzofuran	PeCDF
Reactive oxygen species	ROS
S-glutathionyl -(chloro)hydroquinone reductases	GS-HQRs
S-glutathionyl -p-hydroquinone	GS-p-HQ
S-glutathionyl-(chloro)hydroquinone	GS-HQ
S-glutathionyl-2-hydroxy-p-hydroquinone	GS-OH_2HQ
S-glutathionyl-dichloro-p-hydroquinone	GS-Di-ClHQ
S-glutathionyl-trichloro-p-hydroquinone	GS-Tri-ClHQ

Tetrachloro-benzoquinone	Tet-ClBQ
Tetrachlorodibenzofuran	TCDF
Tetrachloro-*p*-hydroquinone	Tet-ClHQ
Trichloro-dihydroxy-benzene	Tri-ClOH_2B
Trihydroxy-benzene	Tri-OH_3B

2.0 Pentachlorophenol

Pentachlorophenol (PCP) is a chlorophenol with 5 chlorine ions (Cl^-) and one hydroxyl (OH^-) group, a member of pentachlorobenzenes, a pentachlorophenolate conjugate acid and an ionizable organochlorine pesticide (CHEBI, 2016). A pure PCP exist as colourless to white solid with the needle-like crystals with a molecular weight of 266.34 g mol^{-1} and a density of 1.987 g mol^{-1} at 22 ^oC, while the impure PCP is dark-grey to brown and exists as dust, beads, or flakes (Van Der Zande, 2010).

A pure anhydrous PCP has a melting point of 191^oC (EPA, 2010) while the technical grade has a melting point of between 187 – 189^oC (Van Der Zande, 2010). It has a boiling point of 310 ^oC, Vapour pressure of 1.1×10^{-4} mm Hg (0.02 Pa) at 20 °C (PubChem, 2018), 16 Pa at 100 ^oC (Van Der Zande, 2010), relative vapour density of 9.20 (PubChem, 2018); Log K_{ow} of between 5.12 – 5.18 (Van Der Zande, 2010) and a dissociation constant (pKa) of 4.7 at 25 °C (IARC, 2019; WHO, 2003).

PCP is slightly soluble in water (80 mg l^{-1} at 20 °C), due to the acidic pH generated by its dissociation (Becker et al., 2008; Van Der Zande, 2010); soluble in acetone and benzene; very soluble in diethyl ether, ethanol, and methanol respectively (EPA, 2010). However, at high pH, PCP forms highly water-soluble salts and at close to neutral pH, PCP is more than 99% ionised (Becker et al., 2008; Van Der Zande, 2010).

2.1 Production of pentachlorophenol

Pentachlorophenol is produced via alkaline hydrolysis of hexachlorobenzene or chlorination of phenols using iron (lll) chloride or anhydrous aluminium chloride as a catalyst (EPA, 2010; IARC, 2019). In the past, PCP is usually produced in an analytical grade by a series of purification processes to remove the impurities that were generated during the thermal processes (EPA, 2010). PCP has been manufactured in Poland, Germany, the Netherlands, Switzerland, the United Kingdom, Spain, (IARC, 2019).

Although, PCP production has been discontinued in many countries, however, several companies: Mexico (2), China (1), USA (10), the United Kingdom (3), India (1), Switzerland (2), Hong Kong Special Administrative Region (1); Germany (2), Canada (1), South Africa (1),

Israel (1), Japan (1) and the Netherlands (1), were registered as producers of PCP in 2016 (Chem Sources, 2016; IARC, 2019). Moreover, multi-stages chlorination processes involved in the production of PCP generates impurities such as dioxins, furans, HxCDD, HpCDD, OCDD, PeCDF, HxCDF, HpCDF, and OCDF (McLean et al., 2009).

2.2 Uses of pentachlorophenol

PCP had been used extensively as a wood preservative, herbicide, defoliant, algicide, defoliant, germicide, fungicide, and molluscicide, and could be found in ropes, textiles, paints, adhesives, canvas, leather, insulation, electric poles and brick walls (IARC, 2019; UNEP, 2013). PCP had also been used in tanneries (Mikoczy & Hagmar, 2005); molluscicide (Zheng et al., 2012); production of pentachlorophenol laurate (United Nations, 2010); as a preservative in adhesives, oil-based paint, glues, mushroom farms (UNEP, 2013). PCP has also been used as a pesticide in rice and sugar production, water treatment and pre-harvest defoliant in cotton (UNEP, 2013). The current use of PCP and its salts is limited to industrial areas in the treatment of utility poles, cross arms, railroad crossties, fibres heavy-duty textiles not intended for clothing, wooden pilings, fence posts, lumber/timbers for construction and as recipe in chemical synthesis (EPA, 2010; IARC, 2019; UNEP, 2013).

2.3 Toxicological properties of pentachlorophenol

PCP is very toxic, Short term exposure to PCP can damage the central nervous system, while chronic exposure could lead to liver and kidney damage (Fisher, 1991; Sing et al., 2014). PCP has the potential for dermal adsorption (NIOSH, 2005; OSHA, 2018), dermal exposure to PCP has been linked to non-Hodgkin's lymphoma, multiple myeloma, kidney cancer with reported cases of physical and neuropsychological health effects in former sawmill workers that persisted long after PCP exposure had ceased (Demers et al., 2006; McLean et al., 2009; NIOSH, 2019). PCP inhibit oxidative phosphorylation and make cell membranes permeable to protons, resulting in dissipation of the transmembrane pH gradients and electrical potentials (McAllister et al., 1996).

Globally harmonized system (GHS) for classification and Labelling of chemicals suspected PCP to be a cancer-causing agent (European Parliament, 2008); possibly carcinogenic to humans (IARC, 2012); likely to be carcinogenic to humans (US EPA, 2014) and confirmed animal

carcinogen (ACGIH, 2014). As a result of its toxicological profile, PCP has been listed as a priority pollutant (CDC, 2018; ATSDR, 2017). Many developed and developing countries have banned and/or severely restricted their uses, but the disposal of woods treated with PCP and PCP-contaminated soils are still a threat to aquatic, terrestrial and humans, and the environment in general, as the compound still remains a recalcitrant pesticide from a toxicological point of view (Proudfoot., 2003; Sing et al., 2014).

2.4 Incidence of pentachlorophenol in the environment

The volatility, mobility and solubility of the ionized form of PCP in water and its long-range dissemination have led to the widespread contamination of all environments (UNEP, 1996). PCP could get to the environment via production and during application in accordance to the manufacturer's specification (UNEP, 2013). PCP has been specifically detected in soils, air, lakes, rivers, basins, snow, rainwater, drinking water, plants, sediments, aquatic organisms, bacteria, fungi, eggs, in mammalians milk, urine, blood and adipose tissue (ATSDR, 2001; EPA, 2008).

The use of PCP in dip treatment of lumber that is usually done in an open-air basin has led to water and soil pollution in the affected areas (McAllister et al., 1996). Also, soils samples from abandoned sawmill were also contaminated with chlorophenols that further leached into the surrounding soil (Persson et al., 2007). The persistence and sorption of PCP to many soil types and its association with non-aqueous phase liquids has led to the detections of high concentrations of PCP in soils and groundwater’s (Lumar and Glaser, 1994).

PCP is substantially released into the environment through its manufacturing process, wood treatment processes, use, and disposal of treated wood and wood shafts (UNEP, 2013, 2017). PCP may be emitted into the environment (air, soil, and water) from wood treatment facilities; volatilisation into air by evaporation of PCP from treated wood products; runoff from wood surface to soil during transfer of treated wood from dipping tanks for drying; volatilisation into the air and leachates to soil during the drying process; runoff from the wood surfaces into soil from leachates and from outdoor storage of treated wood; as wood waste of treated wood, solid waste and sludge from the bottom of dipping/treatment tank (UNEP, 2013, 2017).

In Mexico, about 0.0029 kg of PCP were released into the soil in 2005, approximately 38.0 kg were emitted into the air between 2006 - 2009, and not less than 17 776 kg was release into the environment by incineration of unspecified objects which could either be an active ingredient, treated material, or end-use products in 2008 (OSHA, 2018; UNEP, 2013, 2017; US EPA, 2019). PCP is also a metabolite of other organochlorines such as quintozene, hexachlorobenzene and lindane, but, the extent of these potential sources cannot be quantified (UNEP, 2013).

Although it's used has been discontinued in most Countries, chlorophenols are still present in the environment due to its extensive use in the past (Ma *et al*., 2005). For instances, PCP was detected in about 84% of human urine samples in American patients (Bravo *et al.*, 2005); urine of children with parent-reported attention deficit hyperactivity disorder (Lenters et al., 2019; X. Xu et al., 2011), and soils around sawmill and waste dump sites (Kitunen *et al*., 1987). Previous and recent studies have reported the presence of PCP and its congeners in South African fresh and estuary water bodies (Belchik, 2009; Olujimi et al., 2012; Yahaya et al., 2019). Bioaccumulation of PCP in aquatic systems has also been reported (Galve *et al*., 2002; Belchik, 2009).

2.5 Fate of pentachlorophenol in the environment

Pentachlorophenol is very resistant to natural remediation, so its fate on the environment is of great concern. The fate of PCP in the soil includes volatilization, leaching, adsorption, fixing, abiotic and biotic transformation (Puglisi et al., 2009; Wong & Bidleman, 2011). Photolysis in aqueous solution is regarded as the fastest means of PCP transformation which may lead to total transformation of PCP in water within hours (UNEP, 2011b). In waters, sediments and soils, where depth and turbidity prevent the exposure of PCP to light, biodegradation is the relevant strategies of remediating such sites (UNEP, 2011b). Under certain environmental conditions, microbes could adapt to their new environment and biodegrade PCP present, with half-lives of < 4, 10 and 20 weeks in waters, soils and sediments respectively why the half-life in anaerobic sediment ranged between 13 to 144 days (UNEP, 2011b; USEPA, 2008). PCP-degrading microbes are abundant in many contaminated sites and some of them have assembled pathways to degrade chemical either aerobic or anaerobically (UNEP, 2011a).

Some microorganisms preferentially transform PCP to PCA (UNEP, 2013), the transformation of PCP to PCA appears to be a detoxification step that allows metabolism of otherwise toxic levels of PCP (Chung & Aust, 1995). Moreover, the rate of PCA formation from PCP could be high under aerobic process (Rigot & Matsumura, 2002; Walter et al., 2004). Trace amounts of PCA have also been detected from microbial degradation of PCP using mixed microbial consortia (Ford et al., 2007; Machado et al., 2005; Rubilar et al., 2011). However, demethylation of chloroanisoles back to PCP has also been reported (D'Angelo & Reddy, 2000; Rubilar et al., 2011). Biodegradation through dehalogenation is the principal pathway in anaerobic biotransformation of PCP in sludges, soils, and aquatic habitat once in the sediment (UNEP, 2013).

Pentachloroanisole (PCA) is expected to split into fragments and volatilise into the air over time in an aquatic environment, however, PCP is biomethylated to PCA and both compounds split into fragments and remained in the sediments (Pierce and Victor, 1978). Also, there was evidence that PCP transformed to lower congeners such as tetrachlorophenol, trichlorophenols) and tetrachloroanisole (UNEP, 2013; USEPA, 2008). The behaviour, fate and toxicity of PCP and its salts are quite similar (Van Der Zande, 2010).

2.6 Biological transformation of pentachlorophenol

PCP can be degraded via physicochemical or microbial processes, several conventional physicochemical strategies such as adsorption (Mathialagan & Viraraghavan, 2009), mixing coagulation, extraction, photochemical oxidation, supersonic chemistry process, hydrogenolysis, radiolysis (Lan et al., 2016; Zhang et al., 2008). Furthermore, ions exchange, liquid-liquid extraction, chemical oxidation, advanced oxidation processes (Pera-Titus et al., 2004; Ren et al., 2016) and Zero-valent metal-based degradation (Kim & Carraway, 2000) have also been used to transform chlorophenol.

Although, these techniques are very fast, they are very expensive and not environmentally friendly due to the formation of more toxic intermediates (Olaniran and Igbinosa, 2011), and require further processing for complete mineralization (Herrera et al., 2008; Patel & Kumar, 2016). On the contrary, bioremediation is an effective, non-invasive, and eco-friendly method of removing chlorophenols from the environment (Dong et al., 2009; Megharaj et al., 2011;

Olaniran and Igbinosa, 2011). Moreover, biological techniques such as biosorption (Mathialagan & Viraraghavan, 2009), bioremediation (Ammeri et al., 2017; Wang et al., 2016) and enzymatic oxidation (Li et al., 2011), is cheaper, effective, and eco-friendly for PCP degradation (Field & Sierra-Alvarez, 2008; Juwarkar et al., 2010; Rubilar et al., 2008).

Biodegradation of PCP has gained worldwide attention due to the complete degradation of PCP by autochthonous microbes in the environment (Arora & Bae, 2014). Complete mineralization of PCP occurs via oxygenolysis and hydroxylation by aerobic or reductive dechlorination via the anaerobic process (Field & Sierra-Alvarez, 2008).

Microbes have overtime evolved strategies for complete transformation of PCP in aerobic or anaerobic conditions either by circumventing the toxicity of the compound by removing PCP from their cells via efflux; transforming PCP into a non-toxic metabolite(s); and/or by using PCP as their sole carbon source (McAllister et al., 1996). The ability of microbes to degrade PCP into less toxic compound(s) depends on environmental variable which including but not limited to pH and composition of the organic matter (Cea et al., 2005), water content (Seech et al., 1991), temperature (Valo et al., 1985), humic contents (Rüttimann-Johnson & Lamar, 1997), electron acceptors and oxygen (D'Angelo & Reddy, 2000). Anaerobic transformation of PCP occurs via reductive dechlorination, where the chloride ions are sequentially replaced by hydrogen ions until it is completely transformed into either phenol, benzoate, acetate, carbon dioxide or methane (Mohn & Tiedje, 1992).

Bacteria and fungi degrade PCP by incorporating one or two atoms of oxygen into the compound through oxygenase process (Reddy & Gold, 2000) because the aromatic ring is deficient in electrons and less susceptible to electrophilic attack by molecular oxygen (Sahm et al., 1986). Aerobic biodegradation can also occur via hydroxylation of PCP into Tet-ClBQ by substituting chlorine atom with a hydroxyl (OH) group (Crawford et al., 2007; Xun et al., 2010). Microbial degradation of PCP has also occurred via methylation (McAllister et al. 1996), however, methylation takes place mainly in co-metabolism (Bosso & Cristinzio, 2014).

Many fungal species have been reported to degrade PCP through methylation using lignin-degrading enzymes (Rubilar et al., 2008; McAllister et al., 1996). Fungi have specifically used laccases and peroxidases to convert PCP to PCA (Rubilar et al. 2008; McAllister et al. 1996).

PCA has a low toxicity due to it lipophilic property, however, it can cross the cell membrane barrier and bioaccumulate in microorganisms (Bosso & Cristinzio, 2014). Furthermore, some microbes have an affinity to bind to PCP which enhances its adsorption into bacterial and fungal biomass using the charge attraction between the PCP and biomass (Rubilar et al., 2012). Bacteria and fungi that have been directly involved in the biotransformation of PCP and their degradation kinetics are highlighted in Table 1.

Table 1: Microorganisms that are involved in the biodegradation of pentachlorophenol

Microorganism	Sources	Method	Medium	PCP (mg l^{-1})	% PCP Removal	Kinetics (mg l^{-1} h^{-1})	References
Bacteria							
Bacillus cereus AOA-CPS1	Sludge	Cometabolism	MSM	500	98.22	1.064	Aregbesola et al., (2020)
Flavobacterium sp. (ATCC 39723) now *S. chlorophenolicum* L-1	Sediment	Cometabolism	MSM	200	83	0.15± 0.01	Brown et al., 1986
M. chlorophenolicum PCP1	Soil	Cometabolism	Soil	100	80	0.167	Miethling & Karlson, 1996
S. chlorophenolica RA2	Soil	Cometabolism	Soil	100	83	0.125	Miethling & Karlson, 1996
ATCC 33790	Soil	Biostimulation	Soil	361	60.6	0.0015	Pu & Cutright, 2007
ATCC 33790	Soil	Bioaugmentation	Soil	361	98.9	0.01125	Pu & Cutright, 2007
ATCC 21918	Soil	Biostimulation	Soil	263	59.6	0.00175	Pu & Cutright, 2007
ATCC 21918	Soil	Bioaugmentation	Soil	263	79.2	0.0025	Pu & Cutright, 2007
ATCC33790 + ATCC21918	Consortium	Biostimulation	Soil	262	61.6	0.0017	Pu & Cutright, 2007
ATCC33790 + ATCC 21918	Consortium	Bioaugmentation	Soil	262	98.2	0.0083	Pu & Cutright, 2007
Acinetobacter sp. ISTPCP-3	Sediment	Liquid	MSM	20 - 200	50 - 100	NI	Sharma et al., 2009
Escherichia coli PCP1	Sediment	Chemostat	MSM	100	45	NI	Sharma & Thakur, 2008
Acinetobacter sp PCP3	Sediment	Chemostat	MSM	100	> 80	NI	Sharma & Thakur, 2008
P. aeruginosa PCP2	Sediment	Chemostat	MSM	100	60	NI	Sharma & Thakur, 2008
Bacillus megaterium CL3	Sludge	Batch	MSM	600	> 90	NI	Karn et al., 2010
Bacillus pumilus CL5	Sludge	Batch	MSM	600	> 90	NI	Karn et al., 2010
Bacillus thuringensis CL11	Sludge	Batch	MSM	600	> 90	NI	Karn et al., 2010
P. mendocina NSYSU	Sediment	Batch	Msm	140	100	NI	Kao et al., 2004

Table 1: Microorganisms that are involved in the biodegradation of pentachlorophenol (cont'd)

Bacteria	Sources	Method	Medium	PCP (mg l^{-1})	% PCP Removal	Kinetics (mg l^{-1} h^{-1})	References
Pseudomonas sp. strain SR3	Soil	Batch	Msm	100	> 70	NI	D'Angelo & Reddy, 2000
Brevibacterium casei TVS-3	Effluent	Batch	Msm	1000	72	NI	Verma & Singh, 2013
C. testosteroni CCM 7530	Sludge	Batch	Soil	100	57	NI	Vítková et al., 2011
Bacillus sp	Effluent	Batch	Msm	500	56.5	NI	Tripathi et al., 2011
Enterobacter sp.	Sludge	Batch	Msm	532	85	NI	Karn & Geetanjali, 2014
Janibacter sp FAS23	Sediment	Cometabolism	Msm	300	40	NI	Khessairi et al., 2014
S. chlorophenolica DSM7098^{T}	Soil	Cometabolism	Msm	400	89	NI	Yang et al., 2006
Kocuria sp. CL2	Sludge	Batch	Sludge	100.029	58.64	NI	Karn et al., 2011
Kocuria sp. CL2	Sludge	Batch	Msm	600	90.00	NI	Karn et al., 2011
Pseudomonas stutzeri CL7	Sludge	Batch	Msm	600	>90.00	NI	Karn et al., 2010a
Pseudomonas stutzeri CL7	Sludge	Batch	Sludge	100.029	66.80	NI	Karn et al., 2010a
Bacillus cereus (DQ002384)	Sludge	Batch	Msm	300	62.75	NI	Singh et al., 2009
S. marcescens (AY927692)	Sludge	Batch	Msm	300	90.33	NI	Singh et al., 2009
S. marcescens (DQ002385)	Sludge	Batch	Msm	300	86.60	NI	Singh et al., 2009
Bacillus cereus ITRC-S6	Effluent	Cometabolism	Effluent	300	90	NI	Chandra et al., 2009
Serratia marcescens ITRC-S7	Effluent	Cometabolism	Effluent	300	85	NI	Chandra et al., 2009
ITRC-S6 & ITRC-S7	Consortium	Cometabolism	Effluent	300	100	NI	Chandra et al., 2009
Bacillus cereus ITRC-S6	Effluent	Co-Substrates	Msm	300	65	NI	Chandra et al., 2006
S. marcescens ITRC-S7	Effluent	Co-Substrates	Msm	300	43	NI	Chandra et al., 2006

Table 1: Microorganisms that are involved in the biodegradation of pentachlorophenol

Bacteria	Sources	Method	Medium	PCP (mg l^{-1})	% PCP Removal	Kinetics (mg l^{-1} h^{-1})	References
ITRC-S6 + ITRC-S7	Effluent	Co-Substrates	Msm	300	100	NI	Chandra et al., 2006
S. chlorophenolica PCP-1	Soil	Bioreactor	Water	120 - 160	92	NI	Yang & Lee, 2008
Serratia marcescens ITRC S_7	Effluent	Cometabolism	Msm	300	90.33	NI	Singh et al., 2007
Pseudomonas sp. Bu34	Soil	Batch	MSM	4000	73	NI	Lee et al., 1998
Fungi							
Anthracophyllum discolor	Ni	Batch	Soil	250-350	79 – 93	0.091	Rubilar et al., 2011
P. chrysosporium CECT-2798	Ni	Batch	Soil	250-350	64 –80.4	0.048	Rubilar et al., 2011
P. chrysosporium CECT-2798	Ni	Reactor	Aqueous	100	> 90	NI	Jiang et al., 2006
Trametes villosa CCB213	Soil	Cometabolism	Soil	1278-4600	58	NI	Machado et al., 2005
Agrocybe perfecta CCB161	Soil	Cometabolism	Soil	1278-4600	78	NI	Machado et al., 2005
Psilocybe castanella CCB444	Soil	Cometabolism	Soil	1278-4600	64.46	NI	Machado et al., 2005
Trametes versicolor HR131	Ni	Batch	Soil	800–1000	99 - 100	NI	Walter et al., 2005
B. adusta ATTC 90940	NI	batch	soil	300	81	0.0241	Rubilar et al., 2007
Penicillium camemberti	effluents	Reactor	MSM	266.34	86	NI	Taseli & Gokcay, 2005
Byssochlamys fulva	soil	batch	soil	25	20	NI	Scelza et al., 2008
B. adusta ATTC 90940	NI	batch	soil	300	81	0.0241	Rubilar et al., 2007

Key: NI: not indicated; *Phanerochaete chrysosporium*; *Mycobacterium chlorophenolicum*; *Sphingomonas chlorophenolica*; *Serratia marcescens*; *Sphingomonas chlorophenolica; Comamonas testosterone; Pseudomonas mendocina*; *Pseudomonas aeruginosa*; *Bjerkandera adusta*; *Sphingobium chlorophenolicum*.

2.6.1 Enhanced microbial processes for PCP biodegradation

Various strategies have been explored in a bid to enhance microbial PCP degradation, some of which include biostimulation, augmentation, cometabolism, degradation in the presence of earthworm, biodegradation of PCP via electron shuttling, bioelectrochemical degradation, biodegradation of PCP in the presence of redox mediator.

2.6.1.1 Vermi-remediation of pentachlorophenol

Earthworms are a group of dominant fractions of the biomass of many terrestrial ecosystems and they have a substantial influence on the fate of many refractory pollutants in the soil (Hickman et al., 2008; Rodriguez-Campos et al., 2014). The migration and burrowing of earthworms into the soil increase transport and distribution of microbes via bioturbation; enhance contacts between microbes and pollutants also improve soil aeration (Lin et al., 2012; Shan et al., 2011). Earthworms can increase the bioavailability of organic matter for microbes; increase microbial proliferation and distribution, which aids biodegradation of recalcitrant compounds by secretion of mucilaginous compounds and digestion of soil organic nutrients (Hickman et al., 2008; Rodriguez-Campos et al., 2014).

Bioturbation and movement of earthworms can increase contacts between microbes and PCP, thereby increasing the soil sub-surface oxygenation and porosity (Schaefer et al., 2005). Earthworms can also ingest soil with organic matter, accelerate its decomposition and secrete metabolites that enhance proliferation of the microbes in the soil (Zhang et al., 2000). Moreover, PCP adsorbed to soil particles could be transformed by the digestive enzymes and intestinal flora in the gut of the earthworms (Hickman et al., 2008). Earthworms have been found to enhance polychlorinated phenol transformation by increasing the distribution of polychlorophenols-degrading microbes and the proliferation of degrading organisms (Luepromchai et al., 2002). For instance, *Amynthas robustus Perrier* and *Eisenia fetida Savigny* accelerated soil PCP biodegradation by transferring viable bacteria from the guts of the earthworms to the soil and increase proliferation and the numbers of PCP-degrading microbes in the soil (Li et al., 2015).

2.6.1.2 Biochar assisted pentachlorophenol biodegradation

Biochar is charcoal-like substance produced via pyrolysis of organic materials from agricultural and forestry wastes biomass (Ding et al., 2016). Biochar contains aromatic and quinone structures (Joseph et al., 2010), redox-active (Klüpfel et al., 2014), and could

influence abiotic redox reactions (Saquing et al., 2016; Wang et al., 2017). Biochar enhances microbial Fe^{3+} reduction by electron transfer between a wide range of microorganisms and Fe^{3+} minerals (Kappler et al., 2014; Zhou et al., 2016). Moreover, reductive dechlorination is the primary pathway for environmental removal of PCP in soil under anaerobic condition (Zhu et al., 2018).

The reductive dehalogenation process has been found to be associated with other soil redox processes of typical biogenic elements such as carbon, iron and sulfur (Zhu et al., 2018). For instance, biochar mediates activation of aged nano zerovalent iron (nZVI) by *Shewanella putrefaciens* strain CN32 to enhance the degradation of PCP (Li et al., 2019). Biochar also enhanced the regeneration of aged nZVI by *S. putrefaciens* CN32, and prolong the reduction of PCP (Pereira et al., 2016; Xie et al., 2017). Moreover, biochar significantly accelerates electron transfer from *Geobacter sulfurreducens* to PCP, thereby facilitating the reductive dechlorination of PCP (Yu et al., 2015).

The effects of biochar on soil redox reactions and reductive dechlorination of PCP in a polluted soil under anaerobic condition was explored, biochar enhanced dissimilatory sulphate and iron reduction while simultaneously decreased PCP concentration significantly (Zhu et al., 2018). However, maximum removal of PCP was facilitated by biochar when microbial sulphate reduction was suppressed by the addition of molybdate as an inhibitor (Zhu et al., 2018).

The effects of biochars derived from different raw materials on the microbial reductive dehalogenation of PCP by microbial consortia were investigated, only caragana-derived biochar showed stable electron transfer activity for PCP dechlorination (Zhang et al., 2019). The beneficial effects of biochar's on PCP dechlorination depended on the electrical conductivity (EC) and not redox reactions nor any redox functional groups, possibly because high EC enabled the highest electron transfer, and thus influence reductive dechlorination of PCP (Zhang et al., 2019).

Biochar stimulates the growth of dehaloginating microbes and significantly enhance PCP bioremediation in paddy soil under anoxic conditions (Tong et al., 2014). Transferred of extracellular electrons in biochar amended soils enhance PCP transformation, by promoting microbial proliferation and metabolism in the soil (Tong et al., 2014). However, the relative abundance of the microbes in the soil dependent biochar concentration (Tong et al., 2014).

2.6.1.3 Biodegradation of PCP in a Microbial fuel cell

A microbial fuel cell (MFC) is a bio-electrochemical system that utilizes the natural metabolisms of microorganisms to produce electrical energy (Li, 2013). In a typical MFC system, microbes metabolize the nutrients in their environment and release a portion of the energy contained in the food in the form of electrical energy (Khan et al., 2017; Li, 2013). MFC is one of the smart and efficient technologies recently developed for the treatments of a wide range of environmental pollutants with a very little expenditure of external energy (Li et al., 2014). The energy recovered from MFC as a by-product can offset the treatment cost if the technology is properly harnessed (Khan et al., 2018).

MFC function is based on the simultaneous reduction and oxidation where substrates in the anodic chamber are oxidized to release electrons, while electron acceptor consumes the electron released in the cathodic chamber which is separated from the anodic chamber by a membrane (Ghasemi et al., 2013). The movement of electrons from the anode to the cathode generates electric current (Logan et al., 2006) and the reaction is wholly catalysed by microbes that use part of the substrate for their growth while converting the others into energy in form of electric current (Khan et al., 2015).

The combined effects of enrichment culture and non-fermentable or fermentable co-substrate on the performance and bacterial community for PCP degradation in MFCs were evaluated with simultaneous addition of glucose as growth substrate yielded the shortest acclimation period and the most endurance to heavy PCP shock loads (Wang et al., 2012). Also, combined anaerobic-aerobic conditions in air-cathode single-chamber MFCs were used to completely mineralize PCP in the presence of acetate or glucose (Huang et al., 2011). Moreover, bio-electrochemical degradation of PCP in a single and dual chambered MFC with the concurrent generation of electric current has also been evaluated, PCP was more effectively degraded under aerobic conditions in dual-chambered MFC (Khan et al., 2018).

Furthermore, bio-electrochemical stimulation of microbial reductive dechlorination of PCP using solid-state redox mediator (humin) immobilization has also been investigated. It was observed that immobilized solid-phase humin on a graphite electrode set at -500 mV significantly enhanced microbial reductive dehalogenation of PCP, as a stable solid-phase redox mediator in bio-electrochemical systems (Zhang et al., 2014). The immobilized system transformed PCP at a higher efficiency and the electrons required for the bacteria

dechlorination of PCP were primarily from the humin-immobilized electrode (Zhang et al., 2014).

2.6.1.4 Biodegradation of PCP in cometabolism

Microbial cometabolic bioremediation can be described as the biodegradation of organic compounds in the presence of primary carbon or energy sources (Luo et al., 2014). Although, the carbon and energy sources that enhance bacteria growth may not necessarily be from the conversion of organic compounds from the metabolism of the primary substrates (Atashgahi et al., 2018; Kracke et al., 2015). The universal microbial cometabolism offers effective strategies to clean-up persistent xenobiotics in our environment, but cometabolic biodegradation of recalcitrant compounds has not been widely utilized due to the concealed factors underlying the fundamentals of the process and/or applications of microbial cometabolism (Luo et al., 2014).

Microbial cometabolism is divided into three types: microbial interspecific synergistic degradation of organic pollutants; conversion of the non-growth substrate by the aid of growth substrate metabolism; and transformation of the non-growth substrate by the resting cells without acquirement of extracellular energy sources (Frascari et al., 2015; Luo et al., 2014). A typical cometabolic bioremediation system consists of microbe, growth substrate and non-growth substrate. Transformation yield and transformation capacity are used to evaluate the efficiency of cometabolic processes (Elango et al., 2011).

The major factors influencing microbial co-metabolism includes but not limited to the degrading enzymes, the toxicity of the compounds, competitive inhibition, and ATP/energy generation. The non-specific enzyme produced by microbes during consumption of growth substrate could also initiate the metabolism of the refractory pollutants and restrict the reaction rate of cometabolic degradation (Elango, 2010). The expression of the key degrading enzymes is induced by growth substrates, and their catalytic activity levels are synonymous to the concentrations of growth substrate (Li et al., 2014).

The enzymes can catalyse the biotransformation of both growth and non-growth substrates, however, transformation of growth substrate provides the essential carbon and energy sources for growth of the prospective microbes, while the non-growth substrate requires the utilization of additional energy source(s) and usually generates toxic metabolites (Li et al., 2014; Nzila, 2013; Shukla et al., 2010).

Enzyme active site is the region of an enzyme where the substrate binds and undergo a chemical reaction, and, in most cases, it's very limited (Kuang et al., 2018). In some cases, the enzyme active site could be occupied by both growth and non-growth substrates which could lead to cometabolic degradation of pollutants (Kuang et al., 2018). The competition for the available but limited enzymes active sites have been reported to cause competitive inhibition of substrate degradation when the concentration of the substrates is not at equilibrium (Li et al., 2014). Therefore, growth and non-growth substrates concentration ratio should be maintained in a range that can enhance the maximum degradation of the recalcitrant compounds. Moreover, non-competitive inhibition of substrates degradation in cometabolism processes have been reported (Keenan et al., 1994).

2.6.1.5 Enhance biodegradation via pH stabilization

Biodegradation of PCP in an aerobic pathway leads to the formation of HCL as shown in equation 1 (Crawford & Crawford, 1996; Yang et al., 2006). The formation and release of HCL into the medium decrease the pH of the medium, thereby slow down the rate of growth and degradation of the compound and may partially or completely inhibit the catalytic power of the degrading enzymes. For instance, incomplete PCP degradation by *S. chlorophenolica* DSM 7098^{T} in batch culture was attributed to the decrease in pH of the medium which in turn inhibited the activity of the isolate (Yang et al., 2006). Biodegradation of PCP in cometabolism with NaCl has been found to stabilize the pH of the medium and enhance degradation of PCP in a concentration dependent manner. In a typical batch degradation model, NaCl was found to increase PCP degradation at an optimum concentration of 1% but significantly hindered dechlorination at > 1% NaCl (Khessairi et al., 2014).

$$C_6Cl_5OH + 4.5O_2 + 2H_2O \rightarrow 6O_2 + 5HCl \qquad \text{Eq.1}$$

2.6.1.6 Surfactants-assisted pentachlorophenol biodegradation

Surfactant-assisted biotransformation of recalcitrant environmental pollutants is a growing technology for the clean-up of sites impacted with chlorophenols and other recalcitrant compounds (Arca-Ramos et al., 2018). Surface-active agents that can significantly increase the bulk aqueous concentrations of organic contaminants by partitioning the components of the soil slurry via interaction with amphipathic surfactant molecules have been reported to either inhibit (Cort and Bielefeldt, 2002; Van Hoof and Rogers, 1992) or promote substrates biotransformation rates (Aronstein, and Alexander, 1992; Cort and Bielefeldt, 2002).

Surfactants have the potential to increase the biodegradation rate of hydrophobic organic compounds in contaminated environments (Khessairi et al., 2014). Non-ionic surfactants are usually used in the bioavailability studies due to their relatively low toxicity compared to ionic surfactants (Yeh et al., 1998). Surfactant enhances substrate transport via microbial cell wall (Noordman et al., 2002). Biodegradation of PCP in the presence of non-ionic surfactants have been greeted with divergent reports. For instance, *S. chlorophenolicum* sp. strain RA2 has been used to study the degradation of PCP in simplified aqueous systems in the presence of surfactants such as an ether type polyglycol and a non-ionic surfactant Tergitol NP-10 (Cort & Bielefeldt, 2002).

It was observed that a high concentration of Tergitol NP-10 (2000 mg l^{-1}), impeded biodegradation of PCP by *S. chlorophenolicum* strain RA2 (Cort & Bielefeldt, 2002). However, the isolate can degrade PCP in the presence of 1500 mg l^{-1} Tergitol NP-10 (Cort & Bielefeldt, 2002). The incubation time required by the organism to consume 90% of the substrate rose from between 50 to 86 h to between 140 to 350 h in the presence of ≥ 1000 mg l^{-1} of TNP10 and all TNP10 concentrations tested slowed down the rates of biotransformation of PCP by the test isolates (Cort & Bielefeldt, 2002). Furthermore, Tween 80, enhanced the growth of *Janibacter* sp. and biodegradation of PCP by 30% after 72 h of incubation (Khessairi et al., 2014).

2.6.1.7 Biodegradation of PCP in the presence of redox mediators

Microbial degradation of PCP faces stiff limitation because of the persistence of the compound in the environment (El-Bialy et al., 2018), which can be attributed to the need for a suitable environmental condition, presence of electron donors or acceptors, specific growth promoting nutrients, toxicity of the compound (D'Angelo & Reddy, 2000), or a low number of autochthonous microorganisms harbouring PCP catabolic genes (Thakur et al., 2002). One of the strategies that have been exploring in the quest for effective and efficient bioremediation of PCP is via electron shuttling 'ES' (El-Bialy et al., 2018). Electron shuttling plays prominent roles in subverting the persistent nature of some organic compounds to biodegradation (Watanabe et al., 2009).

Electron transfer in bacterial takes place via direct and mediated methods, the former depends on the direct contact between the cell surface electron transfer proteins of the bacterial cells

and the electron acceptor while the latter relies on the addition of exogenous or endogenous redox mediators (Karns et al., 1983), and in both cases, redox mediators bridge the gap between the electrons and the substrates (Watanabe et al., 2009). Electron shuttling relies on the presence of redox mediators which is an integral part of electron transfer processes in microbes (El-Bialy et al., 2018). Redox mediators act as an electron's carrier between electrons donors, bacterial outer membrane proteins and electrons acceptors (Watanabe et al., 2009).

Electron shuttling has been reported to be involved in the bioremediation of halogenated organic compounds (Watanabe et al., 2009), azo dyes (Hong et al., 2007), domestic wastewater effluents treatments and electricity generation in microbial fuel cells (Fathy et al., 2016; Kiely et al., 2012). These redox mediators can be produced either naturally (Xu et al., 2014), and can be added as supplements (Tong et al., 2015). Stimulating biodegradation media with redox mediators is considered an efficient and cost-effective strategy for improving biotransformation processes (El-Bialy et al., 2018). Moreover, biochar (Tong et al., 2014) and acetate (Kao et al., 2004) has been used independently to improve overall PCP biodegradation. They have also been used in combination with Fe^{3+} (Xu et al., 2014), and a mixture of both nitrate and Fe^{3+} (Yu et al., 2014).

Acetate is a common electron donor that can drive microbial reductive processes in anaerobic environments (Xu et al., 2019). Apart from acting as a terminal electron acceptor, PCP has antimicrobial properties, but little is known about its effect on the anaerobic microbial community during biodegradation (Xu et al., 2019). Anaerobic processes are the dominant bioprocesses in flooded organic-enriched soils (such as peatlands, mangroves, wetlands, sediments and rice paddies) and their impacts on global geochemical dynamics and nutrient cycling cannot be underestimated (Dassonville et al., 2004; Preston et al., 2012).

Redox reactions occur in sequence along a redox gradient and it is based on the redox potential of the contributing redox mediators (Borch et al., 2010). Redox processes can be extremely abrogated by introducing exogenous biologically active substrates, such as organochlorines, particularly when introduced at levels that exceed the thresholds of microbial physiological tolerance (Xu et al., 2019). This could have both positive and negative effects on the autochthonous community structure and function, leading to a modification of soil redox actions (Borch et al., 2010; Nemir et al., 2010).

PCP can act as an electron acceptor under anaerobic conditions and has the capacity to interfere with electron transfer processes of indigenous microbial redox processes (Mun et al., 2008). The responses of microbes to PCP are complex and little is known about the effects of PCP degradation on soil redox reactions under PCP-stressed conditions (Bosso & Cristinzio, 2014; Chen et., 2016).

The roles of *Pseudomonas chengduensis* and *P. plecoglossicida* associated with paddy soil and ES compounds (acetate and pyruvate) in PCP biodegradation processes have been evaluated (El-Bialy et al., 2018). Both strains were found to degrade 93.5% and 94.88% of PCP starting concentrations respectively in the presence of citrate (El-Bialy et al., 2018). Furthermore, Addition of pyruvate yielded about 80% PCP degradation by *P. chengduensis* under aerobic condition while *P. plecoglossicida* also degraded about the same amount with or without pyruvate under anaerobic condition (El-Bialy et al., 2018).

Biogeochemical transformation and associated microorganisms involved in PCP biodegradation in mangrove sediments was evaluated (Xu et al., 2019), using a Stable Isotope Probing (^{13}C-labeled acetate) which is often used for bacterial anaerobic respiration, methanogenesis and biogeochemical redox reactions in soils (Ding et al., 2015; Vandieken & Thamdrup, 2013). PCP alters acetate-assimilating microbial community and redox shuttling in anaerobic soils and greatly inhibited sulphate reduction via suppression on sulphate-reducing organisms such as Desulfarculaceae, Desulfobulbaceae and Desulfobacteraceae (Xu et al., 2019).

Microbial community structures and characteristics of PCP dehalogenation in a flooded mangrove soil was evaluated in the presence of nitrate and molybdate, nitrate supply and sulphate-reducing suppression facilitated the removal of PCP in a flooded mangrove soil (Cheng et al., 2019). Biostimulation via nutrients addition was considered an alternative strategy for improved bioremediation in nitrate-limited and sulphur-accumulated soils contaminated by PCP. This can be achieved by regulating the growth of the core functional groups by coordinating the interaction between dechlorination and its coupled soil redox processes due to shifts of more available electrons to PCP dechlorination (Cheng et al., 2019).

2.7 Genetic aspects of bacteria degradation of pentachlorophenol

The genes associated with PCP degradation are borne on either chromosomal DNA or plasmids (Arora & Bae, 2014). PCP degrading genes (*pcpA,B,C,D,E*), and regulatory

(*pcpFMR*) genes that are involved in PCP-biodegradation in *S. chlorophenolicum* ATCC 39723 have been detected, identify, and characterized (Cai & Xun, 2002). PCP-degrading genes encode Di-ClHQ-1,2-dioxygenase (PcpA); PCP-4-monooxygenase (PcpB); Tet-ClHQ-reductive dehalogenase or glutathione-s-transferase (PcpC); Tet-ClBQ-reductive dehalogenase (pcpD) and maleylacetate reductase (pcpE) respectively (Anandarajah et al., 2000; Copley, 2010; Copley et al., 2012; Dai et al., 2003; Hlouchova et al., 2012; Kiefer et al., 2002; Rokicki, 2015; Rudolph et al., 2014; Su et al., 2008; Xu et al., 1999; Xun et al., 2010; Yadid et al., 2013).

These genes have been cloned and the proteins expressed in competent cells, purified to near homogeneity, and characterized. The PCP catabolic genes are clustered on two different fragments, *pcpB, pcpD*, and pcpR assembled as a cluster with the same orientation (Cai & Xun, 2002). While *pcpB* and *pcpD* are possibly co-transcribed, pcpR is not likely to be co-transcribed with *pcpBD* due to a rho-independent terminator sequence located downstream of *pcpD* (Cai & Xun, 2002).

The *pcpA*, pcpC, pcpE and pcpM genes are clustered on another fragment, but they are at discrete locations and their transcription is initiated from different promoters, this discrete orientation is not common for genes of a single degradation pathway in bacteria (Cai & Xun, 2002). As a result of the discrete orientation of these genes, it has been speculated that the discrete arrangement of the catabolic and the regulatory genes may reflect the recent acquisition of the PCP catabolic genes in *S. chlorophenolicum* ATCC 39723 for the degradation of PCP (Cai & Xun, 2002).

2.7.1 Pentachlorophenol-4-monooxygenase (PcpB)

The rate-limiting step of PCP biodegradation in *S. chlorophenolicum* L-1 is the hydroxylation of PCP to Tet-ClBQ, a reaction catalyzed by pcpB (Orser et al., 1993). PcpB is a flavin monooxygenase that catalyses reactions of various substituted aromatic compounds and has diverse substrates specificity (Xun et al., 1992b). In addition to PCP, 2,4,6-TCP, Tet-ClP, pcpB catalyzed a primary attack on different substituted phenols by hydroxylating the substrate at the *para* position with removal of halogen, amino, nitro and cyano groups to liberate halides, hydroxylamine, nitrite, and cyanide respectively (Xun et al., 1992a). However, 2 mol of NADPH or NADH is needed to liberate 1 mol of a halide, nitro or cyano groups while those of hydrogen and amino groups require only 1 mol of NADPH or NADH (Xun et al., 1992a).

Sequence analysis of homologous genes from sphingomonads: Sphingomonas, Sphingobium, Novosphingobium and Sphingopyxis (Takeuchi et al., 2001) and other microorganisms were used to evaluate the evolution of the *pcp*B gene (Tiirola et al., 2002). Evidence of natural horizontal transfer of *pcp*B gene in the evolution of polychlorophenols-degrading Sphingomonads was established in a phylogenetically diverse group of poly-chlorophenol degrading bacteria isolated from contaminated groundwater (Tiirola et al., 2002). A comparative analysis of the 16S rDNA and *pcp*B gene trees suggested that horizontal transfer of the *pcp*B gene was involved in the evolution of the catabolic pathway in the sphingomonads (Tiirola et al., 2002).

Gene similar to *S. chlorophenolicum* L-1 *pcp*B gene has also been found in *S. chlorophenolicum* strains RP-2, SR-3 and ATCC 39723 (Ederer et al., 1997; Karlson et al., 1996); *Novosphingobium* sp. strain MT1, a poly-chlorophenol degrading strain in a groundwater remediation system (Tiirola et al., 2002) and *Sphinogomonads* strain UG-30 which involve in p-nitrophenol and other chlorophenols degradation (Cassidy et al., 1999; Leung et al., 1999). Recently, the *pcp*B gene has also been detected in *Bacillus cereus* strain AOA-CPS1.

2.7.2 Tetrachloro-1,4-benzoquinone reductase (PcpD)

The primary step in the PCP biodegradation by *S. chlorophenolicum* resulted in a formation of a very toxic and highly reactive intermediate (Tet-ClBQ) which happens to be a potent alkylating agent that can react with cellular thiols at a diffusion-controlled rate (Yadid et al., 2013). The catalytic role of PcpD in PCP degradation by *S. chlorophenolicum* ensures that Tet-ClBQ formed by PcpB is sequestered until it is transformed to a less toxic intermediate (Tet-ClHQ), thereby protecting the bacterium from the cytotoxic effects of Tet-ClBQ and maintaining flux through the degradation pathway (Yadid et al., 2013). It has also been observed that the toxicity of Tet-ClBQ might have necessitated a selective pressure to maintain a slow turnover of PcpB (0.02 s^{-1}), so that a transient, regulatory and coordinated interactions between PcpB and PcpD can occur before Tet-ClBQ is released from the active site of PcpB (Yadid et al., 2013).

PcpD is an FMN- and NADH-dependent reductase that catalyzes the transformation of Tet-ClBQ to Tet-ClHQ, a previously unrecognized step in PCP biodegradation in *S. chlorophenolicum* (Dai et al., 2003). pcpD occur mainly as trimers and aggregates of trimers

but homo-trimer might be its physiological form *in vivo* (Chen & Yang, 2008). pcpD is active at a neutral pH, its activity is usually very low but crucial to the transformation of Tet-ClBQ to Tet-ClHQ (Chen & Yang, 2008). Moreover, PcpD is constitutively expressed, it is also Tet-ClHQ and PCP inducible at concentrations dependent manner, as high concentrations of these substrates is inhibitory to pcpD (Chen & Yang, 2008).

Due to the relatively high catalytic activity of the downstream enzyme (pcpC), Tet-ClHQ is not likely to accumulate in high concentrations, which means that PcpD would possibly be stimulated by Tet-ClHQ under *in vivo* conditions (Chen & Yang, 2008). The catalytic activity of pcpD is regulated by PCP and Tet-ClHQ using '*Yin-Yang'* mechanism which maintained Tet-ClBQ at a level that would neither affect biodegradation of PCP nor elicits cytotoxicity to the bacterial cells (Chen & Yang, 2008).

Structural analysis of pcpD has shown that its Tet-ClBQ binding site is slightly positively charged and it is located in a deep pit on the surface adjacent to the co-factor flavin mononucleotide and the 2Fe2S cluster (Chen & Yang, 2008). Moreover, It has also been noted that 1 mole of PcpD contained 1.7 moles of iron (Chen & Yang, 2008).

2.7.3 Tetrachloro-1,4-hydroquinone reductive dehalogenase (PcpC)

PcpC is a member of the GST, it uses GSH to dehalogenates Tet-ClHQ to Tri-ClHQ and Di-ClHQ (Orser et al., 1993; Xun et al., 1992). GSTs usually transfer GSH to the electrophilic centres of toxic and hydrophobic compounds to form a complex, with the complex commix being more soluble and less toxic than the parent compound (Sheehan et al., 2001). PcpC belongs to the zeta class of GST based on the sequence similarity of amino-acid residues at the active site and shared enzymatic activities which include dichloroacetate dehalogenase and maleylacetone isomerase activities (Anandarajah et al., 2000). The native pcpC exists as a monomer and a dimer, it utilises glutathione as its reducing agent and not NADPH, NADH, dithiothreitol, or ascorbic acid (Xun et al., 1992b).

The cytosolic GSTs has a thioredoxin-like N-terminal domain of four β-sheets and three α-helixes arranged as $\beta_1\alpha_1\beta_2\alpha_2\beta_3\beta_4\alpha_3$ and all-helical C-terminal domain (Sheehan et al., 2001). The N-terminal domain is mainly responsible for GSH binding 'G-site' and the C-terminal domain is involved in the binding of the second (hydrophobic) substrate 'H-site' (Sheehan et al., 2001). The Cys14 residue at the N-terminal region of pcpC is involved in the

transformation of Tet-ClHQ and Tri-ClHQ respectively (Belchik & Xun, 2012; Huang et al., 2008). When the Cysteine residue of pcpC is oxidatively damaged, the damaged PcpC produces GS-Tri-ClHQ and GS-Di-ClHQ allies respectively (McCarthy et al., 1996). These GS-Tri-ClHQ and GS-Di-ClHQ commix are considered a dead-end intermediate that cannot be transformed by PcpC until the discovery of the maintenance role of PcpF (Belchik & Xun, 2012; Huang et al., 2008).

2.7.4 2,6-Dichloro-*p*-hydroquinone 1,2-dioxygenase (PcpA)

PcpA is a member of the Fe^{2+}-dependent ring-cleaving hydroquinone dioxygenases (HQDOs) of the vicinal oxygen chelate (VOC) superfamily proteins, alongside 2-chlorohy-*p*-droquinone 1,2-dioxygenase (Lin E) from *Sphingomonas paucimobilis* (UT26) (Miyauchi *et al.*, 1999) and hydroquinone 1,2-dioxygenase (MnpC) from *Cupriavidus necator* JMP 134 (Yin and Zhou, 2010). However, unlike the extradiol, intradiol, and gentisate dioxygenases that are inactivated by chlorinated substrates (Dai et al., 2002). PcpA catalyses the cleavage of aromatic rings between a hydroxyl and chlorine group (Ohtsubo *et al.*, 1999; Xu et al., 1999) and the chlorinated substrate is the native substrates for the enzyme (Machonkin et al., 2010).

Di-ClHQ is a common intermediate of several chloro-aromatic compounds, such as PCP, Tet-ClP, and 2,4,6-TCP, by various organisms (Xun et al., 1992). Ring cleavage is an integral part to complete mineralization of aromatic compounds. For microorganism to completely degrade and/or obtain the needed energy from consumption of polychlorophenols, such as PCP, the benzene ring of the compound has to be open before its cleavage product(s) can be a channel to a TCA cycle for complete mineralization (Hayes et al., 2014). The ring cleavage and dehalogenation reactions catalysed by PcpA are crucial steps in the biotransformation of chlorophenols (Machonkin et al., 2010; Sun et al., 2011).

In *S. chlorophenolicum*, PcpA dehalogenates and oxidizes 2,6-Di-ClHQ and other 2,6-substituted-*p*-hydroquinones (Machonkin and Doerner, 2011), and its dioxygenase activity is affirmed by the inactivation of the enzyme in the absence of oxygen (Chanama and Crawford, 1997; Sun et al., 2011). Two pathways are known for Di-ClHQ transformation: direct cleavage to 2-chloromaleylacetate and hydroxyquinol pathway in which Di-ClHQ is converted to chloro-hydroxyquinol (Lee and Xee, 1997; Ohtsubo et al., 1999). Both

pathways are found in *S. chlorophenolicum*. However, the pathway that converts Di-ClHQ to 2-ClMA is the primary pathway for PCP degradation (Ohtsubo et al., 1999).

PcpA requires Fe^{2+} for its activity (Cai and Xun, 2002), it is PCP-inducible and its mRNA expression level increases in the presence of PCP as an inducer (Cai and Xun, 2002; Sun et al., 2011). Moreover, PcpA is highly substrate specific, it only uses 2,6-Di-ClHQ as its preferred substrate (Machonkin and Doerner, 2011). Since Di-ClHQ is a common metabolic intermediate of several haloaromatic compounds, ring cleavage and dehalogenation reactions catalysed by PcpA are crucial in the biotransformation of chlorophenols (Machonkin et al., 2010; Sun et al., 2011).

2.7.5 Maleylacetate Reductase (PcpE)

PcpE is the last enzyme in the PCP biodegradation pathway in *S. chlorophenolicum* L-1, it catalyzes reductive dehalogenation of 2-ClMA to MA and subsequent reduction of MA to 3-oxoadipate, using one molecule of either NADH or NADPH as the co-substrate (Chen et al., 2017). However, PcpE activity is also inhibited by PCP in a concentration-dependent manner like that of PcpD (Chen et al., 2017). PcpE belongs to the iron-containing alcohol dehydrogenase superfamily (Chen et al., 2009). PcpE contains about 352 amino acid residues and is encoded by the gene pcpE which is located in the *pcp*EMAC operon in *S. chlorophenolicum* L-1 genome (Cai & Xun, 2002).

PcpE consists of an N-terminal α/β and a C-terminal α–helical domains with the catalytic site located at the interface of the two domains (Chen et al., 2017). A structural model-based site-directed mutagenesis and steady-state kinetics study of PcpE showed that the putative catalytic site of pcpE is positively charged and it is in a solvent channel at the interface of the two domains and the binding of 2-CIMA/MA involves seven amino acids which are: His172, His236, His237, His241, His251, Lys140 and Lys238 respectively (Chen et al., 2017). Moreover, PcpE is at an early stage of molecular evolution, the mutagenetic analysis showed that His172 and Lys238 are germane to the catalytic activity of PcpE but mutation of His236 to an alanine can increase the catalytic activity of PcpE by more than 2-fold (Chen et al., 2017). Studies have also shown that PcpE has an extremely low alcohol dehalogenase activity toward ethanol which coincides with an earlier belief that pcpE evolved from an iron-containing alcohol dehydrogenase (Chen et al., 2017). Maleylacetate Reductase (MacA) of *Rhodococcus opacus* 1CP has also been cloned, expressed, and characterized (Seibert et al., 1998).

2.8 Genetic regulation of pentachlorophenol biodegradation

At least two potential *LysR*-type transcriptional regulators (PcpR and PcpM) and pcpF, a S-glutathionyl-(chloro)hydroquinone reductase (Belchik & Xun, 2011), were identified and reported to regulate the expression of the PCP catalytic enzymes in *S. chlorophenolicum*.

2.8.1 The regulatory role of PcpR and PcpM

PcpR regulates the expressions of *pcp*A, *pcp*B and *pcp*E while the role of PcpM in PCP degradation is not crucial for PCP degradation (Cai & Xun, 2002). The complete loss of the capacity to degrade PCP and the lack of transcription of pcpB and pcpE in the pcpR mutant suggested that PcpR regulates the expressions of *pcp*A, *pcp*B and *pcp*E respectively (Cai & Xun, 2002). The induction of *pcp*B by several polychlorinated phenols provides evidence that polychlorophenols are co-inducers of *pcp*R (Cai & Xun, 2002).

The conserved binding sites of PcpR in the regulatory regions of *pcp*A, *pcp*B and *pcp*E have been identified and characterized, a LysR-type regulator generally binds to a 15 bp region of disrupted dyadic sequence centred near position -65 from the transcriptional start codon (Cai & Xun, 2002). There is a conserved 5'-T-N_{11}-A-3' motif called the LysR motif within the dyadic sequence (Cai & Xun, 2002; Goethals et al., 1992; Schell, 1993).

With these strategies, a conserved region (5'-ATTC-N_7-GAAT-3') has been detected in the promoter sequences of *pcp*B, *pcp*A, and *pcp*E (Cai & Xun, 2002). The presence of conserved binding sites in the promoter regions of *pcp*A, *pcp*B and *pcp*E also implies that these genes are regulated as a regulon with PcpR as the activator and polychlorophenols as co-inducers (Cai & Xun, 2002).

2.8.2 The dual roles of S-glutathionyl-(chloro)hydroquinone reductases (pcpF)

GS-HQRs are a new class of GSTs that catalyze oxidoreductions (Belchik & Xun, 2011). GS-HQRs reduce GS-HQs (GS-Tri-ClHQ, GS-Di-ClHQ, GS-OH_2HQ, and GS-p-HQ), which are phylogenetically more related to each other than to other GSTs, and they share a Cys-Pro motif at the GSH-binding site (Belchik & Xun, 2011). Hydroquinones can be auto-oxidized to benzoquinones by molecular oxygen (O_2), which spontaneously react with GSH to form GS-HQs by Michael's addition (Belchik & Xun, 2011).

Moreover, the GS-HQRs are expected to re-direct GS-HQs formed either enzymatically (by oxidatively damaged pcpC) or spontaneously back to hydroquinones (Belchik & Xun, 2011). However, if the hydroquinones are intermediates of metabolic pathways, it means that the GS-HQRs play a maintenance role for the pathways by channelling the GS-HQs formed back into the pathway for them to be reconverted back to hydroquinones (Belchik & Xun, 2011).

PcpF is the first GS-HQR identified and reported to plays a maintenance role in the metabolic pathway of PCP in *S. chlorophenolicum* ATCC 39723 (Huang et al., 2008). The physiological role of PcpF in PCP degradation has been evaluated; it plays a maintenance role in PCP degradation (Belchik & Xun, 2011). The glutathionyl-Tri-ClHQ and glutathionyl-Di-ClHQ conjugates produced by oxidatively damaged PcpC are directed back to the metabolic pathway by PcpF during PCP degradation (Belchik & Xun, 2011; Huang et al., 2008). PcpF was discovered due to its proximity to a pcpC gene on the chromosome, the pcpF gene is physically linked to pcpC and when the pcpF gene is disrupted in *S. chlorophenolicum* ATCC 39723, the mutant degrades PCP more slowly and becomes more sensitive to PCP (Belchik & Xun, 2011).

The recombinant PcpF over-produced in competent cells showed an ability to convert GS-Tri-ClHQ and GS-Di-ClHQ to Tri-ClHQ and Di-ClHQ conjugates at the expense of GSH and PcpF re-direct the conjugates back to the metabolic pathway (Belchik & Xun, 2011). However, when pcpF activity was inhibited, Tri-ClHQ and Di-ClHQ commix accumulate (Belchik & Xun, 2011). The quinones in the Tri-ClHQ and Di-ClHQ conjoins are expected to undergo conditional redox cycling (Buffinton et al., 1989), and generate ROS that slows down cell growth and PCP degradation. Thus, PcpF plays two roles of re-directing Tri-ClHQ and Di-ClHQ associates back to the metabolic pathway and prevent the accumulation of these harmful complexes (Belchik & Xun, 2011).

2.9 Pathways for microbial degradation of chlorophenol

Biodegradation of PCP occurs via hydroxylation, oxygenolysis and reductive dehalogenation (McAllister et al., 1996). Two different pathways for bacterial degradation of PCP have been described (Lopez-Echartea et al., 2016). Of all microorganisms capable of degrading PCP, *Sphingomonas chlorophenolicum* strain L-1 is the most widely studied strain (Lopez-Echartea et al., 2016). In *S. chlorophenolicum* L-1, pcpB converts PCP to a more toxic intermediate (Tet-ClBQ) via hydroxylation at the para-position (Chanama & Chanama, 2011; Orser et al., 1993; Xun et al., 1992), and then reduced to Tet-ClHQ by removal of one

chlorine ion from the same carbon 4 of the phenolic ring. Reductive dehalogenation of Tet-ClBQ to Tet-ClHQ is catalysed by pcpD (Cai and Xun, 2002; Chanama & Chanana, 2011; Chen & Yang, 2008; Crawford et al., 2007; Dai et al., 2003).

Tetrachloro-p-hydroquinone (Tet-ClHQ) is sequentially degraded to Di-ClHQ by pcpC with the liberation of Tri-ClHQ as an intermediate (Chanama & Crawford, 1997; Chanama & Chanana, 2011; Joshi *et al*., 2015; Orser et al., *et al*., 1993; Xun et al., 1992). In the next stage, Di-ClHQ is cleaved into 2-ClMA by pcpA (Chanama & Chanana, 2011; Chanama & Crawford, 1997; Ohtsubo et al., 1999; Xu et al., 1999; Xun et al., 1999). In the final stage, pcpE transformed 2-ClMA into succinyl Co-enzyme A and acetyl-CoA respectively before entering into the Kreb's cycle (Cai and Xun, 2002; Chanama & Chanama, 2011; Chen et al., et al., 2009; Joshi et al., 2015), where they are finally used as carbon and energy sources by the isolate (Scheme 1).

In *Mycobacterium chlorophenolicum* strain PCP-1 and *M. fortuitism* CG-2, PCP is hydroxylated to Tet-ClHQ by cytochrome P450 (Uotila et al., 1991; Uotila et al., 1992). Tet-ClHQ is subject to hydrolytic dehalogenation followed by successive reductive dehalogenation (Tri-ClBOH$_3$, Di-ClOH$_3$B and ClOH$_3$B), yielding TriOH$_3$B (Apajalahti & Salkinoja-Salonen, 1987). TriOH$_3$B is subject to ring cleavage by 2,6-Dichloro-*p*-hydroquinone 1,2-dioxygenase (TcpC) to form maleylacetate (Rieble et al., 1994). The maleylacetate is further converted to 3-oxoadipic acid and/or other isoforms by maleylacetate reductase (TcpD), before entering into the TCA cycle where it will be used as sources of carbon and energy (Scheme 2).

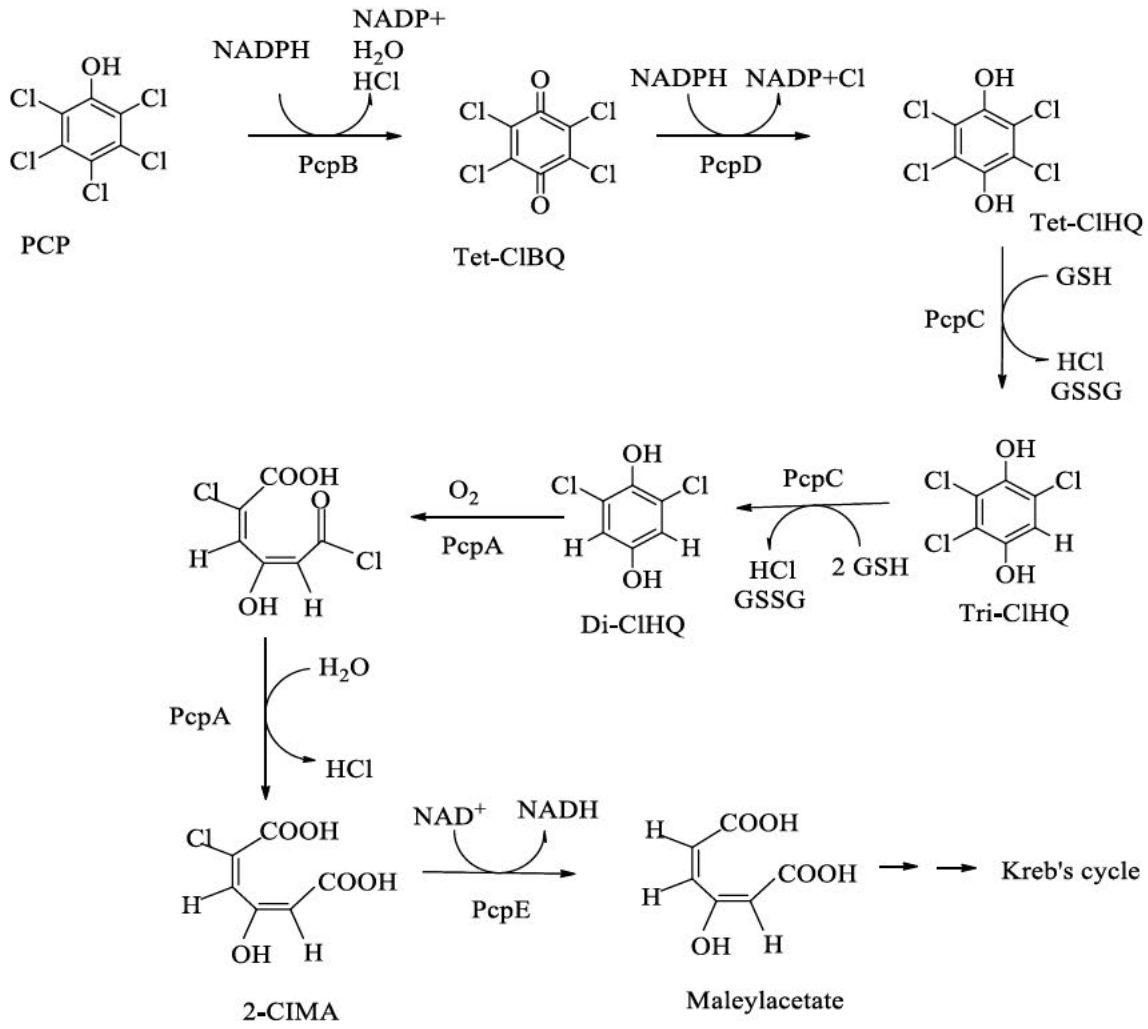
Pathway for PCP biodegradation has also been mapped in *Phanerochaete chrysosporium* (Scheme 3) via sequential identification of primary metabolites produced during PCP degradation and subsequent identification of the secondary metabolites following the addition of primary metabolites to fungal cultures (Reddy & Gold, 2000). The first step in PCP biotransformation by *P. chrysosporium* is the oxidation of PCP to Tet-ClBQ either by lignin peroxidase (LiP) or manganese peroxidase (MnP), which is further transformed into Tet-ClHQ via two parallel pathways with cross-linking steps (Reddy & Gold, 2000). In the first pathway, TeClBQ is reduced to Tet-ClHQ either by enzymic or non-enzymic process (Reddy & Gold, 2000). Tet-ClHQ further undergoes sequential reductive de-chlorination to form Tri-ClOH$_2$B, Di-ClOH$_2$B and ClOH$_2$B and Di-OH$_3$B respectively (Reddy & Gold,

2000, 1999). The 1,4-dihydroxybenzene is o-hydroxylated to Tri-OH_3B (Reddy et al., 1998; 2000).

Alternatively, Tet-ClBQ is transformed into Tri-$ClOH_2B$, by a 1,4- addition reaction (Joshi & Gold, 1994; Reddy and Gold, 2000). Tri-$ClOH_2B$ can undergo three successive reductive de-halogenations to form Tri-OH_3B (Reddy & Gold, 2000). However, at the trichloro-, dichloro- and chloro-dihydroxy-benzene stages, hydroxylation reaction can cross-link the metabolic flow from one pathway to the other (Reddy & Gold, 2000). In the final state, Tri-OH_3B ring is cleaved to maleylacetate by a Tri-OH_3B dioxygenase (Rieble et al., 1994). It should be noted however that, this pathway is restricted to fungus rather than by bacterial contaminants (Reddy et al., 1998).

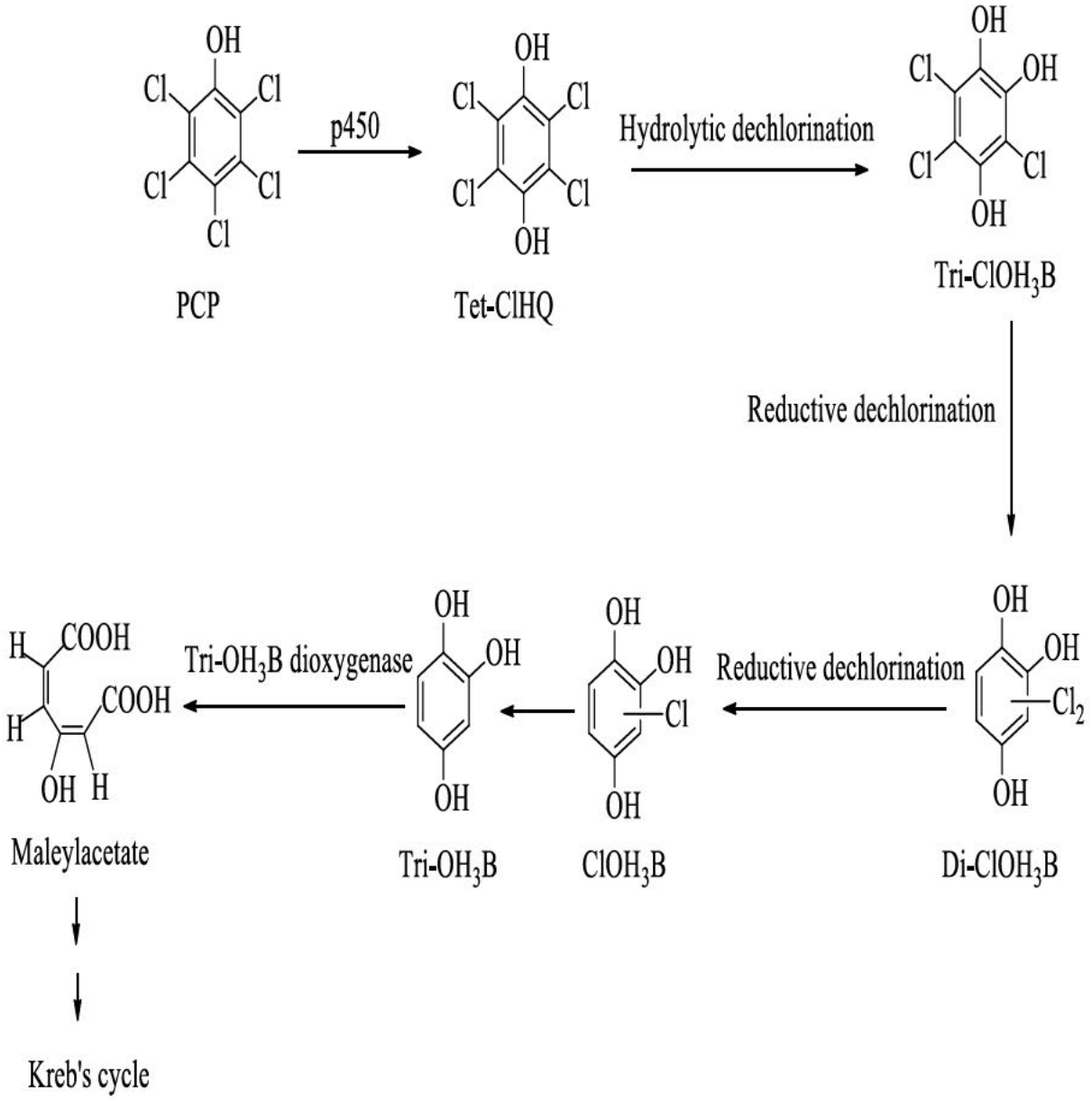
Two pathways were observed in PCP dehalogenation in an anaerobic sludge digestion process (Scheme 4). In the first pathway, PCP dehalogenation was initiated by the removal of the chloride ion at the *ortho* position followed by sequential dechlorination to liberate 2,3,4,5-TetCP, 3,4,5-TriClP and 3,5-DiCP respectively (Chen, Hsu, & Berthouex, 2006). The other pathway follows sequential dechlorination to generate 2,3,5,6-TetCP, 2,3,5-TriCP, 3,5-DiCP and 3-MClP respectively (Chen et al., 2006).

Scheme 1



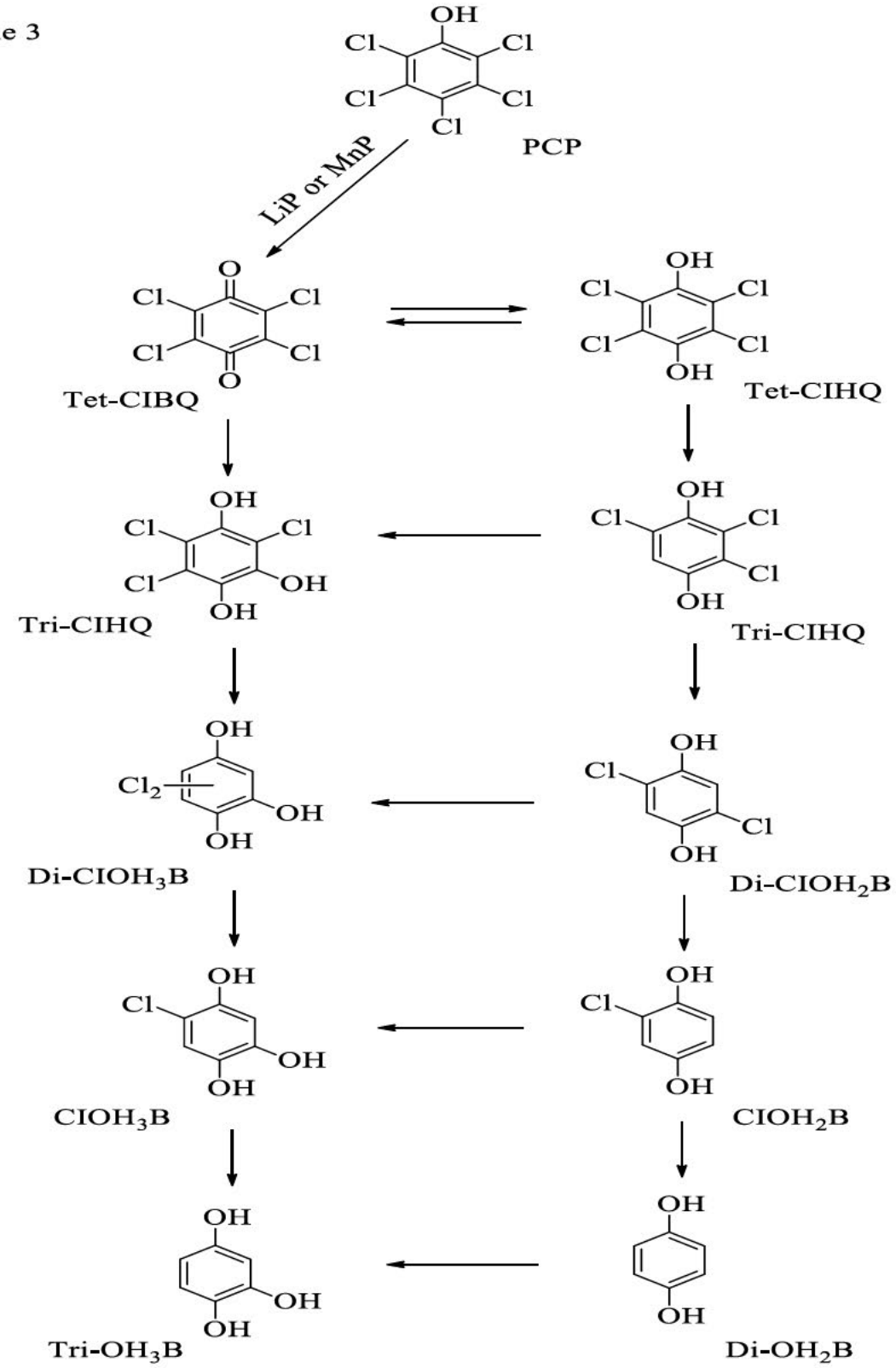
Scheme 1: Pentachlorophenol degradation pathway in *Sphingobium chlorophenolicum* L-1.

Scheme 2



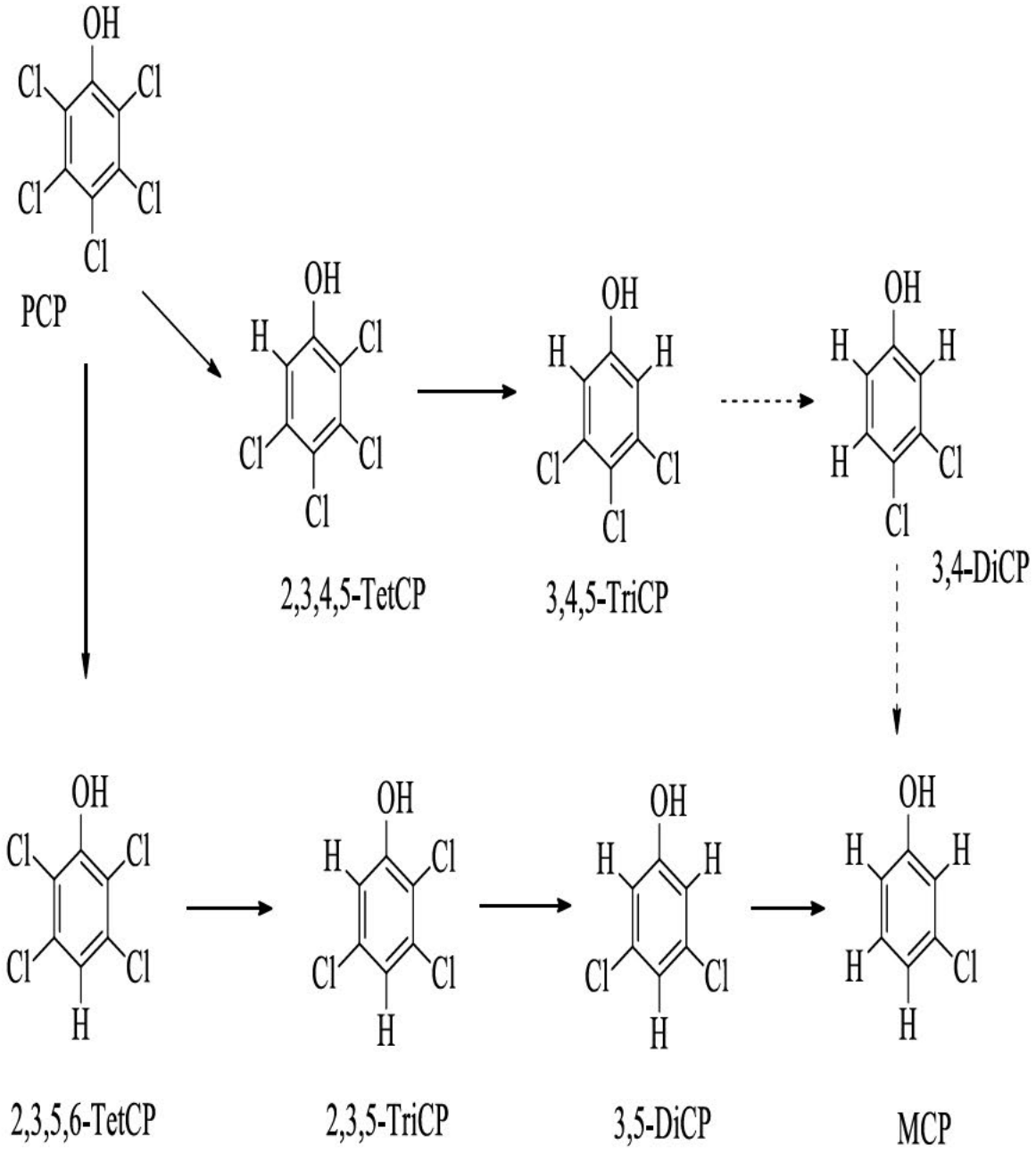
Scheme 2: Pentachlorophenol degradation in *Mycobacterium chlorophenolicum* strain PCP-1 and *Mycobacterium fortuitism* CG-2.

Scheme 3



Scheme 3: Pentachlorophenol degradation in *Phanerochaete chrysosporium*

Scheme 4



Scheme 4: Pentachlorophenol degradation in an anaerobic sludge digestion process

2.10 Challenges associated with microbial PCP degradation strategies

The problems associated with the use of nZVI in bioremediation is scarcely reported, studies on the effects of nZVI on microbes show that nZVI can penetrate or adsorb on cell membranes of bacteria (Auffan et al., 2008). This often blocks cellular ducts, alter structural changes to the membranes, or inhibit mobility and nutrient intake which can eventually lead to the death of the organism (Xiu et al., 2010).

Studies have also shown that nanoscale zero-valent iron (nZVI) have effects on aquatic and soil organisms such as *Heterocypris incongreuens* (El-Temsah and Joner, 2013), *Oryzias latipes* (Chen et al., 2011), earthworms (El-Temsah et al., 2013; El-Temsah and Joner, 2012). Nanoscale zerovalent iron also has some effects on mammalian cells, for instance, human bronchial epithelium cells died in the presence of nZVI dissolved in physiological saline (Keenan et al., 2009; Phenrat et al., 2009), nZVI also cause neurotoxic effects and neuron microglia in rodent (Phenrat et al., 2009).

Microbial degradation of PCP is masked with pockets of challenges such as the limited number of profiled PCP-degrading microbes (Li et al., 2015). PCP-degrading strains are usually isolated via culture enrichment (Szewczyk & Dlugonski, 2009); however, significant numbers of the degrading microbes could not be enriched nor isolated in nature as pure culture (Gans et al., 2005; Li et al., 2015).

Also, complete biodegradation of PCP by autochthonous soil microbes within a shortest time under natural growth conditions is sometimes difficult (Li et al., 2015; Puglisi et al., 2009; Wong & Bidleman, 2011), due to low bioavailability of soil pollutants, sub-optimum nutrient levels, oxygen depletion, and competition between the autochthonous and allochthonous microbes (Tyagi et al., 2011).

Appreciable number of microorganisms have now been profiled as been able to degrade PCP as a sore carbon and energy source or in the presence of a growth substrate(s). However, only a few studies reported the kinetics of these organisms towards PCP, this makes upscale of the bio-processes for bulk bioremediation difficult.

The sorption of PCP has a deficiency of transferring only PCP from waste materials to adsorbent without its actual treatment (Khan et al., 2018), and the conventional

electrochemical treatments of wastewater are laborious and energy intensive (Khan et al., 2017). Biodegradation on the other hand has the challenges of generating excessive sludge (Wang et al., 2012). The use of redox mediators such as citrate is concentration dependent, and high concentration of redox mediators impede cell growth (Lee et al., 2013). Some of the redox mediators also turned out to be competitive inhibitors if not used them in the right proportion (Milton & Minteer, 2017).

The efficacy of MFC for PCP degradation is embroiled with numerous challenges which are not limited to: electron transfer mechanism in an anodic chamber; electron acceptor in a cathodic chamber; electrodes material and the shape of electrodes; geometric design of the MFC; types of membrane used to partition the electrodes; environmental variables such as conductivity temperature, pH and salinity; PCP concentration, the microbes and their metabolisms (Aghababaie et al., 2015; Sultana et al., 2015).

Also, scaling up the MFC system is often problematic, and required technological know-how to achieve maximum results (Aghababaie et al., 2015). Vermi-remediation seems promising, however, most of the reported vermi-remediation experiments were limited to pilot's scale and were evaluated either in the laboratory or mesocosms, with little or no knowledge of the kinetics, feasibility, and efficacy of the strategy on the field (Rodriguez-Campos et al., 2014).

2.11 Strategies for improvement

Due to its toxicological properties, for effective and efficient bioreactors for bulk biodegradation of PCP to be designed, the biokinetic parameters such as the inhibition and PCP utilization constants of the participating microorganisms need to be known. The half-saturation or substrate utilization constant (K_s) is an affinity coefficient of an organism to the substrate, a low K_S is an indication that a particular organism has a high affinity for a compound and vice versa (Patel and Kumar, 2016).

On the other hand, the PCP inhibition constant shows the inhibitory effects of PCP on the organism, a higher K_{si} indicate the tolerance level of the organism to PCP and its high potential for bulk PCP bioremediation *in-situ* or *ex-situ* (Patel and Kumar, 2016). Also, the kinetic parameters of the organisms to the biostimulants, redox mediators and other adjuncts used as growth substrates and co-metabolites to improve the bio-process should also be determined. Since some of these supplements are concentrations dependent, they also act as inducers and inhibitors of the degrading enzymes.

The addition of glucose as a growth substrate seems to be more promising, however, cheap glucose or other simple carbon and energy sources as growth substrates should be sourced to make the process more efficient and cost-effective. Glucose should be added in the right proportion since it can sometimes act as a metabolite repressor and its metabolism is known to alter the pH of the medium.

The use of exo-electrogenic microorganisms that can readily oxidize organic matter and transfer electrons to the anode without using redox mediators is highly essential for an efficient PCP dechlorination via MFC (Aghababaie et al., 2015). At the same time, temperature, and pH optimal enhance proliferation of microorganisms and in turn can improve MFC performance (Aghababaie et al., 2015). High salinity and ionic strength can increase the conductivity of substrate and enhance MFC efficiency but should be optimised due to the inhibitory effects of high salt concentration on the growth and physiology of microbes. Also, enzymatic degradation of PCP is an aspect that is poorly explored.

2.12 Conclusion

For improved bioreactors for bulk remediation to be designed, the biokinetic parameters of the participating microbes in the presence of the substrates and adjuncts used for improved bioremediation, need to be documented. Since some of the supplements needed in concentrations dependent manners, some act as inducers and competitive inhibitors of the degrading enzyme functions. The use of exo-electrogenic microorganisms that can readily oxidize organic matter and transfer electrons to the anode without using redox mediators is highly essential for an efficient PCP dechlorination via MFC, however other environmental variables must be kept at an optimum. Strains improvement should also be adopted to enhance the capacity and tolerance of the degrading strains. Although biodegradation has its own challenges, it remains the best option of cleaning up PCP impacted sites.

2.13 References

ACGIH. (2014). Pentachlorophenol. In: TLVs and BEIs: based on the documentation of threshold limit values for chemical substances and physical agents and biological exposure indices. Cincinnati, OH: American Conference of Governmental Industrial Hygienists.

Aghababaie, M., Farhadian, M., Jeihanipour, A. & Biria, D. (2015). Effective factors on the performance of microbial fuel cells in wastewater treatment – a review. *Environmental Technology Reviews*, *4*(1), 71–89. https://doi.org/10.1080/09593330.2015.1077896

Ammeri, R. W., Mehri, I., Badi, S., Hassen, W. & Hassen, A. (2017). Pentachlorophenol degradation by Pseudomonas fluorescens. *Water Quality Research Journal of Canada*, *52*(2), 99–108. https://doi.org/10.2166/wqrj.2017.003

Anandarajah, K., Kiefer, P. M., Donohoe, B. S. & Copley, S. D. (2000). Recruitment of a Double Bond Isomerase To Serve as a Reductive Dehalogenase during Biodegradation of Pentachlorophenol †. *Biochemistry*, *39*(18), 5303–5311. https://doi.org/10.1021/bi9923813

Apajalahti, J. H. & Salkinoja-Salonen, M. S. (1987). Complete dechlorination of tetrachlorohydroquinone by cell extracts of pentachlorophenol-induced *Rhodococcus chlorophenolicus*. *Journal of Bacteriology*, *169*(11), 5125–5130. https://doi.org/10.1128/jb.169.11.5125-5130.1987

Arca-Ramos, A., Eibes, G., Feijoo, G., Lema, J. M. & Moreira, M. T. (2018). Enzymatic reactors for the removal of recalcitrant compounds in wastewater. *Biocatalysis and Biotransformation*, 36(3), 195–215. https://doi.org/10.1080/10242422.2017.1315411.

Aregbesola, O. A., Mokoena, M. P., Olaniran, A. O. (2020). Biotransformation of pentachlorophenol by an indigenous *Bacillus cereus* AOA-CPS1 isolated from wastewater effluent in Durban, South Africa. *Biodegradation*, 31(4-6):369-383. https://doi.org/10.1007/s10532-020-09915-w

Aronstein, B. N. and Alexander, M. (1992). Surfactant at low concentrations stimulate biodegradation of sorbed hydrocarbons in samples of aquifer sands and soil slurries. *Environ Toxicol Chem*, *11*, 1227–1233.

Arora, P. & Bae, H. (2014). Bacterial degradation of chlorophenols and their derivatives. *Microbial Cell Factories*, *13*(1), 31. https://doi.org/10.1186/1475-2859-13-31

Atashgahi, S., Liebensteiner, M. G., Janssen, D. B., Smidt, H., Stams, A. J. M. & Sipkema, D. (2018). Microbial Synthesis and Transformation of Inorganic and Organic Chlorine

Compounds. *Frontiers in Microbiology*, *9*, 3079. https://doi.org/10.3389/fmicb.2018.03079

ATSDR. (2001). Toxicological profile for pentachlorophenol. Retrieved from https://www.atsdr.cdc.gov/ToxProfiles/tp51.pdf

ATSDR. (2017). Agency for Toxic Substances and Disease Registry. Substance priority list (candidates for toxicological profiles)., (October), 190.

Auffan, M., Achouak, W., Rose, J., Roncato, M.-A., Chanéac, C., Waite, D.T., Masion, A., Woicik, J. C., Wiesner, M. R., Bottero, J-Y. (2008). Relation between the redox state of iron-based nanoparticles and their cytotoxicity toward Escherichia coli. *Environ. Sci. Technol.*, *42*, 6730–6735.

Becker, J., Hopkins, S., Jones, A. & Kiely, T. A. (2008). *Qualitative Economic Impact Assessment of Alternatives to Pentachlorophenol as a Wood Preservative. U.S. EPA. Washington D.C. April, 2008*.

Belchik, S. M. (2009). *Biochemical characterization of 2,4,6-Trichlorophenol degradation in bacterium Cupriavidus Necator Jmp134*. Retrieved From http://www.dissertations. wsu. edu/Dissertations/Spring2009/s_belchik_041409.pdf

Borch, T., Kretzschmar, R., Kappler, A., Cappellen, P. V., Gindervogel, M., Voegelin, A. & Campbell, K. (2010). Biogeochemical redox processes and their impact on contaminant dynamics. *Environmental Science & Technology*, *44*, 15–23.

Bosso, L. & Cristinzio, G. (2014). A comprehensive overview of bacteria and fungi used for pentachlorophenol biodegradation. *Reviews in Environmental Science and Bio/ Technology*, *13*, 387–427.

Bosso, L. & Cristinzio, G. (2014). A comprehensive overview of bacteria and fungi used for pentachlorophenol biodegradation. *Reviews in Environmental Science and Biotechnology*, *13*(4), 387–427. https://doi.org/10.1007/s11157-014-9342-6

Bravo, R., Caltabiano, L. M., Fernandez, C., Smith, K. D., Gallegos, M., Whitehead, R. D., Weerasekera, G., Restrepo, P., Bishop, A. M., Perez, J. J., Needham, L. L. & Barr, D. B. (2005). Quantification of phenolic metabolites of environmental chemicals in human urine using gas chromatography-tandem mass spectrometry and isotope dilution quantification. *J Chromatogr B*, *820*, 229-236.

Buffinton, G. D., Ollinger, K. & Brunmark, A. C. E. (1989). DT-diaphorase-catalysed reduction of 1,4- naphthoquinone derivatives and glutathionyl-quinone conjugates. Effect of substituents on autoxidation rates. *Biochem J*, *257*, 561–571.

Cai, M. & Xun, L. (2002). Organization and Regulation of Pentachlorophenol-Degrading

Genes in Sphingobium chlorophenolicum ATCC 39723. *Society*, *184*(17), 4672–4680. https://doi.org/10.1128/JB.184.17.4672

Cassidy, M. B., Lee, H., Trevors, J. T. & Zablotowicz, R. (1999). Chlorophenol and nitrophenol metabolism by Sphingomonas sp. UG30. *J Ind Microbiol Biotechnol*, *23*, 232–241.

Cea, M., Seaman, J. C., Jara, A. A., Mora, M. L. & Diez, M. C. (2005). Describing chlorophenols sorption on variable-charge soil using the triple-layer model. *J Colloid Interface Sci*, *292*, 171–178.

Centers for Disease Control. (2018). *Fourth National Report on Human Exposure to Environmental Chemicals Updated Table* (Vol. 1). Retrieved from http://www.cdc.gov/exposurereport/pdf/FourthReport_UpdatedTables_Sep2013.pdf

Chanama, M. & Chanana, S. (2011). expression of pentachlorophenol degradative genes of *sphingobium chlorophenolicum* ATCC39723 in *Escherichia coli*. *Asian Journal of Public Health*, *2*(2), 78–83.

Chanama, S. & Crawford, R. L. (1997). Mutational analysis of pcpA and its role in pentachlorophenol degradation by *Sphingomonas (Flavobacterium) chlorophenolica* ATCC 39723. *Applied and Environmental Microbiology*, *63*(12), 4833–4838.

CHEBI. (2016). *Pentachlorophenol (CHEBI:17642). Retrieved April 28, 2019, from* https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:17642

Chem Sources. (2016). *Pentachlorophenol. Chem Sources Online. Pendleton (SC), USA: Chemical Sources International, Inc. Available from: http://www. chemsources.com, accessed October 2016.*

Chen, L., Maloney, K., Krol, E., Zhu, B. & Yang, J. (2009). Cloning, Overexpression, Purification, and Characterization of the Maleylacetate Reductase from *Sphingobium chlorophenolicum* Strain ATCC 53874. *Current Microbiology*, *58*(6), 599–603. https://doi.org/10.1007/s00284-009-9377-z

Chen, L., Krol, E. S., Sakharkar, M. K., Khan, H. A., Alhomida, A. S. & Yang, J. (2017). Residues His172 and Lys238 are Essential for the Catalytic Activity of the Maleylacetate Reductase from *Sphingobium chlorophenolicum* Strain L-1. *SCientifiC RepoRts*, *7*. https://doi.org/10.1038/s41598-017-18475-8

Chen, M. J., Liu, C. S., Chen, P. C., Tong, H., Li, F. B., Qiao, J. T. & Lan, Q. (2016). Dynamics of the microbial community and Fe(III)-reducing and dechlorinating microorganisms in response to pentachlorophenol transformation in paddy soil. *Journal of Hazardous Materials*, *312*, 97–105.

Chen, P-J., Su, C-H., Tseng, C-Y., Tan, S-W. & Cheng, C-H. (2011). Toxicity assessments of nanoscale zerovalent iron and its oxidation products in medaka (Oryzias latipes) fish. *Marine Pollution Bulletin*, *63*, 339–346. https://doi.org/10.1016/j.marpolbul.2011.02.045.

Chen, L. & Yang, J. (2008). Biochemical characterization of the tetrachlorobenzoquinone reductase involved in the biodegradation of pentachlorophenol. *International Journal of Molecular Sciences*, *9*(3), 198–212. https://doi.org/10.3390/ijms9030198

Chen, S-T., Hsu, C-Y. & Berthouex, P. M. (2006). Fate and Modeling of Pentachlorophenol Degradation in a Laboratory-Scale Anaerobic Sludge Digester. *Journal of Environmental Engineering*, *132*(7), 795–802. https://doi.org/10.1061/(ASCE)0733-9372(2006)132:7(795)

Cheng, J., Xue, L., Zhu, M., Feng, J., Shen-Tu, J., Xu, J., Brookes, P. C. He, Y. (2019). Nitrate supply and sulfate-reducing suppression facilitate the removal of pentachlorophenol in a flooded mangrove soil. *Environmental Pollution*, *244*, 792–800. https://doi.org/10.1016/j.envpol.2018.09.143

Chung, N. & Aust, S. D. (1995). Degradation of pentachlorophenol in soil by *Phanerochaete chrysosporium*, *Journal of Hazardous Materials*, *41*, 177–183.

Copley, S. D. (2010). Evolution of Efficient Pathways for Degradation of Anthropogenic Chemicals. *PMC*, *5*(8), 559–566. https://doi.org/10.1038/nchembio.197.Evolution

Copley, S. D., Rokicki, J., Turner, P., Daligault, H., Nolan, M. & Land, M. (2012). The whole genome sequence of *sphingobium chlorophenolicum* L-1: Insights into the evolution of the pentachlorophenol degradation pathway. *Genome Biology and Evolution*, *4*(2), 184–198. https://doi.org/10.1093/gbe/evr137

Cort, T. L. & Bielefeldt, A. R. (2002). A Kinetic Model for Surfactant Inhibition of Pentachlorophenol Biodegradation, *2*. https://doi.org/10.1002/bit.10257

Crawford, R. L. & Crawford, D. L. (1996). *Bioremediation : principles and applications. In: Puhakka, J.A., Melin, E.S. (Eds.), Bioremediation of Chlorinated Phenols. University of Idaho, Moscow, ID, USA, pp. 254–299.* Cambridge University Press. Retrieved from https://books.google.co.za/books/about/Bioremediation.html?id=F3zftwAACAAJ&redir_esc=y

Crawford, R. L., Jung, C. M. & Strap, J. (2007). The recent evolution of pentachlorophenol (PCP)-4-monooxygenase (PcpB) and associated pathways for bacterial degradation of PCP. *Biodegradation*, *18*, 525–539.

D'Angelo, E. M. & Reddy, K. R. (2000). Aerobic and anaerobic transformations of pentachlorophenol in wetland soils. Soil. *Science Society of America Journal*, *64*(3),

933-943.

Dai, M., Rogers, B. J., Warner, J. R. & Copley, S. D. (2003). A Previously Unrecognized Step in Pentachlorophenol Degradation in Sphingobium chlorophenolicum Is Catalyzed by Tetrachlorobenzoquinone Reductase (PcpD). *Journal of Bacteriology*, *185*(1), 302–310. https://doi.org/10.1128/JB.185.1.302-310.2003

Dai, S., Vaillancourt, F. H., Maaroufi, H., Drouin, N. M., Neau, D. B., Snieckus, V., Bolin, J. T. & Eltis, L. D. (2002). Identification and analysis of a bottleneck in PCB biodegradation. *Nat. Struct. Biol.*, *9*, 934.

Demers, P. A, Davies, H.W., Friesen, M. C., Hertzman, C., Ostry, A., Hershler, R. T. K. (2006). Cancer and occupational exposure to pentachlorophenol and tetrachlorophenol (Canada). *CCC*, *17*, 749–758.

Ding, L. J., Su, J. Q., Xu, H. J., Jia, Z. J. & Zhu, Y. G. (2015). Long-term nitrogen fertilization of paddy soil shifts iron-reducing microbial community revealed by RNA-13C-acetate probing coupled with pyrosequencing. *The ISME Journa*, *l 9*, 721–734.

Ding, Y., Liu, Y., Liu, S., Li, Z., Tan, X., Huang, X., Zeng, G., Zhou, L. & Zheng, B. (2016). Biochar to improve soil fertility. A review. *Agronomy for Sustainable Development*, *36*(2), 36. https://doi.org/10.1007/s13593-016-0372-z

Dong, Y. L., Zhou, P. J., Jiang, S. Y., Pan, X. W., Zhao, X. H. (2009). Induction of oxidative stress and apoptosis by pentachlorophenol in primary cultures of Carassius carassius hepatocytes. *Comp. Biochem. Physiol. C*, *150*(2), 179–185.

Ederer, M. M., Crawford, R. L., Herwig, R. P. & Orser, C. S. (1997). PCP degradation is mediated by closely related strains of the genus Sphingomonas. *Mol Ecol*, *6*, 39–49.

El-Bialy, H. A., Khalil, O. A. A. & Gomaa, O. M. (2018). Bacterial-mediated biodegradation of pentachlorophenol via electron shuttling. *Environmental Technology*, 1–9. https://doi.org/10.1080/09593330.2018.1442501

El-Temsah, Y. S., Oughton, D. H. and Joner, E. J. (2013). Effects of nano-sized zero-valent iron on DDT degradation and residual toxicity in soil: A column experiment. *Plant and Soil*, *368*, 189–200. https://doi.org/10.1007/s11104-012-1509-8.

El-Temsah, Y.S. & Joner, E. J. (2012). Ecotoxicological effects on earthworms of fresh and aged nano-sized zero-valent iron (nZVI) in soil. *Chemosphere*, *89*, 76–82. https://doi.org/10.1016/j.chemosphere.2012.04.020.

El-Temsah, Y. S. and Joner, E. J. (2013). Effects of nano-sized zero-valent iron (nZVI) on DDT degradation in soil and its toxicity to collembola and ostracods,. *Chemosphere*, *92*, 131–137.

Elango, V., Kurtz, H. D. & Freedman, D. L. (2011). Aerobic cometabolism of trichloroethene and cis-dichloroethene with benzene and chlorinated benzenes as growth substrates. *Chemosphere*, *84*(2), 247–253.

Elango, V. (2010). *Biodegradation and bioremediation of hexachlorocyclohexane isomers, chlorinated ethenes, chlorinated benzenes and benzene*. Retrieved from http://tigerprints.clemson.edu/all_dissertations

EPA. (2008). *Reregistration Eligibility Decision for pentachlorophenol. U.S. EPA report, september 2008, available via http://www.epa.gov/oppsrrd1/REDs/pentachlorophenol_red.pdf.*

EPA. (2010). Toxicological review of pentachlorophenol (CAS No. 87-86-5). In support of summary information on the Integrated Risk Information System (IRIS). Washington (DC), USA: United States Environmental Protection Agency.

European Parliament. (2008). Council of the European Union [2008]. Regulation (EC) No 1272/2008 of the European Parliament and of the Council of 16 December 2008 on classification, labeling and packaging of substances and mixtures, amending and repealing Directives 67/548/.

Fathy, R., Gomaa, O. M., Ali, A. E-H., Abd El Kareem, H., Abou Zaid, M. (2016). Neutral red as a mediator for the enhancement of electricity production using domestic wastewater double chamber microbial fuel cell. *Ann Microbiol*, *66*, 695–702.

Field, J. A. & Sierra-Alvarez, R. (2008). Microbial degradation of chlorinated phenols. *Reviews in Environmental Science and Bio/Technology*, *7*(3), 211–241. https://doi.org/10.1007/s11157-007-9124-5

Fisher, B. (1991). Pentachlorophenol: toxicology and environmental fate. *J Pesticide Reform*, *11*(1), 2–5.

Frascari, D., Zanaroli, G. & Danko, A. S. (2015). In situ aerobic cometabolism of chlorinated solvents: A review. *Journal of Hazardous Materials*, *283*, 382–399. https://doi.org/10.1016/J.JHAZMAT.2014.09.041

Galve, R., Sanchez-Baeza, F., Camps, F. & Marco, M. P. (2002). Indirect competitive immunoassay for trichlorophenol determination - Rational evaluation of the competitor heterology effect. *Anal Chim Acta*, *452*, 191-206.

Gans, J., Wolinsky, M. & Dunbar, J. (2005). Computational improvements reveal great bacterial diversity and high metal toxicity in soil. *Science*, *309*, 1387-1390.

Ghasemi, M., Wan Daud, W. R., Ismail, M., Rahimnejad, M., Ismail, A. F., Leong, J. X., Miskan, M. Liew, K. B. (2013). Effect of pre-treatment and biofouling of proton exchange membrane on microbial fuel cell performance. *International Journal of*

Hydrogen Energy, *38*(13), 5480–5484. https://doi.org/10.1016/j.ijhydene.2012.09.148

Goethals, K. M., Montagu, V. and Holsters, M. (1992). Conserved motifs in a divergent nod box of *Azorhizobium caulinodans* ORS571 reveal a common structure in promoters regulated by LysR-type proteins. *Proc. Natl. Acad. Sci. USA*, *89*, 1646–1650.

Hayes, R. P., Green, A. R., Nissen, M. S., Lewis, K. M., Xun, L. & Kang, C. (2014). Structural characterization of 2,6-dichloro-p-hydroquinone 1,2- dioxygenase (PcpA) from *Sphingobium chlorophenolicum*, a new type of aromatic ring-cleavage enzyme Robert, *154*(11), 2262–2265. https://doi.org/10.1016/j.pain.2013.06.005.Re-Thinking

Grobler, D. F., Badenhorst, J. E. & Kempster, P. L. (1996). PCBs, chlorinated hydrocarbon pesticides and chlorophenols in the Isipingo estuary, Natal, Republic of South Africa. *Mar Pollut Bull*, *32*, 572–575.

Herrera, Y., Okoh, A. I., Alvarez, L., Robledo, N. & Trejo-Hernández, M. R. (2008). Biodegradation of 2,4-dichlorophenol by a Bacillus consortium. *World Journal of Microbiology and Biotechnology*, *24*(1), 55–60. https://doi.org/10.1007/s11274-007-9437-0

Hickman, Z. A. & Reid, B. J. (2008). Earthworm assisted bioremediation of organic contaminants. *Environment International*, *34*, 1072–1081.

Hlouchova, K., Rudolph, J., Pietari, J. M. H., Behlen, L. S. & Copley, S. D. (2012). Pentachlorophenol hydroxylase, a poorly functioning enzyme required for degradation of pentachlorophenol by sphingobium chlorophenolicum. *Biochemistry*, *51*(18), 3848–3860. https://doi.org/10.1021/bi300261p

Hong, Y., Chen, X., Guo, J., Xu, Z., Xu, M. Sun, G. (2007). Effects of electron donors and acceptors on anaerobic reduction of azo dyes by *Shewanella decolorationis* S12. *Appl Microbiol Biotechnol.*, *74*, 230–238.

Huang, Y., Xun, R., Chen, G. & Xun, L. (2008). Maintenance role of a glutathionyl-hydroquinone lyase (PcpF) in pentachlorophenol degradation by Sphingobium chlorophenolicum ATCC 39723. *J Bacteriol*, *190*, 7595–7600.

Huang, L., Gan, L., Zhao, Q., Logan, B. E., Lu, H. & Chen, G. (2011). Degradation of pentachlorophenol with the presence of fermentable and non-fermentable co-substrates in a microbial fuel cell. *Bioresource Technology*, *102*(19), 8762–8768. https://doi.org/10.1016/j.biortech.2011.07.063

IARC. (2012). *Agents reviewed by the IARC monographs. In: IARC monographs on the evaluation of carcinogenic risks to humans. International Agency for Research on Cancer, http://monographs.iarc.fr/ENG/Monographs/ PDFs/index.php.*

IARC. (2019). IARC monographs on the evaluation of carcinogenic risks to humans. In

IARC (Ed.), *Pentachlorophenol and some related compounds* (17th ed., pp. 33–140). Lyon, France.

Joseph, S. D., Camps-Arbestain, M., Lin, Y., Munroe, P., Chia, C. H., Hook, , J., van Zwieten, L., Kimber, S., Cowie, A., Singh, B. P., Lehmann, J., Foidl, N., Smernik , R. J. & Amonette, J. E. (2010). An investigation into the reactions of biochar in soil. *Aust. J. Soil Res*, *48*(7), 501–515. https://doi.org/10.1071/SR10009

Joshi, V. V., Prewitt, M. L., Ma, D.-P. & Borazjani, H. (2015). Enhanced Remediation of Pentachlorophenol (PCP)-Contaminated Groundwater by Bioaugmentation with Known PCP-Degrading Bacteria. *Bioremediation Journal*, *19*(2), 160–170. https://doi.org/10.1080/10889868.2014.995369

Juwarkar, A. A., Singh, S. K. Mudhoo, A. (2010). A comprehensive overview of elements in bioremediation. *Rev Environ Sci Biotechnol*, *9*, 215–288.

Kao, C. M., Chai, C. T., Liu, J. K., Yeh, T. Y., Chen, K. F. Chen, S. C. (2004). Evaluation of natural and enhanced PCP biodegradation at a former pesticide manufacturing plant. *Water Res.*, *38*, 663–672.

Kappler, A., Wuestner, M. L., Ruecker, A., Harter, J., Halama, M. & Behrens, S. (2014). Biochar as an Electron Shuttle between Bacteria and Fe(III) Minerals. *Environmental Science & Technology Letters*, *1*(8), 339–344. https://doi.org/10.1021/ez5002209

Karlson, U., Rojo, F., vanElsas, J. D. and Moore, E. (1996). Genetic and serological evidence for the recognition of four pentachlorophenol degrading bacterial strains as a species of the genus Sphingomonas. *Syst Appl Microbiol*, *18*, 539–548.

Karns, J. K., Killbane, J. J., Duttagupta, S. & Chakrabarty, A. M. (1983). Metabolism of halophenol by 2,3,5-trichlorophenoxyacetic acid degrading *Pseudomonas cepecia*. *Appl Environ Microbiol.*, *46*, 1176–1181.

Keenan, C. R., Goth-Goldstein, R., Lucas, D. and Sedlak, D. L. (2009). Oxidative Stress Induced by Zero-Valent Iron Nanoparticles and Fe(II) in Human Bronchial Epithelial Cells. *Environ. Sci. Techno*, *43*, 4555–4560.

Keenan, J. E., Strand, S. E. & Stensel, H. D. (1994). Degradation kinetics of chlorinated solvents by a propane-oxidizing enrichment culture. In: In ed.). R. E. Hinchee (Ed.), *Bioremediation of Chlorinated and Polycyclic Aromatic Hydrocarbon Compounds (R. E. Hinchee, ed.).* (pp. 1–13). Lewis Publishers, Boca Raton, Florida, USA.

Keener, W. K. & Arp, D. J. (1993). Kinetic studies of ammonia monooxygenase inhibition in Nitrosomonas europaea by hydrocarbons and halogenated hydrocarbons in an optimized whole-cell assay. *Applied and Environmental Microbiology*, *59*(8), 2501–2510.

Khan, M. D., Khan, N., Sultana, S., Joshi, R., Ahmed, S., Yu, E., Scott, K., Ahmad, A. Khan, M. Z. (2017). Bioelectrochemical conversion of waste to energy using microbial fuel cell technology. *Process Biochemistry*, *57*, 141–158. https://doi.org/10.1016/j.procbio.2017.04. 001

Khan, M. Z., Singh, S., Sultana, S., Sreekrishnan, T. R. & Ahammad, S. Z. (2015). Studies on the biodegradation of two different azo dyes in bioelectrochemical systems. *New Journal of Chemistry*, *39*(7), 5597–5604. https://doi.org/10.1039/C5NJ00541H

Khan, N., Khan, M. D., Nizami, A.-S., Rehan, M., Shaida, A., Ahmad, A. & Khan, M. Z. (2018). Energy generation through bioelectrochemical degradation of pentachlorophenol in microbial fuel cell. *RSC Advances*, *8*(37), 20726–20736. https://doi.org/10.1039/C8RA01643G

Khessairi, A., Fhoula, I., Jaouani, A., Turki, Y., Cherif, A., Boudabous, A., Hassen, A. & Ouzari, H. (2014). Pentachlorophenol degradation by *Janibacter* sp., a new Actinobacterium isolated from saline sediment of arid land. *BioMed Research International*, *2014*, 1–9. https://doi.org/http://dx.doi.org/10.1155/2014/296472

Kiefer, P. M., McCarthy, D. L. & Copley, S. D. (2002). The reaction catalyzed by tetrachlorohydroquinone dehalogenase does not involve nucleophilic aromatic substitution. *Biochemistry*, *41*(4), 1308–1314. https://doi.org/10.1021/bi0117495

Kiely, P. D., Regan, J. M. & Logan, B. E. (2012). The electric picnic: synergistic requirements for exoelectrogenic microbial communities. *Curr. Opin. Biotechnol.*, *22*, 378–385.

Kim, Y. H. & Carraway, E. R. (2000). Dechlorination of pentachlorophenol by zero valent iron and modified zero valent irons. *Environmental Science and Technology*, *34*(10), 2014–2017. https://doi.org/10.1021/es991129f

Kitunen, V. H., Valo, R. J. & Salkinojasalonen, M. S. (1987). Contamination of soil around wood-preserving facilities by polychlorinated aromatic-compounds. *Environ Sci Technol*, (21), 96–101.

Klüpfel, L., Keiluweit, M., Kleber, M. & Sander, M. (2014). Redox Properties of Plant Biomass-Derived Black Carbon (Biochar). *Environmental Science & Technology*, *48*(10), 5601–5611. https://doi.org/10.1021/es500906d

Kracke, F., Vassilev, I. & Krömer, J. O. (2015). Microbial electron transport and energy conservation - the foundation for optimizing bioelectrochemical systems. *Frontiers in Microbiology*, *6*, 575. https://doi.org/10.3389/fmicb.2015.00575

Kuang, F., Li, Y., He, L., Xia, Y., Li, S. & Zhou, J. (2018). Cometabolism degradation of lignin in sequencing batch biofilm reactors. *Environmental Engineering Research*,

23(3), 294–300. https://doi.org/10.4491/eer.2017.201

Laha, S. & Luthy, R. G. (1992). Effects of nonionic surfactants on the solubilization and mineralization of phenanthrene in soil-water systems. *Biotechnol Bioeng*, *40*, 46–53.

Lee, D-H., Koh, E-H., Choi, S-R., Kim, S., Kim, D. H., & Lee, N. Y. (2013). Effect of sodium citrate on growth of bacteria in blood culture. *Ann Clin Microbiol.*, *16*, 168–173.

Lee, J-g., Xee, L. (1997). Purification and Characterization of 2, 6-Dichloro- p -Hydroquinone Chlorohydrolase from Flavobacterium sp . *Journal of Bacteriology*, *179*(5), 1521–1524.

Lenters, V., Iszatt, N., Forns, J., Čechová, E., Kočan, A., Legler, J., Leonards, P., Stigum, H. & Eggesbø, M. (2019). Early-life exposure to persistent organic pollutants (OCPs, PBDEs, PCBs, PFASs) and attention-deficit/hyperactivity disorder: A multi-pollutant analysis of a Norwegian birth cohort. https://doi.org/10.1016/j.envint.2019.01.020

Leung, K. T., Campbell, S., Gan, Y., White, D. C., Lee, H. & Trevors, J. T. (1999). The role of the *Sphingomonas* species UG30 pentachlorophenol-4-monooxygenase in p-nitrophenol degradation. *FEMS Microbiol Lett*, *173*, 247–253.

Li, J. (2013). An experimental study of microbial fuel cells for electricity generating: performance characterization and capacity improvement. *Journal of Sustainable Bioenergy Systems*, *03*(03), 171–178. https://doi.org/10.4236/jsbs.2013.33024

Li, W., Zhang, S., Chen, G. & & Hua, Y. (2014). Simultaneous electricity generation and pollutant removal in microbial fuel cell with denitrifying biocathode over nitrite. *Applied Energy*, *126*, 136–141. https://doi.org/10.1016/j.apenergy.2014.04.015

Li, X., Lin, Z., Luo, C., Bai, J., Sun, Y. & Li, Y. (2015). Enhanced microbial degradation of pentachlorophenol from soil in the presence of earthworms: Evidence of functional bacteria using DNA-stable isotope probing. *Soil Biology and Biochemistry*, *81*, 168–177. https://doi.org/10.1016/j.soilbio.2014.11.011

Li, H., Chen, S., Ren, L. Y., Zhou, L. Y., Tan, X. J., Zhu, Y., Belver, C., Bedia, J. & Yang, J. (2019). Biochar mediates activation of aged nanoscale ZVI by *Shewanella putrefaciens* CN32 to enhance the degradation of Pentachlorophenol. *Chemical Engineering Journal*, *368*, 148–156. https://doi.org/10.1016/j.cej.2019.02.099

Li, H., Li, Y., Cao, H., Li, X. & Zhang, Y. (2011). Degradation of pentachlorophenol by a novel peroxidase-catalyzed process in the presence of reduced nicotinamide adenine dinucleotide. *Chemosphere*, *83*(2), 124–130. https://doi.org/10.1016/j.chemosphere.2011.01.014

Li, X., Lin, Z., Luo, C., Bai, J., Sun, Y. & Li, Y. (2015). Enhanced microbial degradation of

pentachlorophenol from soil in the presence of earthworms: Evidence of functional bacteria using DNA-stable isotope probing. *Soil Biology and Biochemistry*, *81*, 168–177. https://doi.org/10.1016/j.soilbio.2014.11.011

Li, Y., Li, B., Wang, C-P., Fan, J-Z. & Sun, H-W. (2014). Aerobic degradation of trichloroethylene by co-metabolism using phenol and gasoline as growth substrates. *International Journal of Molecular Sciences*, *15*(5), 9134–9148. https://doi.org/10.3390/ijms15059134

Lin, Z., Li, X. M., Li, Y. T., Huang, D. Y., Dong, J. & Li, F.B. (2012). Enhancement of two ecological earthworm species (*Eisenia foetida* and *Amynthas robustus E. Perrier*) on removal and degradation processes of soil DDT. *Journal of Environmental Monitoring*, *14*, 1551–1558.

Logan, B. E., Hamelers, B., Rozendal, R., Schröder, U., Keller, J., Freguia, S., Aelterman, P., Verstraete, W. & Rabaey, K. (2006). Microbial Fuel Cells: Methodology and Technology †. *Environmental Science & Technology*, *40*(17), 5181–5192. https://doi.org/10.1021/es06 05016

Lopez-Echartea, E., Macek, T., Demnerova, K., & Uhlik, O. (2016). Bacterial biotransformation of pentachlorophenol and micropollutants formed during its production process. *International Journal of Environmental Research and Public Health*, *13*(11). https://doi.org/10.3390/ ijerph13111146

Luepromchai, E., Singer, A. C., Yang, C. H. & Crowley, D. E. (2002). Interactions of earth-worms with indigenous and bioaugmented PCB-degrading bacteria. *FEMS Microbiology Ecology*, *41*, 191–197.

Lumar, R. T. & Glaser, J. A. (1994). *Field evaluation of the remediation of soils contaminated with wood-preserving chemicals using lignin-degrading fungi.* (O. S. Hinchee RE, Leeson A, Semprini L, Ed.) (Bioremedia). Boca Raton: Lewis Pubisher.

Luo, W., Zhu, X., Chen, W., Duan, Z., Wang, L. & Zhou, Y. (2014). Mechanisms and strategies of microbial cometabolism in the degradation of organic compounds – chlorinated ethylenes as the model. *Water Science & Technology*, *69*(10), 1971. https://doi.org/10.2166/wst.2014. 108

Ma, M., Wang, C. X. & Wang, Z. J. (2005). ssessing toxicities of hydrophobic organic pollutants in Huaihe River by using two types of sampling. *A. J Environ Sci Heal*, *40*, 331–342.

Machonkin, T. E, Holland, P. L., Smith, K. N., Liberman, J. S., Dinescu, A., Cundari, T. R. & Rocks, S. S. (2010). Determination of the Active Site of *Sphingobium chlorophenolicum*

2,6-dichlorohydroquinone dioxygenase (PcpA). *J. Biol. Inorg. Chem.*, *10*, 291.

Machonkin, T. E. & Doerner, A.E., 2011. Substrate Specificity of *Sphingobium chlorophenolicum* 2,6-Dichlorohydroquinone 1,2-Dioxygenase. Biochemistry 50, 8899–8913. *Biochemistry*, *50*(41), 8899–8913. https://doi.org/10.1021/bi200855m

Mathialagan, T. & Viraraghavan, T. (2009). Biosorption of pentachlorophenol from aqueous solutions by a fungal biomass. *Bioresource Technology*, *100*(2), 549–558. https://doi.org/10.1016/j.biortech.2008.06.054

McAllister, K.A.; Lee, H.; Trevors, J. T. (1996). Microbial degradation of pentachlorophenol. *Biodegradation*, *7*, 1–40.

McCarthy, D. L., Navarrete, S., Willett, W. S., Babbitt, P. C. & Copley, S. (1996). Exploration of the relationship between tetrachlorohydroquinone dehalogenase and the glutathione S-transferase superfamily. *Biochem*, *35*, 14634–14642.

McLean, D., Eng, A., Dryson, E., Walls, C., Harding, E., Wong, K. C., Cheng, S., Mannetje, A., Ellison-Loschmann, L., Slater, T., Shoemack, P. P. N. (2009). Morbidity in former sawmill workers exposed to pentachlorophenol (PCP): A cross-sectional study in New Zealand. *Am J Ind Med*, *52*, 271–281.

Megharaj, M., Ramakrishnan, B., Venkateswarlu, K., Sethunathan, N. & Naidu, R. (2011). Bioremediation approaches for organic pollutants: A critical perspective. *Environment International*, *37*(8), 1362–1375. https://doi.org/10.1016/j.envint.2011.06.003

Mikoczy, Z. & Hagmar, L. (2005). Cancer incidence in the Swedish leather tanning industry: updated findings 1958-99. *Occup Environ Med*, *62*(7), 461–4. https://doi.org/10.1136/oem.2004.017038

Milton, R. D. & Minteer, S. D. (2017). Direct enzymatic bioelectrocatalysis: differentiating between myth and reality. *Journal of the Royal Society, Interface*, *14*(131). https://doi.org/10.1098/rsif.2017.0253

Miyauchi, K., Adachi, Y., Nagata, Y. & Takagi, M. (1999). Cloning and Sequencing of a Novel meta-cleavage dioxygenase gene whose product is involved in degradation of γ-hexachlorocyclohexane in *Sphingomonas paucimobilis*. *J. Bacteriol.*, *181*, 6712.

Mohn, W. W. & Tiedje, J. (1992). Microbial reductive dehalogenation. *Microbiol Rev*, *56*, 482–507.

Mun, C. H., He, J. & Ng, W. J. (2008). Pentachlorophenol dechlorination by an acidogenic sludge. *Water Research*, *42*, 3789–3798.

Nemir, A., David, M., Perrussel, R., Sapkota, A., Simonet, P. & Monier, J. M., Vogel, T. M. (2010). Comparative phylogenetic microarray analysis of microbial communities in

TCE-contaminated soils. *Chemosphere*, *80*, 600–607.

NIOSH. (2005). NIOSH pocket guide to chemical hazards. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication No. 2005-149, http:// www.cdc.

NIOSH. (2019). *NIOSH skin notation profile: pentachlorophenol (PCP). By Hudson NL. Cincinnati, OH: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH) Publication N*. https://doi.org/https://doi.org/10.26616/NIOSHPUB2019120

Noordman, W. H., Wachter, J. H. J., de Boer, G. J.& Janssen, D. B. (2002). The enhancement by surfactants of hexadecane degradation by Pseudomonas aeruginosa varies with substrate availability. *Journal of Biotechnology*, *94*(2), 195–212.

Nzila, A. (2013). Update on the cometabolism of organic pollutants by bacteria. *Environmental Pollution*, *178*, 474–482. https://doi.org/10.1016/j.envpol.2013.03.042

Ohtsubo, Y., Miyauchi, K ., Kanda, K ., Hatta, T., Kiyohara, H., Senda, T., Nagata, Y., Mitsui, Y. & Takagi, M . (1999). PcpA, which is involved in the degradation of pentachlorophenol in *Sphingomonas chlorophenolica* ATCC39723, is a novel type of ring-cleavage dioxygenase. *FEBS Letters*, *459*(3), 395–398. https://doi.org/10.1016/S0014-5793(99)01305-8

Olaniran A. O. & Igbinosa E O. (2011). Chlorophenols and other related derivatives of environmental concern: properties, distribution and microbial degradation processes. *Chemosphere*, *83*, 1297–1306.

Olujimi, O., Fatoki, O., Odendaal, J. & Daso, A. (2012). Chemical monitoring and temporal variation in levels of endocrine disrupting chemicals (priority phenols and phthalate esters) from selected wastewater treatment plant and freshwater systems in Republic of South Africa. *Microchem. J.*, *101*, 11–23. https://doi.org/101, 11–23. https://doi.org/10.1016/j. microc.2011. 09.011

Orser, C. S., Lange, C. C., Xun, L., Zahrt, T. C. & Schneider, B. J. (1993). Cloning, sequence analysis, and expression of the Flavobacterium pentachlorophenol-4-monooxygenase gene in Escherichia coli. *J. Bacteriol.*, *175*, 411– 416.

Orser, C. S., Dutton, J., Lange, C., Jablonski, P., Xun, L. & Hargis, M. (1993). *Characterization of a Flavobacterium glutathione S-transferase gene involved in reductive dechlorination*. *Journal of Bacteriology* (Vol. 175). Retrieved from https://www.ncbi.nlm.nih.gov/ pmc/ articles/PMC204566/pdf/jbacter00051-0168.pdf

Orser, C. S., Lange, C. C., Xun, L., Zahrt, T. C. & Schneider, B. J. (1993). Cloning, sequence analysis, and expression of the Flavobacterium pentachlorophenol-4-monooxygenase gene in Escherichia coli. *J Bacteriol*, *175*(2), 411–416.

OSHA. (2018). Pentachlorophenol. In: OSHA occupational chemical database, Washington DC: U.S. Department of Labor, Occupational Safety and Health Administration, https:// www.osha.gov/chemicaldata/chemResult. HTML?recNo=726.

Patel, B. P. & Kumar, A. (2016a). Multi-substrate biodegradation of chlorophenols by defined microbial consortium. *3 Biotech*, *6*(2). https://doi.org/10.1007/s13205-016-0511-x

Patel, B. P. & Kumar, A. (2016b). Optimization study for maximizing 2,4-dichlorophenol degradation by *Kocuria rhizophila* strain using response surface methodology and kinetic study. *Desalination and Water Treatment*, *57*(39), 18314–18325. https://doi.org/10.1080/ 19443994.2015.1091988

Pera-Titus, M., García-Molina, V., Baños, MA., Giménez, J. & Esplugas, S. (2004). Degradation of chlorophenols by means of advanced oxidation processes: a general review. *Appl Catal B Environ*, *47*(4), 219–256.

Pereira, L., Pereira, R., Pereira, M. F. R. & Alves, M. M. (2016). Effect of different carbon materials as electron shuttles in the anaerobic biotransformation of nitroanilines. *Biotechnology and Bioengineering*, *113*(6), 1194–1202. https://doi.org/10.1002/bit.25896

Persson, Y., Lundstedt, S., Oberg, L. & Tysklind, M. (2007). Levels of chlorinated compounds (CPs. PCPPs,PCDEs, PCDFs and PCDDs) in soils at contaminated sawmill sites in Sweden. *Chemosphere*, *66*, 234–242.

Phenrat, T., Long, T. C., Lowry, G. V. & Veronesi, B. (2009). Partial oxidation ("Aging") and surface modification decrease the toxicity of nanosized zerovalent Iron. *Environmental Science and Technology*, *43*, 195–200.

Pierce, R. H. & Victor, D. M. (1978). The Fate of Pentachlorophenol in an Aquatic Ecosystem. In K. R. Rao (Ed.), *Pentachlorophenol: chemistry, pharmacology, and environmental toxicology*. (pp. 27-39.). New York.: Plenum Press.

Preston, M. D., Smemo, K. A., McLaughlin, J. W. & Basiliko, N. (2012). Peatland microbial communities and decomposition processes in the James Bay Lowlands, Canada. *Frontiers in Microbiology*, *3*, 70. https://doi.org/10.3389/fmicb.2012.00070

Proudfoot, A. T. (2003). Pentachlorophenol poisoning. *Toxicol Rev*, *22*, 3–11.

PubChem. (2018). Pentachlorophenol. Compound summary for CID 992. PubChem. Open

Chemistry Database. Bethesda (MD), USA: National Center for Biotechnology Information, United States National Library of Medicine. Available from: https://pubchem.ncbi.nlm.nih.gov/compound/ pe.

Puglisi, E., Vernile, P., Bari, G., Spagnuolo, M., Trevisan, M., de Lillo, E. & Ruggiero, P. (2009). Bioaccessibility, bioavailability and ecotoxicity of pentachlorophenol in compost amended soils. *Chemosphere*, *77*, 80-86.

Reddy, G. V. B. & Gold, M. H. (2000). Degradation of pentachlorophenol by *Phanerochaete chrysosporium* : intermediates and reactions involved. *Microbiology*, *1612*(146), 17–405. Retrieved from http://www.microbiologyresearch. org/docserver/fulltext/micro/146/ 2/1460405a.pdf?expires=1527351713&id=id&accname=guest&checksum=9F6BEEB27E5887ADED4E4CDFB1A3697B

Ren, H., Li, Q., Zhan, Y., Fang, X. & Yu, D. (2016). 2,4-Dichlorophenol hydroxylase for chlorophenol removal: Substrate specificity and catalytic activity. *Enzyme and Microbial Technology*, *82*, 74–81. https://doi.org/10.1016/j.enzmictec.2015.08.008

Rieble, S., Joshi, D. K. & Gold, M. H. (1994). Purification and characterization of a 1,2,4-trihydroxybenzene 1,2-dioxygenase from the basidiomycete *Phanerochaete chrysosporium*. *Journal of Bacteriology*, *176*(16), 4838–4844. https://doi.org/10.1128/jb.176.16.4838-4844.1994

Rodriguez-Campos, J., Dendooven, L., Alvarez-Bernal, D. & Contreras-Ramos, S. M. (2014). Potential of earthworms to accelerate removal of organic contaminants from soil: a review. *Applied Soil Ecology*, *79*, 10–25.

Rokicki, J. F. (2015). *Regulatory Organization and Transcriptional Response of Sphingobium Chlorophenolicum to the Anthropogenic Pesticide Pentachlorophenol*. University of Colorado Boulder.

Rüttimann-Johnson, C. & Lamar, R. T. (1997). Binding of substances in pentachlorophenol to humic soil by the action of white rot fungi. *Soil Biol Biochem*, *29*, 1143–1148.

Rubilar, O., Diez, M. C. & Gianfreda, L. (2008). Transformation of chlorinated phenolic compounds by white rot fungi. *Crit Rev Environ Sci Technol*, *38*, 227–268.

Rubilar, O., Tortella, G. R., Cuevas, R., Cea, M., Rodríguez-Couto, S. & Diez, M. (2012). Adsorptive removal of pentachlorophenol by Anthracophyllum discolor in a fixed-bed column reactor. *Water Air Soil Pollut*, *223*, 2463–2472.

Rubilar, O., Diez, M. C. & Gianfreda, L. (2008). Transformation of chlorinated phenolic compounds by white rot fungi. *Critical Reviews in Environmental Science and*

Technology, *38*(4), 227–268. https://doi.org/10.1080/10643380701413351

Rudolph, J., Erbse, A. H., Behlen, L. S. & Copley, S. D. (2014). A radical intermediate in the conversion of pentachlorophenol to tetrachlorohydroquinone by *sphingobium chlorophenolicum*. *Biochemistry*, *53*(41), 6539–6549. https://doi.org/10.1021/bi5010427

Sahm, H., Brunner, M. & Schoberth, S. (1986). Anaerobic degradation of halogenated aromatic compounds. *Microb Ecol*, *12*, 147–153.

Saquing, J. M., Yu, Y-H. & Chiu, P. C. (2016). Wood-Derived Black Carbon (Biochar) as a Microbial Electron Donor and Acceptor. *Environmental Science & Technology Letters*, *3*(2), 62–66. https://doi.org/10.1021/acs.estlett.5b00354

Schaefer, M., Peterson, S. O. & Filser, J. (2005). Effects of *Lumbricus terrestris*, *Allolobophora chlorotica* and *Eisenia fetida* on microbial community dynamics in oil-contaminated soil. *Soil Biology Biochemistry*, *37*, 2065–2076.

Schell, M. A. (1993). Molecular biology of the LysR family of transcriptional regulators. *Annu. Rev. Microbiol*, *47*, 597–626.

Seech, A. G., Trevors, J. T. & Bulman, T. L. (1991). Biodegradation of pentachlorophenol in soil the response to physical chemical and biological treatments. *Can J Microbiol*, *37*, 440–444.

Seibert, V., Kourbatova, E. M., Golovleva, L. A., Schlo¨mann, M. & Schlömann, S. (1998). Characterization of the Maleylacetate Reductase MacA of Rhodococcus opacus 1CP and Evidence for the Presence of an Isofunctional Enzyme. *Journal Of Bacteriology* (Vol. 180). Retrieved from http://jb.asm.org/

Shan, J., Xu, J., Zhou, W. Q., Ji, L. L., Cui, Y. B., Guo, H. Y., Ji, R. (2011). Enhancement of chlorophenol sorption on soil by geophagous earthworms (Metaphire guillelmi). *Chemosphere*, *82*, 156–162.

Sheehan, D., Meade, G., Foley, VM. & Dowd, C. (2001). Structure, function, and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochem J.*, *360*, 1–16.

Shukla, A. K., Singh, R. S., Upadhyay, S. N. & Dubey, S. K. (2010). Kinetics of bio-filtration of trichloroethylene by methanotrophs in presence of methanol. *Bioresource Technology*, *101*(21), 8119–8126. https://doi.org/10.1016/j.biortech.2010.06.040

Sing, N. N., Zulkharnain, A., Roslan, H. A., Assim, Z. & Husaini, A. (2014). Bioremediation of PCP by Trichoderma and Cunninghamella Strains Isolated from Sawdust. *Brazilian Archieves of Biology and Technology*, *57*(6), 811–820.

Su, Y., Chen, L., Bandy, B. & Yang, J. (2008). The Catalytic Product of Pentachlorophenol

4-Monooxygenase is Tetra-chlorohydroquinone rather than Tetrachlorobenzoquinone. *The Open Microbiology Journal*, *2*, 100–106. https://doi.org/10.2174/1874285800802010100

Sun, W., Sammynaiken, R., Chen, L., Maley, J., Schatte, G., Zhou, Y. & Yang, J. (2011). *Sphingobium chlorophenolicum* dichlorohydroquinone dioxygenase (PcpA) is alkaline resistant and thermally stable. *International Journal of Biological Sciences*, *7*(8), 1171–1179. https://doi.org/10.7150/ijbs.7.1171

Szewczyk, R. & Dlugonski, J. (2009). Pentachlorophenol and spent engine oil degradation by Mucor ramosissimus. *International Biodeterioration and Biodegradation*, *63*, 123–129.

Takeuchi, M., Hamana, K. & Hiraishi, A. (2001). Proposal of the genus *Sphingomonas sensu stricto* and three new genera, Sphingobium, Novosphingobium and Sphingopyxis, on the basis of phylogenetic and chemotaxonomic analyses. *Int. J. Syst. Evol. Microbiol.*, *51*, 1405–1417.

Thakur, I. S., Verma, P. & Upadhayaya, K. (2002). Molecular cloning and characterization of pentachlorophenol-degrading monooxygenase genes of *Pseudomonas* sp. from the chemostat. *Biochem Biophys Res Commun*, *290*, 770–774.

Tiirola, M. A., Mannisto, M. K., Puhakka, J. A. & Kulomaa, M. S. (2002). Isolation and Characterization of *Novosphingobium* sp. Strain MT1, a Dominant Polychlorophenol-Degrading Strain in a Groundwater Bioremediation System. *Applied and Environmental Microbiology*, *68*(1), 173–180. https://doi.org/10.1128/AEM.68.1.173-180.2002

Tong, H., Hu, M., Li, F., Chen, M. & Lv, Y. (2015). Burkholderiales participating in pentachlorophenol biodegradation in iron-reducing paddy soil as identified by stable isotope probing. *Environ Sci Process Impacts*, *17*, 1282–1289.

Tong, H., Hu, M., Li, FB., Liu, C. S. & Chen, M. J. (2014). Biochar enhances the microbial and chemical transformation of pentachlorophenol in paddy soil. *Soil Biol Biochem*, *70*, 142–150.

Tyagi, M., da Fonseca, M. M. R. & de Carvalho, C. C. C. . (2011). Bioaugmentation and biostimulation strategies to improve the effectiveness of bioremediation processes. *Biodegradation*, *22*, 231–241.

UNEP. (1996). *UNEP United Nations Environment Programme. Pentachlorophenol and its salts and esters. Decision guidance document Rome-Geneva, 1991, amended 1996.*

UNEP. (2011a). *Stockholm Convention on Persistent Organic Pollutants (UNEP/POPS/POPRC.7/INF5).*

UNEP. (2011b). *The Stockholm Convention on Persistent Organic Pollutants.*

(UNEP/POPS/POPRC.7/4). The American Journal of International Law (Vol. 95). https://doi.org/10.2307/2668517

UNEP. (2013). *United Nations SC Stockholm Convention on Persistent Organic Pollutants Report of the Persistent Organic Pollutants Review Committee on the work of its ninth meeting Addendum Risk profile on pentachlorophenol and its salts and esters (UNEP/POPS/POPRC.9/13*. Retrieved from https://www.env.go.jp/council/05hoken/y051-155b/ref07.pdf

UNEP. (2017). *Technical guidelines on the environmentally sound management of wastes containing or contaminated with unintentionally produced polychlorinated dibenzo-p-dioxins, polychlorinated dibenzofurans, hexachlorobenzene, polychlorinated biphenyls, pentachlorobenzene or polychlorinated naphthalenes Revised final version (5 May 2017)*

Uotila, J. S., Salkinoja-Salonen, M. S. & Apajalahti, J. H. A. (1991). Dechlorination of pentachlorophenol by membrane bound enzymes of *Rhodococcus chlorophenolicus* PCP-I. *Biodegradation*, *2*(1), 25–31. https://doi.org/10.1007/BF00122422

Uotila, J. S., Kitunen, V. H., Saastamoinen, T., Coote, T., Haggblom, M. M. Salkinoja-Salonen, M. S. (1992). Characterization of aromatic dehalogenases of *Mycobacterium fortuitum* CG-2. *J Bacteriol*, *174*, 5669–5675.

US EPA. (2014). Pentachlorophenol. In: Integrated risk information system, http://www.epa.gov/ iris/subst/0086.htm.

US EPA. (2019). *Resource Conservation and Recovery Act (RCRA) Overview*. Retrieved from https://www.epa.gov/rcra/resource-conservation-and-recovery-act-rcra-overview

USEPA. (2008). *Reregistration eligibililty decision for pentachlorophenol. EPA 739-R-08-008.*

Valo, R. J., Apajalahti, J. & Salkinoja-Salonen, M. S. (1985). Studies on the physiology of microbial degradation of pentachlorophenol. *Appl Environ Microbiol*, *21*, 313–319.

Van Der Zande, A. (2010). *Exploration of management options for Pentachlorophenol (PCP) Paper for the 8th meeting of the UNECE CLRTAP Task Force on Persistent Organic Pollutants, Montreal, 18-20 May 2010 Colophon*. Retrieved from https://www.unece.org/fileadmin/DAM/env/documents/2013/air/PCP.pdf

Vandieken, V. & Thamdrup, B. (2013). Identification of acetate-oxidizing bacteria in a coastal marine surface sediment by RNA-stable isotope probing in anoxic slurries and intact cores. *FEMS Microbiology Ecology*, *84*, 373–386.

Van Hoof, P. L. & Rogers, J. E. (1992). Influence of low levels of nonionic surfactant on the anaerobic dechlrination of hexachlorobenzene. Biosystem technology development

program. *Bioremediation of Harzadous Wastes.*, p105-152.

Walter, M, Boyd-Wilson, K., Boul, L., Ford, C., McFadden, D., Chong, B. & Pinfold, J. (2005). Field-Scale bioremediation of pentachlorophenol by Trametes versicolor. *International Biodeterioration and Biodegradation*, *56*, 51-57.

Walter, M., Boul, L., Chong, R. & Ford, C. (2004). Growth substrate selection and biodegradation of PCP by New Zealand white-rot fungi. *Journal of Environmental Management*, *71*, 361–369.

Wang, N., Xue, X-M., Juhasz, A. L., Chang, Z-Z. & Li, H-B. (2017). Biochar increases arsenic release from an anaerobic paddy soil due to enhanced microbial reduction of iron and arsenic. *Environmental Pollution*, *220*(Pt A), 514–522. https://doi.org/10.1016/j.envpol.2016.09.095

Wang, S., Huang, L., Gan, L., Quan, X., Li, N., Chen, G., Lu, L., Xing, D. & Yang, F. (2012). Combined effects of enrichment procedure and non-fermentable or fermentable co-substrate on performance and bacterial community for pentachlorophenol degradation in microbial fuel cells. *Bioresource Technology*, *120*, 120–126. https://doi.org/10.1016/j.biortech.2012.06.022

Wang, W., Wang, S., Zhang, J., Hu, Z., Zhang, X. & Mu ~ Noz Sierra, J. (2016). Degradation kinetics of pentachlorophenol and changes in anaerobic microbial community with different dosing modes of co-substrate and zero-valent iron. *International Biodeterioration & Biodegradation*, *113*, 126–133. https://doi.org/10.1016/j.ibiod.2015.12.006

Watanabe, K., Manefield, M., Lee, M. & Kouzuma, A. (2009). Electron shuttles in biotechnology. *Current Opinion in Biotechnology*, *20*(6), 633–641. https://doi.org/10.1016/J.COPBIO.2009.09.006

WHO. (2003). Pentachlorophenol in drinking-water: background document for development of WHO Guidelines for Drinking-water Quality. WHO/ SDE/WSH/03.04/62. Geneva, Switzerland: World Health Organization. Available from: https://www. who.int/water_sanitation_health/dwq/.

Wong, F. & Bidleman, T. F. (2011). Ageing of organochlorine pesticides and polychlorinated biphenyls in muck soil, volatilization, bioaccessibility, and degra- dation. *Environment Science & Technology*, *45*, 958–963.

Xie, Y., Dong, H., Zeng, G., Tang, L., Jiang, Z., Zhang, C., Deng, J., Zhang, L. & Zhang, Y. (2017). The interactions between nanoscale zero-valent iron and microbes in the subsurface environment: A review. *Journal of Hazardous Materials*, *321*, 390–407.

https://doi.org/10.1016/j.jhazmat.2016.09.028

Xiu, Z-M., Jin, Z-H., Li, T-L. , Mahendra, S., Lowry, G. V. & Alvarez, P. J. J. (2010). Effects of nanoscale zero-valent iron particles on a mixed culture dechlorinating trichloroethylene,. *Bioresour. Technol.*, *101*, 1141–1146. https://doi.org/10.1016/j.biortech.2009.09.057.

Xu, Y., He, Y., Feng, X., Liang, L., Xu, J., Brookes, PC., Wu, J. (2014). Enhanced abiotic and biotic contributions to dechlorination of pentachlorophenol during Fe(III) reduction by an iron-reducing bacterium *Clostridium beijerinckii* Z. *Sci Total Environ.*, *473–474:*, 215–223.

Xu, Ling, Resing, K., Lawson, S. L., Babbitt, P. C. & Copley, S. D. (1999). Evidence that pcpA encodes 2,6-dichlorohydroquinone dioxygenase, the ring cleavage enzyme required for Pentachlorophenol Degradation in Sphingomonas chlorophenolica Strain ATCC 39723 †. https://doi.org/10.1021/bi990103y

Xu, X., Nembhard, W. N., Kan, H., Kearney, G., Zhang, Z-J. & Talbott, E. O. (2011). Urinary trichlorophenol levels and increased risk of attention deficit hyperactivity disorder among US school-aged children. *Occup Environ Med*, *68*, 557–661. https://doi.org/10.1136/oem.2010.063859

Xu, Y., He, Y., Egidi, E., Franks, A. E., Tang, C. & Xu, J. (2019). Pentachlorophenol alters the acetate-assimilating microbial community and redox cycling in anoxic soils. *Soil Biology and Biochemistry*, *131*, 133–140. https://doi.org/10.1016/j.soilbio.2019.01.008

Xun, L., Topp, E. & Orser, C. S. (1992). Glutathione is the reducing agent for the reductive dehalogenation of tetrachloro-p-hydroquinone by extracts from a Flavobacterium sp. *Biochemical and Biophysical Research Communications*, *182*(1), 361–366. https://doi.org/10.1016/S0006-291X(05)80153-6

Xun, L., Belchik, S. M., Xun, R., Huang, Y., Zhou, H., Sanchez, E., Kang, C. & Board, P. G. (2010). S-Glutathionyl-(chloro)hydroquinone reductases: a novel class of glutathione transferases. *The Biochemical Journal*, *428*(3), 419. https://doi.org/10.1042/BJ20091863

Xun, L., Bohuslavek, J. & Cai, M. (1999). Characterization of 2,6-Dichloro-p-hydroquinone 1,2-Dioxygenase (PcpA) of Sphingomonas chlorophenolica ATCC 39723. *Biochem. Biophys. Res. Commun.*, *266*, 322.

Xun, L., Topp, E. & Orser, C. S. (1992a). Diverse substrate range of a Flavobacterium pentachlorophenol hydroxylase and reaction stoichiometries. *Journal of Bacteriology*, *174*(9), 2898–2902. https://doi.org/10.1128/jb.174.9.2898-2902.1992

Xun, L., Topp, E. & Orser, C. S. (1992b). *Purification and characterization of a tetrachloro-*

p-hydroquinone reductive dehalogenase from a Flavobacterium sp. Journal of Bacteriology (Vol. 174). https://doi.org/10.1128/jb.174.24.8003-8007.1992

Yadid, I., Rudolph, J., Hlouchova, K. & Copley, S. D. (2013). Sequestration of a highly reactive intermediate in an evolving pathway for degradation of pentachlorophenol. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(24), E2182-90. https://doi.org/10.1073/pnas.1214052110

Yahaya, A., Okoh, O. O., Agunbiade, F. O. & Okoh, A. I. (2019). Occurrence of phenolic derivatives in Buffalo River of Eastern Cape South Africa: Exposure risk evaluation. *Ecotoxicology and Environmental Safety*, *171*(November 2018), 887–893. https://doi.org/10.1016/j.ecoenv.2019.01.037

Yang, CF., Lee, CM. & Wang, C. (2006). Isolation and physiological characterization of the pentachlorophenol degrading bacterium Sphingomonas chlorophenolica. *Chemosphere*, *62*(5), 709–714. https://doi.org/10.1016/j.chemosphere.2005.05.012

Yeh, D. H., Pennell, K. D. & Pavlostathis, S. G. (1998). Toxicity and biodegradability screening of nonionic surfactants using sediment-derived methanogenic consortia. *Water Science and Technology*, *38*(7), 55–62.

Yin, Y. & Zhou, N. Y. (2010). Characterization of MnpC, a Hydroquinone Dioxygenase Likely Involved in the meta-Nitrophenol Degradation by Cupriavidus necator JMP134. *Curr. Microbiol.*, *61*, 471. https://doi.org/10.1007/s00284-010-9640-3

Yu, H., Wang, Y., Chen, PC., Li, FB., Chen, MJ., Hu, M., Ouyang, X. (2014). Effect of nitrate addition on reductive transformation of pentachlorophenol in paddy soil in relation to iron (III) reduction. *J Environ Manag.*, *132*, 42–48.

Yu, L., Yuan, Y., Tang, J., Wang, Y. & Zhou, S. (2015). Biochar as an electron shuttle for reductive dechlorination of pentachlorophenol by Geobacter sulfurreducens. *Scientific Reports*, *5*(1), 16221. https://doi.org/10.1038/srep16221

Zhang, B. G., Li, G. T., Shen, T.S., Wang, J. K. & Sun, Z. (2000). Changes in microbial biomass C, N, and P and enzyme activities in soil incubated with the earthworms Metaphire guillelmi or *Eisenia fetida*. *Soil Biology Biochemistry*, *32*, 2055–2062.

Zhang, C., Zhang, N., Xiao, Z., Li, Z. & Zhang, D. (2019). Characterization of biochars derived from different materials and their effects on microbial dechlorination of pentachlorophenol in a consortium. *RSC Advances*, *9*(2), 917–923. https://doi.org/10.1039/C8RA09410A

Zhang, D., Zhang, C., Li, Z., Suzuki, D., Komatsu, D. D., Tsunogai, U. & Katayama, A. (2014). Electrochemical stimulation of microbial reductive dechlorination of

pentachlorophenol using solid-state redox mediator (humin) immobilization. *Bioresource Technology*, *164*, 232–240. https://doi.org/10.1016/j.biortech.2014.04.071

Zhang, J., Liu, X., Xu, Z., Chen, H. & Yang, Y. (2008). Degradation of chlorophenols catalyzed by laccase. *International Biodeterioration & Biodegradation*, *61*(4), 351–356. https://doi.org/10.1016/J.IBIOD.2007.06.015

Zheng, W., Yu, H., Wang, X. & Qu, W. (2012). Systematic review of pentachlorophenol occurrence in the environment and in humans in China: not a negligible health risk due to the re-emergence of schistosomiasis. *Environ Int*, *42*, 105–16. https://doi.org/10.1016/ j.envint. 2011.04.014

Zhou, G-W., Yang, X-R., Li, H., Marshall, C. W., Zheng, B-X., Yan, Y., Su, J-Q. & Zhu, Y-G. (2016). Electron Shuttles Enhance Anaerobic Ammonium Oxidation Coupled to Iron(III) Reduction. *Environmental Science & Technology*, *50*(17), 9298–9307. https://doi.org/10.1021/acs.est.6b02077

Zhu, M., Zhang, L., Zheng, L., Zhuo, Y., Xu, J. & He, Y. (2018). Typical Soil Redox Processes in Pentachlorophenol Polluted Soil Following Biochar Addition. *Frontiers in Microbiology*, *9*, 579. https://doi.org/10.3389/fmicb.2018.00579

CHAPTER THREE

This chapter has been ***published in the Journal: Biodegradation, Springer Nature***

Biotransformation of pentachlorophenol by an indigenous Bacillus cereus *AOA-CPS1 isolated from wastewater effluent in Durban, South Africa*

Oladipupo A. Aregbesola, Mduduzi P. Mokoena & Ademola O. Olaniran

Biodegradation

ISSN 0923-9820

Biodegradation
DOI 10.1007/s10532-020-09915-w

ONLINE FIRST

Springer

Biodegradation
https://doi.org/10.1007/s10532-020-09915-w

ORIGINAL PAPER

Biotransformation of pentachlorophenol by an indigenous *Bacillus cereus* AOA-CPS1 isolated from wastewater effluent in Durban, South Africa

Oladipupo A. Aregbesola · Mduduzi P. Mokoena · Ademola O. Olaniran

Received: 3 March 2020 / Accepted: 23 September 2020

Abstract Pentachlorophenol (PCP) is a recalcitrant biocide that bioaccumulates in the environment due to its persistent nature and has been listed as a priority pollutant due to its toxicological and health effects. In this study, a novel PCP-degrading *Bacillus cereus* strain AOA-CPS1 (*Bc*AOA) was isolated from wastewater and characterized for PCP biotransformation in a batch reactor. The degradation kinetics were elucidated via substrate inhibition models, while PCP biotransformation was established by spectrophotometric and GC–MS analysis. *Bc*AOA shared 95% sequence homology with *Bacillus cereus* strain XS2 and is closely related to some *B. cereus* strains which are previously reported to degrade PCP and other related pollutants. BcAOA degraded 74% of 350 mg l^{-1} of PCP within 9 days in a batch culture. The biotransformation of PCP by *Bc*AOA followed the first and zero-order kinetics at low and high PCP concentration, respectively, with biokinetic constants: maximum biotransformation rate (0.0996 mg l^{-1} - h^{-1}); substrate inhibition constant (723.75 mg l^{-1}); half-saturation constant (171.198 mg l^{-1}) and R^2 (0.98). The genes (pcp*ABCDE*, cytochrome P450) encoding the enzymes involved in the biodegradation of PCP were amplified from the genomic DNA of *Bc*AOA. Further, depending upon the genes amplified and identified metabolites using GC–MS, two different PCP biotransformation pathways were proposed in this study. Cloning and expression of the catabolic genes are underway to map out the concise pathway for PCP biotransformation by *Bc*AOA.

Keywords Pentachlorophenol · *Bacillus cereus* AOA-CPS1 · PCP-degrading genes (pcpABCDE) · Cytochrome p450

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10532-020-09915-w) contains supplementary material, which is available to authorized users.

O. A. Aregbesola · M. P. Mokoena · A. O. Olaniran (✉)
Discipline of Microbiology, School of Life Sciences, College of Agriculture, Engineering and Science, University of KwaZulu-Natal (Westville Campus), Private Bag X54001, Durban 4000, Republic of South Africa
e-mail: olanirana@ukzn.ac.za

Introduction

Pentachlorophenol (PCP) is one of the World's worst chemical ever produced (IPEN POPRC 2013). It is a synthetic organochlorine pesticide, a conjugate acid of pentachlorophenolate and a member of pentachlorobenzenes, which comprises aromatic fungicides and a chlorophenol (CPs) (Kim et al. 2019). PCP is also a metabolite of lindane and other polychlorinated phenolic compounds (Engst et al. 1976). PCP has previously been used as fungicides, pesticides,

Published online: 03 October 2020

defoliant, and wood preservative (Lopez-Echartea et al. 2016), its usage has been discontinued in many countries due to its toxicological profile (Igbinosa et al. 2013). However, PCP is still in used industrially as a wood preservative for railroad ties, wharf pilings and utility poles, but, the sale and use of PCP have been restricted to certified applicators (Kim et al. 2019).

PCP has global health implications since it is found in the bodies of people throughout the world including indigenous people of the arctic (IPEN POPRC 2013). PCP has been listed as a persistent organic pollutant (ATSDR 2017; CDC 2018), and classified as carcinogenic (Group 1) to humans based on sufficient epidemiological data (IARC 2016). Furthermore, it has recently been recognized (Stockholm convention 2019), and published as a group 1 carcinogenic hazard to humans (IARC 2019a, 2019b).

The widespread use of PCP has led to contamination of the environment; and in many wood-processing industries, the surrounding soil is heavily contaminated, impacting surrounding groundwaters, rivers, lake and sea waters (McAllister et al. 1996; Villemur 2013). The presence of this compound in wood and wood materials at the end of their life cycle limits their recycling prospects (Lin et al. 2009; Patachia and Croitoru 2016). Also, Leaching of the compound from chemically preserved wood and wood products constitutes significant environmental contamination (Patachia and Croitoru 2016). Due to its recalcitrant nature and potential for long-range dispersal, PCP and its congeners are still found in areas where use had been discontinued for decades (Gong et al. 2007; Quinn et al. 2011). These residues may pose chronic toxicity to animals and humans via air, water, and food intake (Darko et al. 2008).

Many indigenous microorganisms cannot utilize PCP as their sole carbon source because of its high toxicity and inhibitory effects. Consequently, the compound bioaccumulates in the environment (Patachia and Croitoru 2016). Several conventional physical and chemical techniques (Zhang et al. 2008; Ren et al. 2016) have been used to transform PCP. Although these techniques are very fast, they are very expensive and not environmentally friendly due to the formation of more toxic intermediates, requiring further processing for complete mineralization (Olaniran and Igbinosa 2011; Patel and Kumar 2016a, b, c). On the contrary, bioremediation offers an effective and eco-friendly method of removing chlorophenols from the environment (Olaniran and Igbinosa 2011).

Several bacteria such as *Sphingobium chlorophenolicum* L-1 (Takeuchi et al. 2001), *Shingomonas* sp UG30 (Cassidy et al. 1999), *Acinetobacter* sp. ISTPCP-3 (Sharma et al. 2009), *Pseudomonas stutzeri* CL7 and *Enterobacter* sp. SG1 (Karn et al., 2010), and *Burkholderia cepacia* (Joshi et al. 2015) have been reported to degrade PCP. Of all these organisms, *S. chlorophenolicum* L-1 is the most widely studied strain. Catabolic genes (*pcpABCDE*) encoding PCP degrading enzymes (pcpABCDE) have been reported, cloned and overexpressed (Chanama and Chanama 2011; Dai et al. 2003).

Biodegradation of PCP occurs via hydroxylation, oxygenolysis and/or reductive dehalogenation (McAllister et al. 1996). In *S. chlorophenolicum* L-1, PCP is first converted to tetrachloro-1,4-benzoquinone (TeCBQ) via hydroxylation at the para-position involving PCP-4-monooxygenase (pcpB) (Orser et al. 1993a, b), and then TeCBQ is reduced to tetrachloro-1–4-hydroquinone (TeCHQ) by TeCBQ-reductase (pcpD) (Cai and Xun 2002; Chen and Yang 2008). Glutathione-*S*-transferase (pcpC), a TeCHQ-reductive dechlorinase sequentially transforms TeCHQ to 2,6-Dichloro-p-hydroquinone (Di-CHQ) (Chanama and Chanana 2011; Orser et al. 1993a, b). Di-CHQ is then cleaved to 2-CMA by Di-CHQ 1,2-dioxygenase (pcpA) (Chanama and Crawford 1997; Lopez-Echartea et al. 2016; Xu et al. 1999). Furether, Maleylacetate reductase (pcpE) transforms 2-CMA into succinyl CoA and acetyl-CoA, before entering into the Kreb's cycle (Chen et al. 2009; Joshi et al. 2015).

The previous reports on the incidences of organochlorine pesticides in Africa soil (Thompson et al. 2017) and the unmasking of a hidden reality of pesticide residues in European agricultural soils (Silva et al. 2019), indicated that our environment is not free of these persistent organic pollutants. South Africa is one of the four largest importers of pesticides in sub-Saharan Africa (Osibanjo et al. 2003). More than 500 pesticides were registered for use in South Africa (Quinn et al. 2011), and organochlorine pesticides are one of the major pesticides in use in South Africa (DOH 2005; Quinn et al. 2011). These pesticides are used in almost every aspect of our everyday lives; to ensure the quantity and quality of food we eat and to control insects and pests in our homes (Quinn et al.

2011). According to the South African Government Gazette 2019, all chlorophenols are considered potentially harmful to aquaculture and human health, the recommended maximum limit of PCP in the natural environment is 0.4 μg l^{-1} and 0.1 μg l^{-1} for mariculture.

According to Water Research Commission (WRC), South Africa is currently under threat of a lack of sufficient water, while water quality and availability issues are becoming more acute and worrisome (WRC 2018). About seventeen organochlorine pesticides and/or metabolites were detected in sediment collected in rivers, estuaries and canals in the eThekwini area of kwaZulu-Natal, at concentrations exceeding the method detection limit (WRC 2015). The recent detection of: chlorophenol congeners in Buffalo River of Eastern Cape, South Africa (Yahaya et al. 2019), uMgeni river, kwaZulu-Natal (Gakuba et al. 2018), polychlorinated biphenyls (PCBs) congeners and dichlorodiphenyltrichloroethane (DDT) in fresh root and leafy vegetables (Olatunji 2019), PCP congeners and Polycyclic aromatic hydrocarbons (PAHs) in Nandoni Dam in Limpopo Province of South Africa (Nthunya et al. 2019), showed a wide distribution of organochlorine pesticides contamination in South African environment. There is an urgent need for an environmental friendly strategies such as bioremediation, to ameliorate the contaminated environment. Therefore, the present study reports a novel and efficient PCP-degrading bacterium *Bacillus cereus* strain AOA-CPS1 from a wastewater effluent in South Africa.

Materials and methods

Chemicals and reagents

Pure PCP (98%) and its congeners, *N*-methyl-*N*-trimethylsilyl-trifluoroacetamide (TMS), and other chemicals used for this study were purchased from Merck (Merck & Company, Inc., USA) and are of analytical grade standards. The stock solution of PCP and its congeners were prepared as their sodium salt by dissolving them in 0.05 N sodium hydroxide while 2,4,5-trichlorophenol (2,4,5-TCP) was dissolved in ethyl acetate. PCR reaction master mix was purchased from ThermoFisher Scientific (Waltham, MA, USA).

Sample collection and culture enrichment

The effluent of wastewater was collected from wastewater treatment plant in Durban, South Africa, and transported to the laboratory for use as inoculum for cultural enrichment. The samples were enriched using a minimal salt medium (MSM) prepared as previously described (Saber and Crawford 1985). The enrichment culture in 45 ml of MSM supplemented with 50 mg l^{-1} of PCP and 5 ml of sludge. The flasks were incubated for 7 days, after which 5.0 ml of the enriched culture was transferred into 45.0 ml of fresh 50 mg l^{-1} PCP-supplemented MSM and incubated for another 7 days. Unless otherwise stated, all biotransformation experiments were set up in 250 ml Erlenmeyer flasks, and all biotransformation assays were incubated at 30 °C and 150 rpm in a shaking incubator (Innova 44 series, New Brunswick Scientific, UK). All biotransformation experiments were conducted in triplicates, and the results presented were the means and standard deviations of triplicate experiments. All spectrophotometric readings were taken using UV–Vis spectrophotometer (Cary 60 UV–Vis, Agilent Technologies, USA) and centrifugation was done with either Avanti J-26 XPI centrifuge (Beckman Coulter, USA) or Eppendorf centrifuge 5415D (Hamburg, Germany).

Isolation and screening for PCP degrading bacteria

After three successive sub-culturing, 0.1 ml of the enriched culture was spread inoculated on minimal salt agar (MSM plus 15 g l^{-1} of bacteriological agar) plates supplemented with 50 mg l^{-1} of PCP. The plates were incubated at 30 °C until visible growths were observed. The isolates were purified via successive sub-culturing on sterile nutrient agar (NA) plates until distinct colonies were obtained. Glycerol stocks of the pure isolates were prepared and stored in the bio-freezer at − 80 °C. The isolated bacteria were individually screened in PCP-MSM, and strains with potential for PCP biotransformation were selected for further studies.

Biotransformation study

Biotransformation of PCP by the selected isolate was performed using a low buffered MSM (Saber and Crawford 1985), with some modifications. Four types

of growth media {MSM only (MSM), MSM + glucose (GMSM), MSM + glucose + PCP (GPMSM) and MSM + PCP (PMSM)} were re-composed. The reconstituted MSM only contained (g l^{-1}): K_2HPO_4 (0.065); KH_2PO_4 0.019; $MgSO_4.7H_2O$ (0.1); $NaNO_3$ (0.5); 0.02 M $FeSO_4$ (2 ml); pH 7.0, and 2 ml of micronutrients. The micronutrients contained (mg l^{-1}): $ZnSO_4{\cdot}7H_2O$ (4.0); $MnSO_4{\cdot}4H_2O$ (0.2); H_3BO_3 (0.15) and EDTA (2.5) (Ammeri et al. 2017). The GMSM contained glucose (0.05%); Luria Bertani broth (0.006%) and MSM only; GPMSM contained 100 mg l^{-1} of PCP, glucose (0.05%); Luria Bertani broth (0.006%) and MSM only while PMSM contained 100 mg l^{-1} of PCP and MSM only.

The isolate was grown in nutrient broth overnight, harvested by centrifugation (8000 rpm for 10 min), washed twice with MSM only, resuspended in the same medium and standardized (OD_{600nm} of 1.0). About 90 ml each of GPMSM and PMSM media were inoculated with 10 ml of standardized inoculum and incubated for 9 days. Positive (10 ml of inoculum + 90 ml of GMSM medium), and negative (PMSM medium without the inoculum) controls were assayed along with the experiment to check for growth and abiogenic loss of PCP during biotransformation processes (Patel and Kumar 2016a, b, c). Cell growth and PCP biotransformation were monitored spectrophotometrically everyday. Cells growth was measured at 600 nm while the disappearance of PCP and formation of metabolites were detected by scanning the sample between 350 to 200 nm using a 10 mm quartz cuvette (Hellma Analytics, Germany).

Effects of different initial PCP concentrations on bacterial growth and PCP biotransformation

Effects of different initial PCP concentrations on the isolate was assessed by growing the bacterium in GMSM medium in the presence of different initial concentrations (100, 150, 200, 250, 300, and 350 mg l^{-1}) of PCP. The reaction mixtures were incubated as previously described, and growth and PCP biotransformation were monitored every 24 h (Ammeri et al. 2017).

Kinetic studies

Kinetics studies on the isolate were evaluated in batch experiments. The specific growth rate (μ) of the organism at each initial PCP concentration tested was extrapolated from the exponential growth phase (Eq. 1) as previously described (Monod 1949). Since PCP is an inhibitor for cell growth, biodegradation kinetic constants were calculated using Haldane/Andrew's substrate inhibition model (Andrews 1968; Kargi and Eker 2005) described in Rq. 2. At low PCP concentration ($\leq$ 250 mg l^{-1}), inhibition constant was ignored, Eq. 2 becomes Eq. 3, and in a linear form as Eq. (4).

$$\mu = (\mathrm{Log}_2 X_2 - \log_2 X_1) / (t_2 - t_1) \quad (1)$$

$$\begin{aligned} R_s &= (R_m S/(K_s + S))\,(K_{si}/(K_{si} + S)) \\ &= \{R_m/(1 + (K_s/S))\,(1 + (S/K_{si}))\} \end{aligned} \quad (2)$$

$$R_s = R_m S/(K_s + S) \quad (3)$$

$$1/R_s = (1/R_m) + (K_s/R_m)\,(1/S) \quad (4)$$

where μ is the specific growth rate, X_2 and X_1 were the cell growth at time t_2 and t_1, respective, R_s is the specific degradation rate, R_m is the maximum biodegradation rate, S is the substrate concentration, K_s is the half-saturation constant (mg l^{-1}) and K_{si} is the substrate inhibition constant (mg l^{-1}).

Metabolites detection via GC–MS analysis

The isolate was grown in MSM supplemented with 50 mg l^{-1} of PCP and incubated for 48 h. Metabolites were extracted and derivatized as previously described (Li et al. 2001; Smith 2003). Both derivatized and underivatized samples were analysed on the Agilent 7890A GC System (Agilent Technologies, USA), equipped with a 5975C MS detector. The system was run at 80 °C, raised at 5 °C min^{-1} to 160 °C for 3 min, then raised at 10 °C min^{-1} to 260 °C and held at 260 °C for another 3 min. Mass ranges (m/z) were set at 50–700, ionization energy was set at 70 eV and injection volume of 1 µl. The peaks of the PCP and its metabolites were compared to the National Institute of Standard and Technology (NIST) library database (Sharma et al. 2009).

Identification of the isolate based on 16S rDNA sequence analysis and phylogenetic typing

Genomic DNA of the isolate was extracted from an overnight grown broth culture using a Quick-DNA™

fungal/bacterial miniprep kit (Zymo Research Corporation, USA). Purity and concentration of the extracted DNA were checked (NanoDrop 2000c Spectrophotometer, ThermoScientific). The integrity of the DNA samples was checked by running 200 ng of the sample on 1.0% agarose gel at 80 V for 80 min. The sample was kept in the fridge (− 20 °C) for further analysis. The 16S rDNA fragment of the extracted DNA was PCR amplified in a thermal cycler (T100™, Bio-Rad, USA), using 63F and 1387R universal primers pair (Table S1), using the previously reported amplification conditions (Marchesi et al. 1998). The amplicon was visualized in a 1.0% agarose gel, stained in ethidium bromide (1%) and the image was captured (Syngene G: BOX, gel documentation system, UK). The amplified 16S rDNA gene was sequenced (Inqaba Biotechnical Company, South Africa) and submitted at NCBI GenBank (https://blast.ncbi.nlm.nih.gov/blast/) database.

A phylogenetic tree was constructed via the Neighbor-Joining method (Zhang and Sun 2008). Bootstrap consensus tree from 1000 replicates was used to denote the evolutionary lineage of the taxa analysed (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates were collapsed. Evolutionary distances were computed using the Maximum Composite Likelihood method (Tamura et al. 2004), and were in the units of the number of base substitutions per site. The analysis involved 21 nucleotide sequences, and 1st, 2nd, 3rd codon positions and non-coding region were included. All positions containing missing data and gaps were removed, and a total of 1008 positions were used in the final dataset. Evolutionary relatedness was conducted using a molecular evolutionary genetics analysis software, version 7.0 (Kumar et al. 2016).

Detection of catabolic genes

Complete genomes of closely related strains based on the 16S rDNA homology: *Bacillus thuringiensis* strain L-7601 (CP020002.1), *B. thuringiensis* strain ATCC 10,792 (CP021061.1) and *B. cereus* strain 25 (CP020803.1), were retrieved from the GenBank. Catabolic genes (*pcpABCDE*) encoding PCP degrading enzymes were searched, sequences of each gene found in the genomes of the selected organisms were aligned (DNAMAN-Lynnon Biosoft, USA), and consensus sequences used to design the primers pairs (Table S1), used for PCR amplification of the catabolic genes in the isolate used in this study. The 10 µL reaction mix contained: 1 µL of 10 × buffer, 1.5 mM of $MgCl_2$, 20 µM (each) dNTPs, 1.0 µM of each primer, 1.25 U of *Taq* polymerase, ~20 ng of the template DNA and ddH_2O. The PCR was run at 95 °C for 5 min, followed by 25 cycles at 95 °C for 30 s, annealing at 54 °C (for *pcpA*, *pcpD* and *pcpE*) or 56 °C (for *pcpB* and *pcpC*) for 1 min, 72 °C for 1 min and a final extension at 72 °C for 10 min. The amplicons were analysed in a 1.0% agarose gel at 80 V for 120 min and visualized in a gel documentation system (Syngene G: BOX, gel documentation system, UK).

Results

Identification and characterization of the PCP-degrading bacterial isolate

Blast searches of the 16S rRNA gene sequence of the isolate showed that the isolate belongs to the *B. cereus* group. The strain shared 95% homology with *B. cereus* XS2 (JX448404.1) and 94% with *B. cereus* M7 (MG977683.1), *B. cereus* RD6 (MH114968.1), *B. cereus* F16 (MF135182.1), *B. thuringiensis* FDBS2 (MG827297.1) and *B. anthracis* DOS-ERY-1 respectively. The isolate is designated as *Bacillus cereus* strain AOA-CPS1 (*Bc*AOA). The 16S rDNA partial gene sequence of the isolate has been deposited in the GenBank with an accession number MH504118.1. Phylogenetic analysis of the 16S rRNA gene partial sequence of the isolate (Fig. 1) showed that *Bc*AOA is closely related to *B. cereus* F16 (MF135182.1), an endophyte associated with rice under aluminium toxicity and *B. cereus* strain M7 (MG977683.1), used for bioremediation of winery effluents. *Bc*AOA is also found to be related to *B. cereus* ITRC-S6 (DQ002384), a PCP-degrading strain isolated from pulp and paper mill effluent (Chandra et al. 2009); and *B. cereus* XS2 (JX448404.1), and arsenite oxidizing bacterium from gold mine tailing area of Xinjiang, China (Karn and Pan 2017).

Biotransformation of PCP by BcAOA

Time course growth of *Bc*AOA and PCP biotransformation is shown in Fig. 2a, b, respectively. The

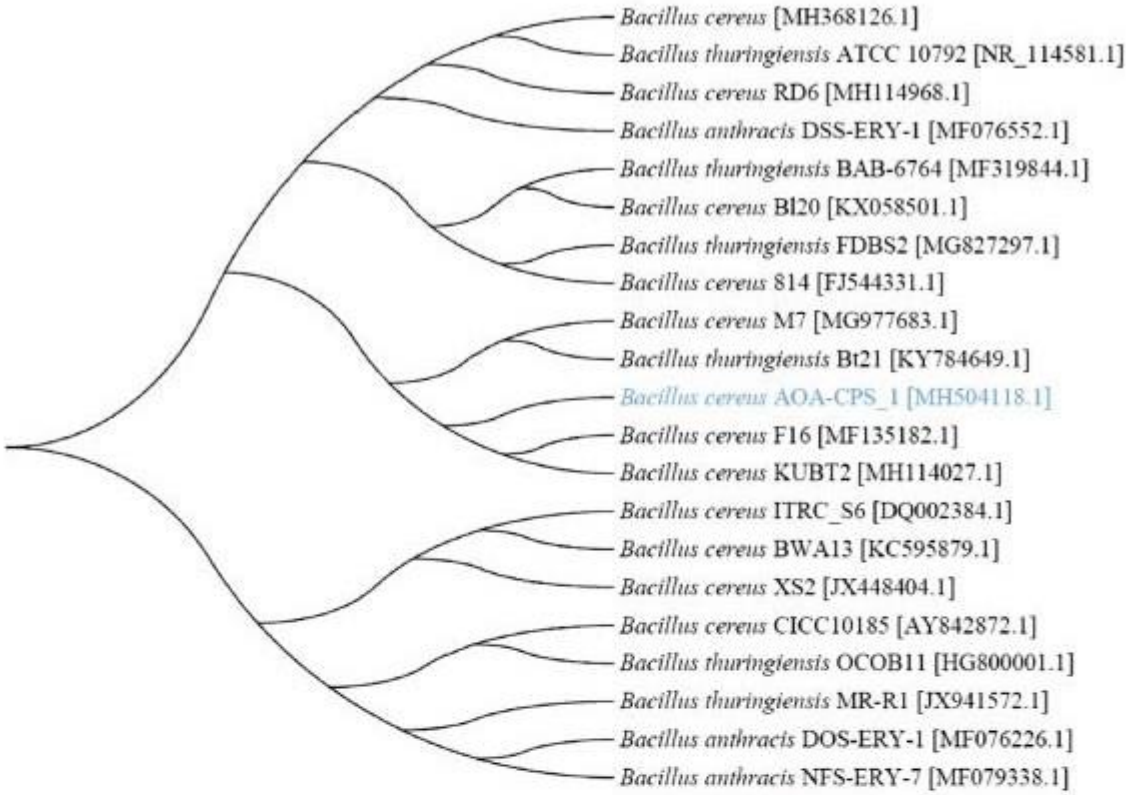


Fig. 1 Phylogenetic analysis of the 16S rRNA gene partial sequence of *B. cereus* AOA-CPS

opimum growth (OD_{600nm} of 1.1782) of *Bc*AOA was observed within 6 days in GMSM but the growth declined marginally during further incubation, whereas maximum growth densities of 0.9616 and 0.2335 (at OD_{600nm}) were obtained at day 6 and day 9 in GPMSM and PMSM, respectively (Fig. 2a). The organism transformed 57.783 mg l^{-1} ($\approx$ 57%) of 100 mg l^{-1} of PCP in GPMSM medium within 5 days, further incubation did not yield a significant increase in PCP biotransformation. *Bc*AOA also transformed 28.893 mg l^{-1} ($\approx$ 29%) of 100 mg l^{-1} of PCP in PMSM medium with inoculum within 9 days. However, no significant reduction in PCP concentration was recorded in a PMSM medium without the inoculum after 9 days of incubation (Fig. 2b).

Apart from PCP, *Bc*AOA also transformed 2,4,6-TCP, 2,4-dichlorophenol (2,4-DCP), 4-chlorophenol (4-CP) and 2-chlorophenol (42-CP). The strain also co-metabolized different concentrations of PCP and 2,4,6-TCP mixture (Table 1). Effects of different initial PCP concentrations on the growth of *Bc*AOA and PCP biotransformation is shown in Fig. 2c, d, respectively. The data shows that the growth of *Bc*AOA decreases (Fig. 2c) while PCP biotransformation increases (Fig. 2d) at increased initial PCP concentrations. The inhibitory effects of PCP on the growth of *Bc*AOA resulted in a lag phase that lasted between 24 and 72 h. It is evident that *Bc*AOA could transform $\approx$55% when inoculated with 100 mg l^{-1} of PCP while $\approx$75% when inoculated with 350 mg l^{-1} of PCP when compared at day 9.

Kinetic studies

The specific growth of *Bc*AOA and PCP biotransformation rates are shown in Fig. 3a. Due to the inhibitory effects of PCP on the growth of *Bc*AOA, the specific growth rate decreased with increase in PCP concentration. However, the specific PCP

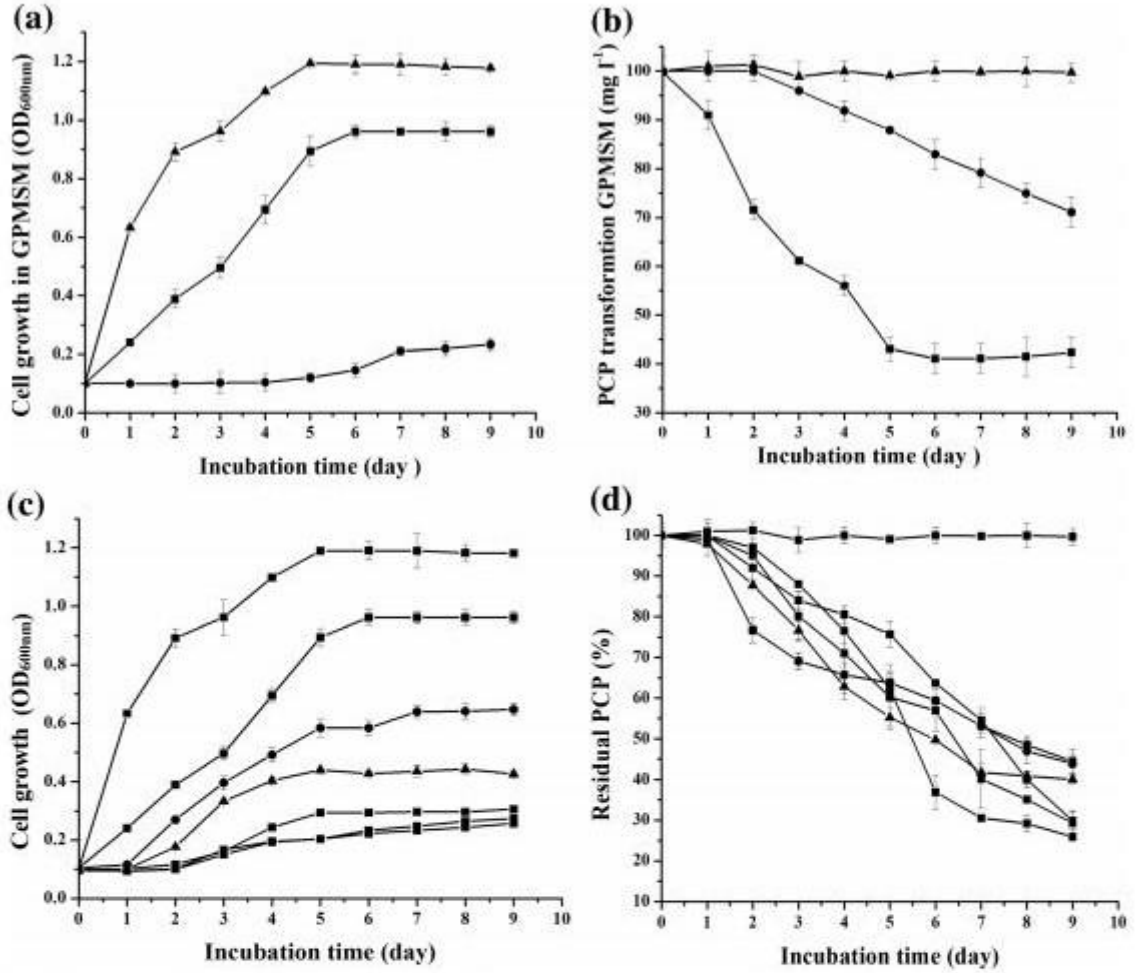


Fig. 2 Growth and PCP biotransformation profile of *B. cereus* AOA-CPS1. **a** Cell growth in GMSM (filled triangle), GPMSM (filled triangle) and PMSM (filled circle) in the presence of 100 mg l^{-1} of PCP; **b** residual of 100 mg l^{-1} of PCP in PMSM medium without inoculum (filled triangle), GPMSM medium (filled square) and PMSM medium with inoculum (filled circle); **c** Growth in different initial PCP concentrations 100 (filled square), 150 (filled circle), 200 (filled triangle), 250 (filled inverted triangle), 300 (filled left triangle), 350 (filled right triangle) mg l^{-1} and GMSM medium (filled diamond) on growth and **d** residual PCP (%) of initial PCP concentrations 100 (filled square), 150 (filled circle), 200 (filled triangle), 250 (filled inverted triangle), 300 (filled left triangle), 350 (filled right triangle) mg l^{-1} and PMSM without the inoculum (filled diamond) over time

biotransformation rate increased with increase in PCP concentrations up to 250 mg l^{-1}, remains at equilibrium/saturated between 250 and 300 mg l^{-1}, and then increased at 350 mg l^{-1} after which the removal rates decreased. A reciprocal plot, between PCP removal rate ($1/R_s$), and substrate concentration (1/S), at low PCP conc ($\leq$ 250 mg l^{-1}), yielded a linear graph with a slope of K_s/R_m and an intercept of $1/R_m$ (Fig. 3b). From a linear line of the best fit plot, R_m (0.0996 mg PCP l^{-1} h^{-1}); K_s (171.198 mg l^{-1}), and R^2 (0.98) were obtained.

PCP inhibits *Bc*AOA growth at a particular concentration. The concentration at which degradation rate dropped is the critical PCP concentration (S*) for the isolate. Critical substrate concentration was obtained from the derivation of Eq. 4 as S approaches zero (Eq. 5). From the degradation rate curve, the maximum PCP concentration S* at which optimum

Table 1 Growth of *Bacillus cereus* strain AOA-CPS1 in different concentrations of chlorophenol congeners

CPS congener (mg l^{-1})	50	100	150	200	250	300	350
PCP	+	+	+	+	+	+	+
2,4,6-TCP	+	+	+	+	+	+	+
2,4,5-TCP	−	−	−	−	−	−	−
2,6-DCP	−	−	−	nt	nt	nt	nt
2,4-DCP	+	+	+	nt	nt	nt	nt
2,4-D	−	−	−	nt	nt	nt	nt
4-CP	+	+	+	nt	nt	nt	nt
3-CP	−	−	−	nt	nt	nt	nt
2-CP	+	+	+	nt	nt	nt	nt
PCP + 2,4,6-TCP	+	+	+	+	+	+	nt

CPS chlorophenol; *PCP* pentachlorophenol; *2,4,6-TCP* 2,4,6-trichlorophenol; *2,4,5-TCP* 2,4,5-trichlorophenol; *2,6-DCP* 2,6-dichlorophenol; *2,4-DCP* 2,4-dichlorophenol; *2,6-D* 2,6-dichlorophenoxyacetic acid; *4-CP* 4-dichlorophenol; *3-CP* 3-dichlorophenol; *2-CP* 2-dichlorophenol; nt, not tested

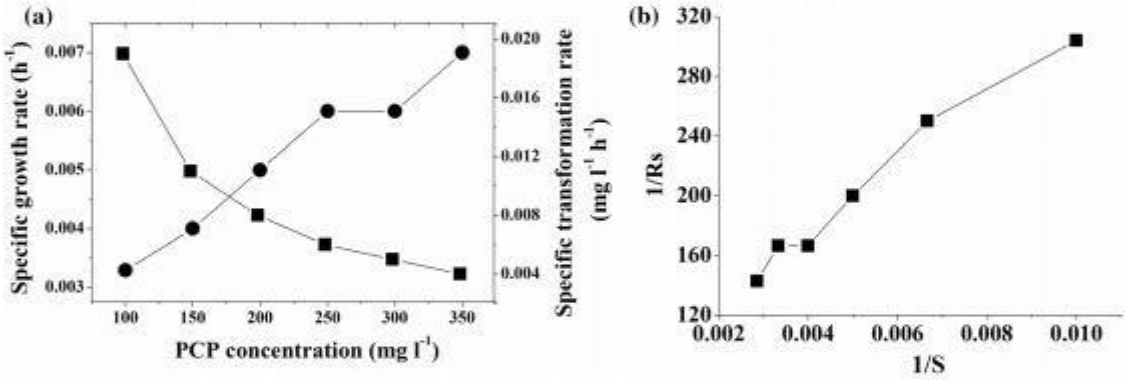


Fig. 3 Kinetic studies of growth and PCP biotransformation by *B. cereus* AOA-CPS1. **a** Specific growth rate (filled circle) and PCP biotransformation rate constants (filled square) at different initial PCP concetration of 100–350 mg l^{-1}; **b** reciprocal plot of biotransformation rate ($1/R_s$) vs PCP concentration (1/S)

PCP biodegradation rate was obtained for *Bc*AOA is 250 mg l^{-1}. Substituting this value and the value of K_s (171.19 mg l^{-1}) into Eq. 6, K_{si} value of 723.75 mg l^{-1} was obtained. The biodegradation kinetic parameters obtained were: R_m (0.996 mg l^{-1} - h^{-1}); K_s (171.19 mg l^{-1}; K_{si} (723.75 mg l^{-1}); and R^2 (0.98). Therefore, PCP degradation kinetic models for the strain can be written as Eq. 7.

$$dRs/dS = 0 \quad (5)$$

$$S^* = \sqrt{K_s K_{si}} \quad (6)$$

$$R_s = \{0.0996/(1 + (171.19/S))(1 + (S/723.75))\} \quad (7)$$

Profiling of PCP-degrading genes in BcAOA and PCP degradation metabolites

Catabolic genes involved in the PCP biotransformation in *Bc*AOA were amplified from the purified genomic DNA. In addition to all the catabolic genes (*pcpABCDE*) encoding PCP degrading enzymes (pcpABCDE) detected and shown in Fig. 4, the cytochrome P450 monooxygenase gene was also detected and amplified. The GC–MS analysis showed

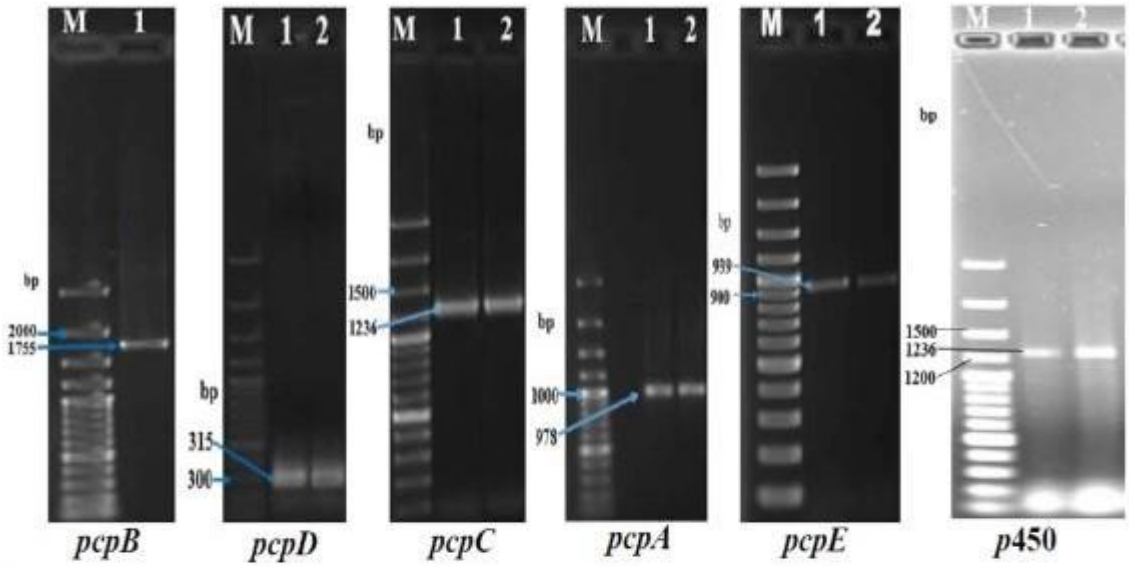


Fig. 4 PCP amplification of catabolic genes involved in the degradation of PCP in *B. cereus* AOA-CPS1

a major peak with a retention time (RT) of 19.795 correspondings to the standard PCP. Appearances of new peaks in GC–MS data led to the identification of some of the PCP degradation metabolites especially the ring cleavage products (Fig. 5), proved that the PCP was degraded by the organism. The retention time(s), mass spectra and relative intensity of both derivatized (TMS) and underivatized products are shown in Table S2 while the spectra for all the identified metabolites are shown in supplementary material (Fig. S1).

Proposed pathway for PCP biotransformation in BcAOA

Based on the catabolic genes and metabolites detected by PCR and GC–MS analysis, respectively, two pathways (Fig. 6) are proposed for PCP biotransformation in *Bc*AOA in this study. In the first pathway, 2,6-bis(1,1-dimethylethyl)phenol (RT:15.310); 1-methoxy-5-trimetylsilyoxyhexane, trimethylsilyl 2-butoxyacetate (RT: 10.225), and Methyl 2-hydroxyl-3-methylbutanoate (RT: 10.225) fragments were found. The metabolites detected for the proposed second pathway include 1,3-dimethyl-4,6-diisopropylbenzene; 2,4-Dimethylbenzenecarboxaldehyde and 2,5-Dimethylbenzaldehyde.

Discussion

Recovery of a chlorophenol degrading bacterium from the wastewater effluent is not surprising based on the history of organochlorine pesticides use in the Republic of South Africa (DOH 2005; Quinn et al. 2011). It is probable that the isolate (*Bc*AOA) acquired the capacity to degrade PCP, through adaptation due to the long-term exposure to the compound as consistent with the previous report that the compound might still be present in the environment (Quinn et al. 2011). Based on the reported presence of chlorophenol and other organochlorine pesticides in South African environment, it is not surprising that *Bc*AOA isolated from wastewater is closely related to isolates that have been used in the bioremediation of various environmental pollutants.

Addition of glucose to the medium is sometimes necessary to boost the initial cell growth, but when it is in excess, its consumption may adversely affect the pH of the medium, due to acid production (Icgen et al. 2002; Patel and Kumar 2016a). Glucose is mostly used as a secondary carbon source, because, it is highly metabolizable and support maximum microbial growth (Singh et al., 2009). The presence of a secondary metabolite in the medium significantly increased PCP degradation, by the isolate, compared to its performance in medium without glucose which

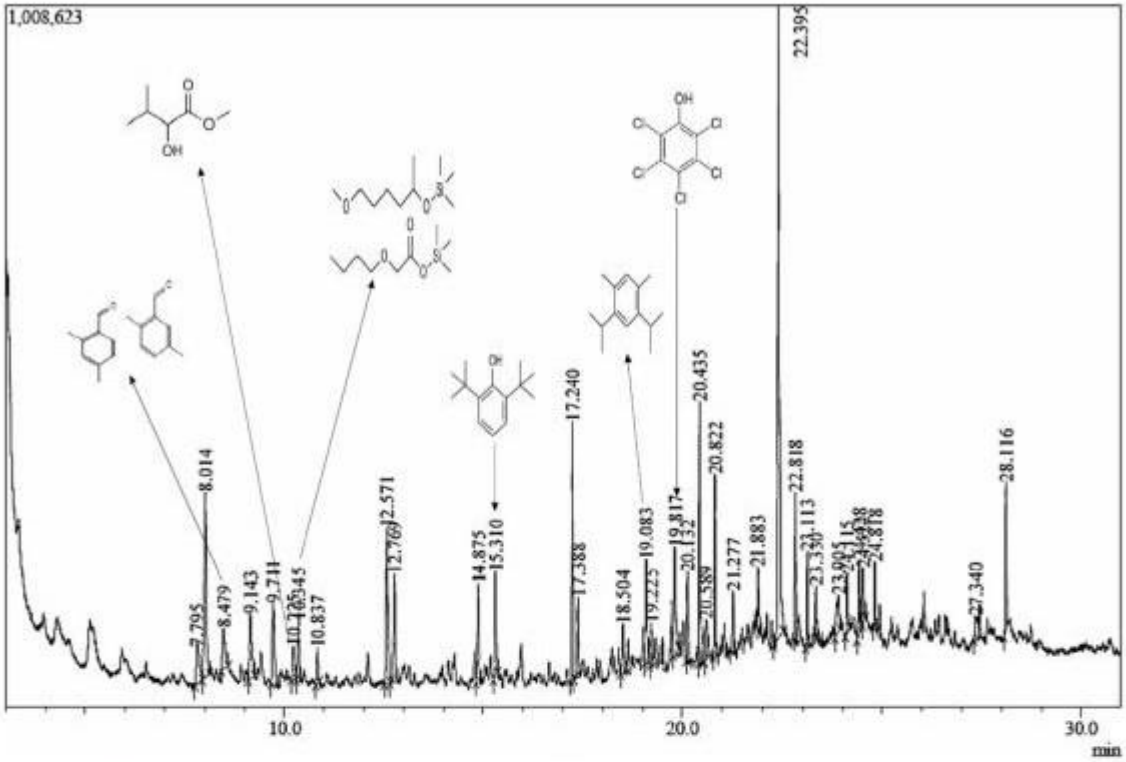


Fig. 5 GC–MS chramatogram of PCP and its metabolites in *B. cereus* AOA-CPS1

corroborates the reports of Ammeri et al. (2017) and Khessairi et al. (2014). Metabolic channelling of glucose into ATP generation, for growth, during biotransformation of xenobiotics, in cometabolism, as expressed by the isolate in this study, has also been reported (Khessairi et al. 2014; Tripathi and Garg 2010). It has also been reported (Patel and Kumar 2016a, b, c), that some degrading microorganism does not obtain energy from metabolism of the compound(s) being degraded; but rather use an alternative source of energy.

Xenobiotics degrading enzymes are expressed from a different array of amino acids. The use of inorganic nitrogen (N) sources such as $NaNO_3$, $(NH_4)_2SO_4$, NH_4NO_3 etc., as a sole N source is insufficient because *B. cereus* group requires some essential amino acids or peptides for the synthesis of functional proteins (Icgen et al. 2002). The combination of inorganic and organic N-sources has been found to be the best for the growth of most strains of *Bacillus cereus* group (Bhowmik 2014).

The isolate transformed other PCP congeners, singly and in cometabolism. Bacterial capability to mineralize mixtures of environmental pollutants have also been reported (Durruty et al. 2011a, b; Patel and Kumar 2016a, b, c). Such strains have great potentials for bio-based environmental remediation. PCP biotransformation by the isolate increased with increase in PCP concentration, this agrees with previous reports (Khessairi et al. 2014; Yuancai et al. 2014), that PCP biotransformation by some degrading strains increases with increase in the substrate concentration. The overall % residual PCP decreases with increase in PCP concentration, this is also in accordance with several reports (Ammeri et al. 2017; Joshi et al. 2015; Lopez-Echartea et al. 2016).

Moreover, maximum PCP biotransformation rates increased with an increase in PCP concentration. this is in consonance with other earlier reports (Ammeri et al. 2017; El-Bialy et al. 2018), that biotransformation of PCP increases with increase in PCP concentration and that biotransformation is not growth-

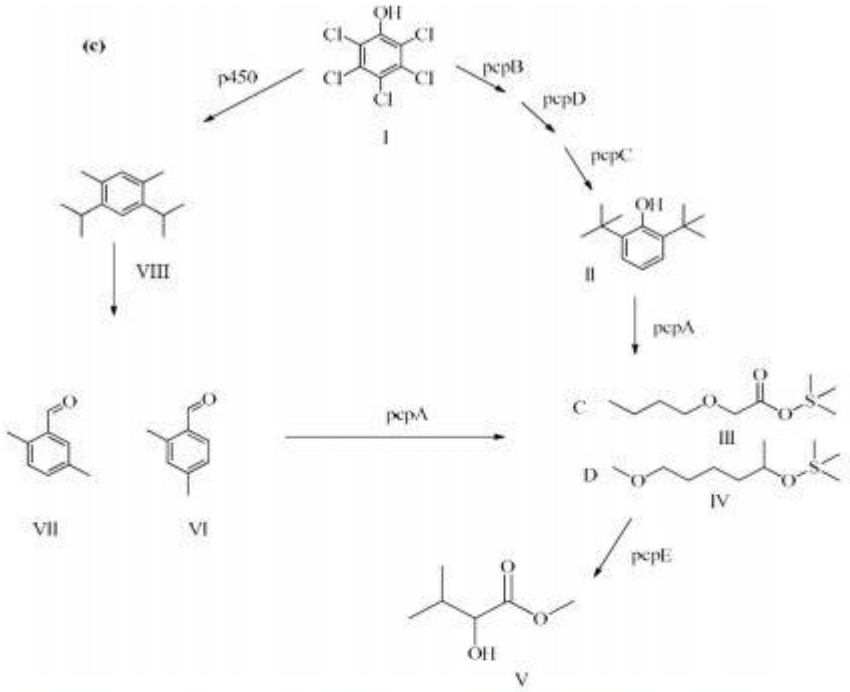


Fig. 6 Proposed PCP degradation pathways in *Bacillus cereus* AOA-CPS1. PCP (I); 2,6-bis(1,1-dimethylethyl) phenol (II); trimethylsilyl 2-butoxyacetate (III); 1-methoxy-5-trimetylsilyoxyhexane (IV); Methyl 2-hydroxyl-3-methylbutanoate (V); 2,4-Dimethylbenzenecarboxaldehyde (VI); 2,5-Dimethylbenzaldehyde (VII); 1,3-Dimethyl-4,6-diisopropylbenzene (VIII)

dependent but substrate concentration-dependent. The inhibitory effects of the compound on growth of the bacterium probably stimulate the synthesis of the biotransformation enzymes during the lag phase, which led to rapid biotransformation of the compound when the organism eventually overcame the inhibitory effects as previously observed (Yang et al. 2006).

Half-saturation or substrate utilization constant (K_s) is an affinity coefficient of an organism to the substrate (Patel and Kumar 2016a, b, c). At high substrate concentrations, degradation kinetics are independent of substrate concentration but depends on maximum specific biotransformation rate (Arnaldos et al. 2015; Jenkins and Wanner 2014). However, at low concentrations, the substrate becomes rate-limiting factor, which is mainly influenced by K_s (Arnaldos et al. 2015). A low K_s obtained in this study showed that the organism has a high affinity for the compound. Inhibition constant symbolizes the inhibitory effects of the compound on an organism. A higher K_{si} indicates the tolerance level of the organism to the compound (Patel and Kumar 2016a, b, c). The high K_{si} value obtained in this study showed that the organism has a high tolerance to the substrate, and therefore has high potential for the bioremediation of PCP at a short incubation time.

*Bc*AOA harbours genes encoding two metabolically active monooxygenases (*pcpB* and *P450*) which could individually or simultaneously initiate PCP biotransformation. It is therefore not surprising that this isolate was able to biotransform PCP and other PCP congeners of environmental importance. This further buttress the suitability of the isolate as a good candidate for biotechnological applications. The GC–MS spectra confirm the new peaks observed in the spectra while monitoring the biotransformation processes with UV–Vis spectrophotometer. The PCP degradation fragments identified in this study have also been reported (Cai and Xun 2002; Lopez-Echartea et al. 2016; Sharma et al. 2009). TeCBQ was not detected in the biotransformation assay analysed by the GC–MS, further affirming that

microbial conversion of TeCBQ to TeCHQ, is well regulated in microbial systems, during biotransformation. In most cases, TeCBQ is not released into the medium without being converted to TeCHQ, because, it is more toxic than the parent compound (Dai et al. 2003).

The fragments detected in the first pathway were similar to those earlier described in the *ortho* pathway for PCP degradation by *S. chlorophenolicum* ATCC 39,723 (Cai and Xun 2002), while the second pathway is similar to PCP degradation pathway by *Rhodococcus chlorophenolicus* PCP-1 (Apajalahti and Salkinoja-Salonen 1987; Uotila et al. 1992) and *Desulfitobacterium hafniense* strain PCP-1 (Mikesell and Boyd 1986; Villemur 2013). Undoubtedly, both 2,6-Di-*tert*-butylbenzoquinone and 1,3-Dimethyl-4,6-diisopropylbenzene cannot be metabolites from the same pathway, it could be that, one of the pathways is initiated by pcpB and the other by cytochrome P450 monooxygenase. The presence of these two monooxygenases in the organism further stresses the potential biotechnological applications of this organism.

Conclusion

Pentachlorophenol (PCP) has been listed as a priority pollutant due to its toxicological and health effects, therefore, its removal from the environmental systems is of global interest. Biodegradation of PCP is preferred over the conventional physical and chemical techniques due to the environmental friendly nature of the process and formation less toxic intermediates. In this, biodegradation of PCP by a newly isolated *Bacillus cereus* strain AOA-CPS1 (*Bc*AOA) indigenous to contaminated wastewater was investigated. *Bc*AOA was found to tolerate and degrade high concentrations of PCP. The kinetic parameter values for PCP degradation and the catabolic genes (cytochrome P450 or pentachlorophenol-4-monooxygenase) involved the degradation process were determined, resulting in the proposed pathways for PCP degradation by *Bc*AOA. Findings from this study suggest posible application of *Bc*AOA in various remediation processes for bulk bioremediation of wastewater contaminated by PCP and its congeners.

Acknowledgments O.A. thank University of KwaZulu Natal for awarding fellowship during PH.D. studies.

Author contributions O.A. and A.O conceived and designed the project. O.A. performed the experiments. M.P.M. contributed reagents and materials. O.A., A.K., M.P.M. and A.O. wrote the manuscript. All the authors have read and approved the manuscript.

Funding National Research Foundation, South Africa (Grant No: 94036 and 92803).

Compliance with ethical standards

Conflict of interest All the authors declare he has no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Ammeri RW, Mehri I, Badi S, Hassen W, Hassen A (2017) Pentachlorophenol degradation by *Pseudomonas fluorescens*. Water Qual Res J Canada 52:99–108. https://doi.org/10.2166/wqrj.2017.003

Andrews JF (1968) A Mathematical model for the continuous culture of microorganisms utilizing inhibitory substrates. Biotechnol Bioeng X:707–723. https://doi.org/10.1002/bit.260100602

Apajalahti JH, Salkinoja-Salonen MS (1987) Complete dechlorination of tetrachlorohydroquinone by cell extracts of pentachlorophenol-induced *Rhodococcus chlorophenolicus*. J Bacteriol 169:5125–5130. https://doi.org/10.1128/jb.169.11.5125-5130.1987

Arnaldos M, Amerlinck Y, Rehman U, Maere T, Van Hoey S, Naessens W, Nopens I (2015) From the affinity constant to the half-saturation index: understanding conventional modeling concepts in novel wastewater treatment processes. Water Res 70:458–470. https://doi.org/10.1016/j.watres.2014.11.046

ATSDR (2017) Agency for toxic substances and disease registry. substance priority list (candidates for toxicological profiles). 190. https://www.atsdr.cdc.gov/. Accessed 30 Sept 2020

Bhowmik A (2014) Development of a cost effective medium for enhanced production of *Bacillus thuringiensis* δ-endotoxin in partial fulfilment of the requirments for the MS degree in biotechnology. BRAC University, Dhaka. https://core.ac.uk/download/pdf/61805939.pdf. Accessed 30 Sept 2020

Cai M, Xun L (2002) Organization and regulation of pentachlorophenol-degrading genes in *Sphingobium chlorophenolicum* ATCC 39723. Society 184:4672–4680. https://doi.org/10.1128/JB.184.17.4672

Cassidy MB, Lee H, Trevors JT, Zablotowicz RB (1999) Chlorophenol and nitrophenol metabolism by *Sphingomonas* sp UG30. J Ind Microbiol Biotechnol 23:232–241. https://doi.org/10.1038/sj/jim/2900749

CDC (2018) Fourth National Report on Human Exposure to Environmental Chemicals Updated Table. URL: https://www.cdc.gov/exposurereport/pdf/FourthReport_UpdatedTables_Jul2014.pdf. Accessed 30 Sept 2020

Chanama M, Chanana S (2011) Expression of pentachlorophenol degradative genes of *Sphingobium chlorophenolicum* ATCC39723 in *Escherichia coli*. Asian J Public Heal 2:78–83

Chanama S, Crawford RL (1997) Mutational analysis of pcpA and its role in pentachlorophenol degradation by *Sphingomonas* (Flavobacterium) *chlorophenolica* ATCC 39723. Appl Environ Microbiol 63:4833–4838

Chandra R, Raj A, Yadav S, Patel DK (2009) Reduction of pollutants in pulp paper mill effluent treated by PCP-degrading bacterial strains. Environ Monit Assess 155:1–11. https://doi.org/10.1007/s10661-008-0413-4

Chen L, Yang J (2008) Biochemical characterization of the tetrachlorobenzoquinone reductase involved in the biodegradation of pentachlorophenol. Int J Mol Sci 9:198–212. https://doi.org/10.3390/ijms9030198

Chen L, Maloney K, Krol E, Zhu B, Yang J (2009) Cloning, overexpression, purification, and characterization of the maleylacetate reductase from *Sphingobium chlorophenolicum* strain ATCC 53874. Curr Microbiol 58:599–603. https://doi.org/10.1007/s00284-009-9377-z

Dai M, Rogers BJ, Warner JR, Copley SD (2003) A previously unrecognized step in pentachlorophenol degradation in *Sphingobium chlorophenolicum* is catalyzed by tetrachlorobenzoquinone reductase (PcpD). J Bacteriol 185:302–310. https://doi.org/10.1128/JB.185.1.302-310.2003

Darko G, Akoto O, Oppong C (2008) Persistent organochlorine pesticide residues in fish, sediment and water from Lake Bosomtwi, Ghana. Chemosphere 72(1):21–24. https://doi.org/10.1016/j.chemosphere.2008.02.052

DOH (2005) Pesticides registered in South Africa, including information on chemical classification, application use and relevant crops as indicated by South African MRL levels (South African Department of Health, 2005)

Durruty I, Okada E, González JF, Murialdo SE (2011a) Multisubstrate monod kinetic model for simultaneous degradation of chlorophenol mixtures. Biotechnol Bioprocess Eng 16:908–915. https://doi.org/10.1007/s12257-010-0418-z

Durruty I, Okada E, González JF, Murialdo SE (2011b) Degradation of chlorophenol mixtures in a fed-batch system by two soil bacteria. Water SA 37:547–552. https://doi.org/10.4314/wsa.v37i4.13

El-Bialy HA, Khalil OAA, Gomaa OM (2018) Bacterial-mediated biodegradation of pentachlorophenol via electron shuttling. Environ Technol 40(18):2416–2424. https://doi.org/10.1080/09593330.2018.1442501

Engst R, Macholz RM, Kujawa M, Jochen HL, Plass R (1976) The metabolism of lindane and its metabolites gamma-2,3,4,5,6-pentachlorocyclohexene, pentachlorobenzene, and pentachlorophenol in rats and the pathways of lindane metabolism. J Environ Sci Heal Part B 11:95–117. https://doi.org/10.1080/03601237609372028

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x

Gakuba E, Moodley B, Ndungu P, Birungi G (2018) Partition distribution of selected organochlorine pesticides in water, sediment pore water and surface sediment from Umgeni river, Kwazulu-Natal, South Africa. Water SA 44:232–249. https://doi.org/10.4314/wsa.v44i2.09

Gong X, Qi S, Wang Y, Julia EB, Lv C (2007) Historical contamination and sources of organochlorine pesticides in sediment cores from Quanzhou Bay, Southeast China. Mar Pollut Bull 54:1434–1440. https://doi.org/10.1016/j.marpolbul.2007.05.006

IARC (2019a) Pentachlorophenol and some related compounds. IARC monographs on the evaluation of carcinogenic risks to humans. URL: https://publications.iarc.fr/Book-And-Report-Series/Iarc-Monographs-On-The-Identification-Of-Carcinogenic-Hazards-To-Humans/Pentachlorophenol-And-Some-Related-Compounds-2019. Accessed 30 Sept 2020

IARC (2019b) International Agency for Research on Cancer monographs on the identification of carcinogenic hazards to humans. In: Agents Classified by the IARC Monographs, Volumes 1–127. URL: https://monographs.iarc.fr/agents-classified-by-the-iarc/. Accessed 30 Sept 2020

IARC (2016) International Agency for Research on Cancer (IARC) Monographs evaluate pentachlorophenol and some related compounds. Lyon, France. URL: https://www.iarc.fr/wp-content/uploads/2018/07/Volume-117_news-item.pdf. Accessed 30 Sept 2020

Icgen Y, Icgen B, Ozcengiz G (2002) Regulation of crystal protein biosynthesis by *Bacillus thuringiensis*: I. Effects of mineral elements and pH. Res Microbiol 153:599–604. https://doi.org/10.1016/s0923-2508(02)01366-9

Igbinosa EO, Odjadjare EE, Chigor VN, Igbinosa IH, Emoghene AO, Ekhaise FO, Igiehon NO, Idemudia OG (2013) Toxicological profile of chlorophenols and their derivatives in the environment: the public health perspective. Sci World J 2013:460215. https://doi.org/10.1155/2013/460215

IPEN POPRC (2013) Pentachlorophenol is one of the world's worst chemicals—Alaska Community Action On Toxics. https://www.akaction.org/pentachlorophenol_2013_10_18/. Accessed 30 Sept 2020

Jenkins D, Wanner J (2014) Activated sludge—100 years and counting. IWA Publishing, London

Joshi VV, Prewitt ML, Ma D-P, Borazjani H (2015) Enhanced remediation of pentachlorophenol (PCP)-contaminated groundwater by bioaugmentation with known PCP-degrading bacteria. Bioremediat J 19:160–170. https://doi.org/10.1080/10889868.2014.995369

Karn SK, Pan X (2017) Diversity of arsenite-oxidizing gene (aioA gene) from arsenic-rich gold mine dump of Xinjiang, China. Bioinfolet 14:149–151

Karn SK, Chakrabarty SK, Reddy MS (2010) Pentachlorophenol degradation by *Pseudomonas stutzeri* CL7 in the secondary sludge of pulp and paper mill. J Environ Sc 22:1608–1612. https://doi.org/10.1016/s0923-2508(02)01366-9

Khessairi A, Fhoula I, Jaouani A, Turki Y, Cherif A, Boudabous A, Hassen A, Ouzari H (2014) Pentachlorophenol degradation by Janibacter sp., a new Actinobacterium isolated from saline sediment of arid land. Biomed Res Int. https://doi.org/10.1155/2014/296472

Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, Zaslavsky L, Zhang J, Bolton EE (2019) PubChem 2019 update: improved access to chemical data. Nucleic Acids Res 47(D1):D1102–D1109. https://doi.org/10.1093/nar/gky1033

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Bio Evol 33(7):1870–1874. https://doi.org/10.1093/molbev/msw054

Li D, Park J, Oh JR (2001) Silyl derivatization of alkylphenols, chlorophenols, and bisphenol A for simultaneous GC/MS determination. Anal Chem 73:3089–3095. https://doi.org/10.1021/ac0014941

Lin LD, Chen YF, Wang SY, MingJer T (2009) Leachability, metal corrosion, and termite resistance of wood treated with copper-based preservative. Int Biodeterior Biodegrad 63:533–538. https://doi.org/10.1016/j.ibiod.2008.07.012

Lopez-Echartea E, Macek T, Demnerova K, Uhlik O (2016) Bacterial biotransformation of pentachlorophenol and micropollutants formed during its production process. J Environ Res Public Health Int. https://doi.org/10.3390/ijerph13111146

Marchesi JR, Sato T, Weightman AJ, Martin TA, Fry JC, Hiom SJ, Wade WG (1998) Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. Appl Environ Microbiol 64:795–799

McAllister KA, Lee H, Trevors JT (1996) Microbial degradation of pentachlorophenol. Biodegradation 7:1–40

Mikesell MD, Boyd SA (1986) Complete reductive dechlorination and mineralization of pentachlorophenol by anaerobic microorganisms. Appl Env Microbiol 52:861–865

Monod J (1949) The Growth of bacterial cultures. Annu Rev Microbiol 3:371–394. https://doi.org/10.1146/annurev.mi.03.100149.002103

Nthunya LN, Khumalo NP, Verliefde AR, Mamba BB, Mhlanga SD (2019) Quantitative analysis of phenols and PAHs in the Nandoni Dam in Limpopo Province, South Africa: A preliminary study for dam water quality management. Phys Chem Earth 112:228–236. https://doi.org/10.1016/j.pce.2019.02.003

Olaniran AO, Igbinosa EO (2011) Chlorophenols and other related derivatives of environmental concern: properties, distribution and microbial degradation processes. Chemosphere 83:1297–1306. https://doi.org/10.1016/j.chemosphere.2011.04.009

Olatunji OS (2019) Evaluation of selected polychlorinated biphenyls (PCBs) congeners and dichlorodiphenyltrichloroethane (DDT) in fresh root and leafy vegetables using GC-MS. Sci Rep 9:538. https://doi.org/10.1038/s41598-018-36996-8

Orser CS, Dutton J, Lange C, Jablonski P, Xun L, Hargis M (1993) Characterization of a Flavobacterium glutathione S-transferase gene involved in reductive dechlorination. J Bacteriol 175:2640–2644

Orser CS, Lange CC, Xun L, Zahrt TC, Schneider BJ (1993) Cloning, sequence analysis, and expression of the Flavobacterium pentachlorophenol-4-monooxygenase gene in *Escherichia coli*. J Bacteriol 175:411–416. https://doi.org/10.1128/jb.175.2.411-416.1993

Osibanjo O, Bouwman H, Bashir NHH, Okond'Ahoka J, Choong Kwet Yve R, Onyoyo HA (2003) Regionally based assessment of persistent toxic substances. Sub- Saharan Africa regional report. UNEP Chemicals, Geneva. http://ee-net.ne.jp/mec2004/UNEP-PTS/RBA%20PTS%20Global%20Report.pdf. Accessed 30 Sept 2020

Patachia S, Croitoru C (2016) Biopolymers for wood preservation. In: Jonkers HM (ed) Biopolymers and biotech admixtures for eco-efficient construction materials. Elsevier, New York, pp 305–332. https://doi.org/10.1016/B978-0-08-100214-8.00014-2

Patel BP, Kumar A (2016a) Multi-substrate biodegradation of chlorophenols by defined microbial consortium. 3 Biotech 6(2):191. https://doi.org/10.1007/s13205-016-0511-x

Patel BP, Kumar A (2016b) Biodegradation of 2,4-dichlorophenol by *Bacillus endophyticus* strain: optimization of experimental parameters using response surface methodology and kinetic study. Desalin Water Treat 57:15932–15940. https://doi.org/10.1080/19443994.2015.1076351

Patel BP, Kumar A (2016c) Optimization study for maximizing 2,4-dichlorophenol degradation by *Kocuria rhizophila* strain using response surface methodology and kinetic study. Desalin Water Treat 57:18314–18325. https://doi.org/10.1080/19443994.2015.1091988

Quinn LP, de Vos BJ, Fernandes-Whaley M, Roos C, Bouwman H, Kylin H, Pieters R, van den Berg J (2011) Pesticide use in South Africa: One of the largest importers of pesticides in Africa, Pesticides in the modern world - pesticides use and management, Stoytcheva M (ed) InTech. http://www.intechopen.com/articles/show/title/pesticide-use-in-south-africa-one-of-the-largest-importers-of-pesticides-in-africa. Accessed 30 Sept 2020

Ren H, Li Q, Zhan Y, Fang X, Yu D (2016) 2,4-Dichlorophenol hydroxylase for chlorophenol removal: substrate specificity and catalytic activity. Enzyme Microb Technol 82:74–81. https://doi.org/10.1016/j.enzmictec.2015.08.008

Saber DL, Crawford RL (1985) Isolation and characterization of *Flavobacterium* strains that degrade pentachlorophenol. Appl Environ Microbiol 50:1512–1518

Sharma A, Thakur IS, Dureja P (2009) Enrichment, isolation and characterization of pentachlorophenol degrading bacterium *Acinetobacter* sp. ISTPCP-3 from effluent discharge site. Biodegradation 20:643–650. https://doi.org/10.1007/s10532-009-9251-5

Silva V, Mol HGJ, Zomer P, Tienstra M, Ritsema CJ, Geissen V (2019) Pesticide residues in European agricultural soils—a hidden reality unfolded. Sci Total Environ 653:1532–1545. https://doi.org/10.1016/j.scitotenv.2018.10.441

Singh S, Singh ABB, Chandra AR, Patel ADK, Rai AV (2009) Synergistic biodegradation of pentachlorophenol by *Bacillus cereus* (DQ002384), *Serratia marcescens* (AY927692) and *Serratia marcescens* (DQ002385). World J Microbiol Biotechnol. https://doi.org/10.1007/s11274-009-0083-6

Smith RM (2003) Before the injection–modern methods of sample preparation for separation techniques. J Chromatogr A 1000:3–27. https://doi.org/10.1016/s0021-9673(03)00511-9

Stockholm convention (2019) Stockholm convention on persistent organic pollutants. (UNEP/POPS/COP.9/INF/16). Geneva, Switzerland. http://chm.pops.int/TheConvention/ConferenceoftheParties/Meetings/COP9/tabid/7521/ctl/Download/mid/20310/Default.aspx?id=5&ObjID=27106. Accessed 30 Sept 2020

Takeuchi M, Hamana K, Hiraishi A (2001) Proposal of the genus *Sphingomonas sensu stricto* and three new genera, Sphingobium, Novosphingobium and Sphingopyxis, on the basis of phylogenetic and chemotaxonomic analyses. Int J Syst Evol Microbiol 51:1405–1417. https://doi.org/10.1099/00207713-51-4-1405

Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Natl Acad Sci USA 101:11030–11035. https://doi.org/10.1073/pnas.0404206101

Thompson LA, Darwish WS, Ikenaka Y, Nakayama MMS, Mizukawa H, Ishizuka M (2017) Organochlorine pesticide contamination of foods in Africa: incidence and public health significance. J Vet Med Sci. https://doi.org/10.1292/jvms.16-0214

Tripathi M, Garg SK (2010) Studies on selection of efficient bacterial strain simultaneously tolerant to hexavalent chromium and pentachlorophenol isolated from treated tannery effluent. Res J Microbiol 5:707–716. https://doi.org/10.3923/jm.2010.707.716

Uotila JS, Kitunen VH, Saastamoinen T, Coote T, Haggblom MM, Salkinoja-Salonen MS (1992) Characterization of aromatic dehalogenases of *Mycobacterium fortuitum* CG-2. J Bacteriol 174:5669–5675

Villemur R (2013) The pentachlorophenol-dehalogenating *Desulfitobacterium hafniense* strain PCP-1. Philos Trans R Soc Lond B Biol Sci 368:20120319. https://doi.org/10.1098/rstb.2012.0319

WRC (2015) Persistent organic pollutants threats: threats posed by the organic pollutants to a tourist destination. http://www.wrc.org.za/wp-content/uploads/mdocs/PB_1977_PoPs%20in%20DBN%20estuaries.pdf. Accessed 30 Sept 2020

WRC (2018) Water research commission: Annual report 2017-2018. http://www.wrc.org.za/about-us/annual-reports/. Accessed 30 Sept 2020

Xu L, Resing K, Lawson SL, Babbitt PC, Copley SD (1999) Evidence That *pcpA* encodes 2,6-Dichlorohydroquinone dioxygenase, the ring cleavage enzyme required for pentachlorophenol degradation in *Sphingomonas chlorophenolica* Strain ATCC 39723. Biochemistry 38:7659. https://doi.org/10.1021/bi990103y

Yahaya A, Okoh OO, Agunbiade FO, Okoh AI (2019) Occurrence of phenolic derivatives in Buffalo River of Eastern Cape South Africa: exposure risk evaluation. Ecotoxicol Environ Saf 171:887–893. https://doi.org/10.1016/j.ecoenv.2019.01.037

Yang CF, Lee CM, Wang C (2006) Isolation and physiological characterization of the pentachlorophenol degrading bacterium *Sphingomonas chlorophenolica*. Chemosphere 62:709–714. https://doi.org/10.1016/j.chemosphere.2005.05.012

Yuancai L, Chen Y, Song W, Hu Y (2014) Enhanced selection of micro-aerobic pentachlorophenol degrading granular sludge. J Hazard Mater 280:134–142. https://doi.org/10.1016/j.jhazmat.2014.07.067

Zhang W, Sun Z (2008) Random local neighbor joining: A new method for reconstructing phylogenetic trees. Mol Phylogenet Evol 47(1):117–128. https://doi.org/10.1016/j.ympev.2008.01.019

Zhang J, Liu X, Xu Z, Chen H, Yang Y (2008) Degradation of chlorophenols catalyzed by laccase. Int Biodeterior Biodegrad 61:351–356. https://doi.org/10.1016/J.IBIOD.2007.06.015

3.12 Supplementary material

Fig. S1: Primers set used for the detection and amplification of PCP-degradation genes

	Genes	Primers sequences (5'-3')	Amplicon size (bp)
1	*cps*1A-dF	TCTGTYMCRAWTTCAAAWA	677
	*cps*1A-dR	ATRADRVAGGWARNSCHGGWA	
2	*cps*1A-F	GGG***CAT ATG***ATGAAC CAATTAAAAGGA	978
	*cps*1A-R	ACG***GGATCC*** TTACTCTTTAATAAATTCCTT	
3	*cps*1B-F	AGA ***CTCGAG***ATGACAAAGAAAACAGAA ATT	1755
	*cps*1B-R	AAA***GGATCC***TCAGTTAATCTTAGCATCATT	
4	*cps*1C-F	TTT***CTCGAG***ATGTCGCAATACATAAGGGAT	1236
	*cps*1C-R	AGT***GGATCC***TCATTTATTTTCCCCCTTCTT	
5	*cps*1D-F	TAA***CATATG***ATGATGCTAAGATTAACTGAA	315
	*cps*1D-R	GTA***GGATCC***TTATTTTCTTATAATTGC	
6	*cps*1E-F	TGT***CATATG***ACAATCAAACGCAAGAAAG	939
	*cps*1E-R	AGA***GGATCC***TTAAACAAGAACTTTCATTAC	
7	*P450*-F	GAG **GGATCC** CAT GTC AAT GAA AAA CAA AGT	
	P450-R	TAT **GCG GCCGC**G AAA GTT AAA GGC AAT TCC	1237
8	*63F*	CAGGCCTAACACATGCAAGTC	
	1387R	GGGCGGTGTGTACAAGGC	1324

Fig. S2: Retention time(s) and mass spectra of derivatized (TMS) and underivatized metabolites of pentachlorophenol biodegradation by *Bacillus cereus* AOA-CPS1

S/N	Compounds	Retention time (min)	Mass spectrum *m/z* (relative intensity)
1	2,4-Dimethylbenzaldehyde	8.480	133.10 (100); 118.90 (0.027); 105.00 (63.92); 91.0 (18.72); 77.05 (49.31); 62.90 (7.05); 51.05 (8.66)
2	3,5-Dimethylbenzaldehyde	8.480	133.10 (100); 118.90 (0.027); 105.00 (63.92); 91.0 (18.72); 77.05 (49.31); 62.90 (7.05); 51.05 (8.66)
3	2,5-Dimethylbenzaldehyde	8.480	143.10 (89.63); 118.90 (0.027);105.00 (63.92); 91.0 (18.72); 77.05 (49.31); 62.90 (7.05); 51.05 (8.66)
4	4-Ethylbenzaldehyde	8.480	143.10 (89.63); 118.90 (0.027); 105.00 (63.92); 91.0 (18.72); 77.05 (49.31); 62.90 (7.05); 51.05 (8.66)
5	Benzene-1,4-bis(1,1-dimethylethyl)	9.145	190.10 (2.39); 175.10 (100); 159.90 (0.62); 145.05 (1.92); 128.05 (3.91); 105.10 (4.39); 91.05 (9.33); 77.00 (3.80); 57.10 (54.38); 51.05 (1.25)
6	1-tert-Butyl-3-isopropyl-5-methylbenzene	9.145	190.10 (2.39); 175.10 (100); 147.10 (8.42); 133.00 (1.74); 119.10 (4.39); 105.10 (4.39); 91.05 (9.33); 80.10 (4.55); 57.10 (54.38)
7	1,3-Dimethyl-4,6-diisopropylbenzene	9.145	190.10 (2.39); 175.10 (100); 147.10 (8.42); 133.00 (1.74); 119.10 (4.39); 105.10 (4.39); 91.05 (9.33); 77.95 (2.13); 65.00 (8.87)
8	Trimethylsilyl phenylacetate	10.225	208.10 (0.40); 192.95 (8.46); 177.10 (0.26); 164.05 (15.88); 137.10 (4.39); 105.10 (0.27); 90.95 (13.00); 73.05 (100.00); 65.00 (8.14)

Fig. S2 (Cont'd): Retention time(s) and mass spectra of derivatized (TMS) and underivatized metabolites of pentachlorophenol biodegradation by *Bacillus cereus* AOA-CPS1

S/N	Compounds	Retention time (min)	Mass spectrum *m/z* (relative intensity)
9	1-Methoxy-5-trimetylsilyoxyhexane	10.225	117.05 (31.78); 106.90 (0.44); 89.05 (12.82); 73.05 (100); 58.95 (1.80)
10	Methyl 2-hydroxyl-3-methylbutanoate	10.225	90.00 (16.45); 73.05 (100.00); 58.95 (1.80)
11	Trimethylsilyl 2-butoxyacetate	10.225	161.10 (0.06); 145.10 (0.10); 117.05 (31.78); 103.00 (0.17); 89.05 (12.82); 73.05 (100.00); 57.05 (21.22)
12	2,4-Di-tert-butylphenol	15.310	206.05 (16.65); 191.10 (100); 163.05 (7.34); 135.10 (2.8); 107.05 (5); 91.05 (5.33); 74.05 (7.93); 57.10 (31.97); 50.90 (1.48)
13	2,6-Bis(tert-butyl) phenol	15.310	206.05 (16.65); 191.10 (100.00); 163.05 (7.34); 147.00 (3.40); 131.00 (1.47); 105.15 (2.71); 91.05 (5.33); 74.05 (7.93); 57.10 (31.97); 55.10 (1.97)
14	3,5-Di-t-butylphenol	15.310	206.05 (16.65); 191.10 (100); 163.05 (7.34); 135.10 (2.8); 107.05 (5.00); 91.05 (5.33); 74.05 (7.93); 57.10 (31.97); 55.10 (1.97)
15	2,5-bis(1,1-Dimethylethyl) phenol	15.310	206.05 (16.65); 191.10 (100); 163.05 (7.34); 135.10 (2.790); 107.05 (5); 88.05 (3.59); 73.05 (6.04); 57.10 (31.97); 55.10 (1.97)
16	Pentachlorophenol	19.815	265.75 (100.00); 263.75 (57.84); 229.80 (18.83); 201.80 (15.62); 168.90 (8.84); 164.90 (30.24); 129.85 (15.10); 115.00 (3.17); 95.00 (21.19); 83.10 (5.63); 60.00 (9.59)

Fig. S2 (Cont'd): Retention time(s) and mass spectra of derivatized (TMS) and underivatized metabolites of pentachlorophenol biodegradation by *Bacillus cereus* AOA-CPS1

S/N	Compounds	Retention time (min)	Mass spectrum *m/z* (relative intensity)
17	2,6-Di-tert-butylbenzoquinone	21.808	220 (100); 205 (15.08); 192 (1.64); 177 (34.07); 163 (0.95); 149 (3.17); 135 (1.05); 121 (1.24): 107 (0.09); 95 .00 (1.42); 77. 05 (2.15); 67.01 (1.04)

TMS: trimethylsilated

CHAPTER FOUR

This chapter ***has been submitted to the Journal: Bioremediation***

Optimization of pentachlorophenol degradation by a newly isolated *Bacillus tropicus* strain AOA-CPS1 using response surface methodology and elucidation of the biodegradation kinetic parameters

Oladipupo A. Aregbesola, Mduduzi Paul Mokoena and Ademola O. Olaniran*

Abstract

In this study, pentachlorophenol (PCP) degradation by a newly isolated *Bacillus tropicus* AOA-CPS1 (BtAOA) was enhanced by optimizing some environmental variables via response surface methodology (RSM). BtAOA was isolated from a contaminated wastewater sample using culture enrichment technique and the transformation of PCP (100 $mg.L^{-}1$) was monitored spectrophotometrically through scanning of the samples collected from fermentation media at relevant wavelengths. Optimization of biodegradation process parameters were assessed by RSM using the rotatable central composite experimental design. At RSM optimized conditions, *Bacillus tropicus* strain AOA-CPS1 was able to grow and degrade PCP at different initial PCP concentrations of up to 500 $mg.L^{-}1$ (at an initial PCP concentrations of 100 $mg.L^{-}1$) and the removal rate increases with an increase in PCP concentration. Thus, the maximum PCP biodegradation (98.2 %) was recorded at 500 $mg.L^{-}1$ of PCP, while the least (84.9 %) was obtained at 100 $mg.L^{-}1$. The optimized conditions (glucose (0.05% w/v), NaNO3 (0.7 $g.L^{-}1$), pH 7.4, incubation time of 6 days) increased cell growth and PCP degradation by 38.07% and 59.2%, respectively, compared to the unoptimized conditions. The degradation rates followed the first-order kinetics but switched to a zero-order when PCP concentration was increased beyond the concentration at which maximum rates were attained. This indicates that degradation was no longer dependent on substrate concentration but highly dependent on the maximum specific growth and degradation rates. The biodegradation kinetic constants for PCP biodegradation were found to be 1.064±0.114 $mg.L^{-}1.h^{-}1$ (maximum biodegradation rate); 229±19.5 $mg.L^{-}1$ (half-saturation constant); 535 $mg.L^{-}1$ (inhibition constant); and R2=0.96. The low-affinity coefficient and high inhibition constant obtained in this study showed that the bacterium has a high affinity

and tolerance for PCP, which could be explored for the biotechnological applications of the organism for transformation of high concentrations PCP.

Keywords:
Biodegradation; biodegradation kinetics; half-saturation constant; inhibition constant; response surface methodology; Substrate affinity.

4.1 Introduction

Pentachlorophenol (PCP) is a synthetic organochlorine pesticide, a conjugate acid of pentachlorophenolate and a member of pentachlorobenzenes, which comprises aromatic fungicides and a chlorophenol (CPs) (Kim et al., 2019), and is one of the World's worst chemical ever produced (IPEN, 2013). Pentachlorophenol is also a metabolite of lindane and other polychlorinated phenolic compounds (Engst et al., 1976) and it is a refractory environmental pollutant that had been widely used as a wood preservative, herbicide, defoliant, algaecide, defoliant, germicide, fungicide, and molluscicide (UNEP 2013; IARC 2019). The current use of PCP is limited to industrial areas in the treatment of utility poles, cross arms, railroad crossties, and impregnation of fibres heavy-duty textiles not intended for clothing, wooden pilings, fence posts, and timbers for construction and as a recipe in chemical synthesis (EPA 2010; IARC 2019). Based on its volatility, mobility, solubility of ionized form in water and a long-range dissemination, PCP has been detected in soils, air, lakes, rivers, basins, snow, rainwater, drinking water, plants, sediments, aquatic organisms, bacteria, fungi, eggs, as well as in mammalian milk, urine, blood, and adipose tissue (ATSDR 2001; EPA 2008). One of the major problems with the recycling of wood and waste wood materials is the potential presence of PCP used to preserve them (Yu and Kim 2012).

Pentachlorophenol is toxic to all forms of life (Igbinosa et al. 2013; Lopez-Echartea, et al., 2016), carcinogenic to humans (IARC 2019a, 2019b; Stockholm convention 2019), and resistant to microbial degradation (Lopez-Echartea et al. 2016). Furthermore, due to its persistent nature and potential for long-range dispersal, PCP and its congeners are still found in areas where use had been banned for decades (Gong et al. 2007; Quinn et al. 2011; WRC 2015; Yahaya et al. 2019). These residues may pose chronic toxicity to humans and animals through air, food intake

and water (Darko et al. 2008). In spite of its effects on human health, PCP is still been used by certified industries (Kim et al. 2019) to protect products (such as utility poles) that would in one way or the other come in contact with animals and humans. For instance, PCP, polychlorinated dibenzo-p-dioxins, and polychlorinated dibenzofurans were recently found in surface soil surrounding PCP-treated utility poles on the Kenai National Wildlife Refuge, Alaska USA and Montreal, Quebec in Canada (Verbrugge et al. 2018).

Several conventional physical and chemical techniques such as adsorption, mixing coagulation, extraction, photochemical oxidation, supersonic chemistry process, hydrogenolysis, radiolysis (Zhang *et al*. 2008), ion exchange, liquid–liquid extraction, and chemical oxidation and advanced oxidation processes have been used to remove PCP and other chlorophenol congeners from wastewater (Pera-Titus *et al*. 2004; Olaniran & Igbinosa 2011; Ren *et al*. 2016). Although these techniques are fast and reliable, they are very expensive and not environmentally friendly due to the possible formation of more toxic intermediates, that requires further processing for complete mineralization (Olaniran and Igbinosa 2011). On the contrary, bioremediation is an effective and eco-friendly method of removing chlorophenols from the environment (Olaniran and Igbinosa, 2011).

Many indigenous microorganisms cannot utilise PCP as their sole carbon source because of its high toxicity and inhibitory effects (Patachia and Croitoru 2016), thus leading to the bioaccumulation of PCP in the environment. Nonetheless, PCP have been reported to be degraded by some microorganisms, including; *Sphingobium chlorophenolicum* strain L-1 (Takeuchi *et al*. 2001), *Sphingomonas* sp UG30 (Cassidy *et al*. 1999), *Pseudomonas* sp. strain RA2 (Radehaus and Schmidt 1992), *Acinetobacter* sp. strain ISTPCP-3 (Sharma *et al*. 2009), *Pseudomonas stutzeri* CL7 and *Enterobacter* sp. SG1 (Karn *et al*. 2010), *Burkholderia cepacia* (Joshi *et al*. 2015), *Mycobacterium chlorophenolicum* PCP-1 (Uotila *et al*. 1991), *M. fortuitum* CG-2 (Uotila *et al*. 1992), *Desulfitobacterium hafniense* PCP-1 (Bisaillon *et al*. 2010; Boyer *et al*. 2003; Villemur 2013), *D. dehalogenans* (van de Pas *et al*. 1999) and *D. chlororespirans* (Krasotkina *et al*. 2001; Loffler and Sanford 1996). Some of these organisms have also evolved pathways for the complete transformation of PCP (Atashgahi et al. 2018). However, biodegradation of PCP is limited by PCP toxicity, nutrients availability and other abiotic factors such as temperature, pH etc. *In vitro* optimization of these nutrients and environmental variables

can influence microbial growth and efficient degradation of recalcitrant compounds (Patel and Kumar 2016). Optimization of nutrients and environmental variables for maximum biodegradation and enzyme production is crucial for making the industrial process more efficient, economical, and cost-effective (Singh et al. 2011).

Environmental parameters influence biotransformation potential of microbes by altering growth and the physicochemical properties of the compound (Czaplicka 2004; Patel and Kumar 2016). Conventional optimization process involves varying one parameter at a time (Singh *et al*. 2016), which is stressful and laborious, and does not account for the interactive effects of the various environmental variables on the overall performance of the organism (Patel and Kumar 2016a). However, optimization using a statistical model such as response surface methodology offers a more effective alternative to overcome these drawbacks. Response surface methodology is an empirical modulization method used for the determination of the interrelationship of a set of controlled experimental variables and observed outcomes, which requires a prior knowledge of the processes to achieve statistical model (Sridevi et al. 2011) or an integration of experimental strategies, mathematical methods, and statistical inference to determine the optimal response (Elboughdiri *et al*. 2015).

Response surface methodology also reduces the number of experimental trials needed to evaluate multiple variables and their interactive effects, which makes the technique less laborious and less time consuming than other approaches (Elboughdiri *et al*. 2015; Ravikumar et al. 2006). Response surface methodology is more suitable as the model simultaneously optimizes different variables at a time and determine their interaction effects at different coded levels (Ravikumar *et al*. 2006; Elboughdiri *et al*. 2015). Response surface methodology also provide model equation that relates the response parameters to the process variables and optimizes the same variables (Owolabi *et al*. 2018); and reduces the number of experimental trials needed to evaluate multiple variables and their interactive effects (Ravikumar *et al*. 2006; Elboughdiri *et al*. 2015). Response surface methodology has been widely used to optimize phenol degradation by *Pseudomonas aeruginosa* NCIM 2074 (Chandanalakshmi *et al*. 2010), while 23% increase in 4-chlorophenol biodegradation has been observed at RSM optimized conditions (Sahoo *et al*. 2011).

The occurrences of dangerous organochlorine pesticides in the European agricultural soils (Silva et al. 2019) and in African soils (Thompson et al. 2017) showed that our environment is contaminated with different recalcitrant compounds. South Africa is one of the four largest importers of pesticides in sub-Saharan Africa (Osibanjo et al. 2002). More than 500 pesticides were registered for use in South Africa (PAN 2010), and organochlorine pesticides are one of the major pesticides in use in South Africa (PAN 2010). These pesticides are used in almost every aspect of our everyday lives; to ensure the quantity and quality of food we eat and to control insects and pests in our homes (Quinn et al. 2011).

The detection of about seventeen organochlorine pesticides and their metabolites in sediments collected in rivers, estuaries, and canals in the eThekwini area of kwaZulu-Natal, at concentrations exceeding the method detection limit (WRC 2015) is worrisome. Also, the recent detection of chlorophenols in uMgeni river, kwaZulu-Natal (Gakuba et al. 2018), Buffalo River of Eastern Cape, South Africa (Yahaya et al. 2019), PCP congeners and Polycyclic aromatic hydrocarbons (PAHs) in Nandoni Dam in Limpopo Province of South Africa (Nthunya et al. 2019), polychlorinated biphenyls (PCBs) congeners and dichlorodiphenyltrichloroethane (DDT) in fresh root and leafy vegetables (Olatunji 2019), showed an outspread of organochlorine pesticides contamination in South African environment and require serious intervention programmes.

Recently, we reported on PCP transformation by an indigenous *Bacillus tropicus* AOA-CPS1 isolated from wastewater effluent in Durban, South Africa (Aregbesola et al. 2020). In this study, optimization strategies for PCP transformation by this organism was assessed in the presence of secondary carbon and inorganic nitrogen sources as well as pH conditions. The interactive effects of the various process parameters were evaluated, and PCP degradation under optimized RSM conditions was determined. Effects of the different initial concentration of PCP on the growth of *Bacillus tropicus* AOA-CPS1, % degradation and biodegradation kinetic parameters at the optimized conditions were also elucidated.

4.2 Materials and Methods

4.2.1 Materials and Reagents

Pure PCP, 98% analytical grade, sodium nitrate ($NaNO_3$), ammonium sulphate [$(NH_4)_2SO_4$], ammonium nitrate NH_4NO_3, Iron(II) sulphate heptahydrate ($FeSO_4.7H_2O$), potassium dihydrogen orthophosphate (KH_2PO_4), dipotassium dihydrogen phosphate (K_2HPO_4), magnesium sulphate heptahydrates ($MgSO_4.7H_2O$), D(+)-glucose anhydrous and sodium hydroxide pellets (≥97.0 %), zinc sulphate heptahydrate ($ZnSO_4$. $6H_2O$), manganese (II) sulphate tetrahydrate ($MnSO_4.4H_2O$), boric acid (H_3PO_4), ethylenediaminetetraacetic acid (EDTA) were purchased from Merck (Merck and Company, Inc., USA) and were all ACS reagents. Other chemicals and reagents used in this study were either purchased from Sigma-Aldrich or Merck and were of analytical grade standards. A stock solution of PCP was prepared as its sodium salt by dissolving PCP in 0.05N sodium hydroxide. Stock solution of sodium hydroxide (1.0 N) was prepared by dissolving sodium hydroxide pellets in sterile distilled water.

4.2.2 Sample collection, Enrichment, Isolation, and identification of *Bacillus tropicus* strain AOA-CPS1

Bacillus tropicus strain AOA-CPS1 was isolated from wastewater effluent, via culture enrichment and purified through successive sub-culturing on sterile nutrient agar plates until distinct colonies were obtained. The effluent of wastewater was obtained from wastewater treatment plants in Durban, South Africa, and transported to the laboratory for use in cultural enrichment. The samples were enriched using a minimal salt medium (MSM) as previously described (Saber and Crawford, 1985). The enrichment culture contained 45 mL of MSM supplemented with 50 $mg{\cdot}L^{-1}$ of PCP and 5 mL of sludge. The flasks were incubated for 7 days, after which 5.0 mL of the enriched culture was transferred into 45.0 mL of fresh PCP-supplemented MSM and incubated for another 7 days. Unless otherwise stated, all degradation experiments were set up in 250 mL Erlenmeyer flasks, and all degradation assays were incubated at 30 °C and 150 rpm in a shaking incubator (Innova 44 series, New Brunswick Scientific, UK).

All biodegradation experiments were conducted in triplicates, and the results presented were the means and standard deviations of triplicate experiments. All spectrophotometric readings were

taken using UV-Vis spectrophotometer (Cary 60 UV-Vis, Agilent Technologies, USA) and centrifugation was done with either Avanti J-26 XPI centrifuge (Beckman Coulter, USA) or Eppendorf centrifuge 5415D (Hamburg, Germany). After three successive transfer, 0.1 mL of the enriched culture was spread inoculated on minimal salt agar (MSM plus 15 $g \cdot L^{-1}$ of bacteriological agar) plates supplemented with 50 $mg \cdot L^{-1}$ of PCP. The plates were incubated at 30 °C until visible growths were observed. The isolates were purified via successive sub-culturing on sterile nutrient agar (NA) plates until distinct colonies were obtained. Glycerol stalks of the pure isolates were prepared and kept in the bio-freezer at -80 °C for further studies. The isolated bacteria were individually screened in PCP-MSM, and strains with potential for PCP degradation were selected for further studies.

The 16S rDNA fragment was amplified from the genomic DNA of the isolated via PCR using 63F and 1387R universal primers pair (Marchesi et al., 1998), sequenced (Inqaba Biotech., Pretoria, South Africa) and submitted at NCBI BLASTn server (Camacho et al., 2009) for the identification of the pure culture. Further, the whole genome data was generated (Inqaba Biotech, Pretoria, South Africa) using a combination of Sequel II System, PacBio Single-Molecule Real-Time (SMRT Link Version 7.0.1.66975) sequencing technology (Moine -Scientist and Applications Support, 2019), FALCON assembler and Hierarchical Genome Assembly Process 4 (HGAP4) de novo assembly analysis application. This Whole Genome Shotgun project of *Bacillus tropicus* strain AOA-CPS1 has been deposited in GenBank under accession number CP049019 (version CP049019.1) (unpublished data).

4.2.3 Biodegradation study

Biotransformation of PCP was evaluated using MSM as previously described (Saber and Crawford, 1985), with some modifications. The modified MSM contained ($g \cdot L^{-1}$): KH_2PO_4 (0.017); $MgSO_4 .7H_2O$ (0.1); $NaNO_3$ (0.5); K_2HPO_4 (0.065); 0.02 M $FeSO_4.7H_2O$ (3.0 mL), glucose (0.05 % w/v), Luria Bertani broth (0.006 %, w/v), pH 7.0 and 2 mL of micronutrients. The micronutrients contained ($mg \cdot L^{-1}$): $ZnSO_4 \cdot 7H_2O$ (4.0); $MnSO_4 \cdot 4H_2O$ (0.2); H_3BO_3 (0.15); EDTA (2.5). Ninety ml of MSM supplemented with PCP (100 $mg \cdot L^{-1}$) was inoculated with 10 mL of the standardized inoculum and the flasks were incubated for 9 days. Positive (inoculum + MSM without PCP), and negative (PCP-MSM without the inoculum) controls were set up along

with the experiment to check for cell growth and abiogenic loss of PCP during the remediation processes (Patel and Kumar 2016).

Biomass and PCP transformation were followed spectrophotometrically on a 24 h basis. Cell growth was measured at wavelength 600 nm while the transformation of PCP and formation of new peaks were monitored by scanning the sample between wavelength 350 to 200 nm using a 10 mm quartz cuvette (Hellma Analytics, Germany), after centrifugation to remove the cell pellets. Unless otherwise stated, all degradation experiments were set up in 250 mL Erlenmeyer flasks, degradation assays were incubated at 30 °C and 150 rpm in a shaking incubator (Innova 44 series, New Brunswick Scientific, UK). Biodegradation experiments were conducted in triplicates, and the results presented were the means and standard deviations of triplicate experiments. All spectrophotometric readings were taken using UV-Vis spectrophotometer (Cary 60 UV-Vis, Agilent Technologies, USA). Centrifugations were done with either Avanti J-26 XPI centrifuge (Beckman Coulter, USA) or Eppendorf microcentrifuge 5415D (Hamburg, Germany).

4.2.4 Determination of suitable secondary carbon and inorganic nitrogen sources

Suitable secondary carbon and inorganic nitrogen sources for growth and biotransformation of PCP by the organism were determined via conventional one variable at a time technique. Glucose, mannitol, citrate, glycerol, and glutamate were the carbon sources tested while $NaNO_3$, $(NH_4)_2SO_4$ and NH_4NO_3 were the inorganic nitrogen sources screened. Each carbon and nitrogen sources were assessed by incorporating each factor into PCP-MSM individually while all other factors were kept constant during the optimization process.

4.2.5 Optimization of experimental variables using the rotatable central composite design of experiment of the RSM

Optimization of biodegradation process parameters for effective and efficient transformation of PCP by the bacterium was studied via RSM. Variables optimized were secondary carbon source (glucose), $NaNO_3$, pH, and incubation time at different PCP concentrations. Rotatable central composite design (RCCD) of experiment of the RSM was used to design the optimization experiment. The RCCD is widely used for designing and modelling a second-order response surface and a design is termed rotatable when the variance of the simulated response at any point

depends only on the distance of the point from the centre point of design and allow for easy determination of the curvature (Sahoo and Barman 2012). Each numeric factor was set at 5 levels, coded as maximum, minimum, low, high and the mean (Online Resource 1). The total number of experimental runs was determined using (Eq. 1). Runs were generated with a rotatable alpha 2.38 for (k<6), 3 blocks, and the centre points. The centre points in the design include 1 replicate of factorial points, 4 centre points in each factorial block, 1 replicate of axial points, and 3 centre points at each axial block, making *Do* to be 8. The variables were fitted using the second-order polynomial regression model (Eq. 2) while (Eq. 3) was used to establish the quadratic model equation and the relationship between coded and the actual degradation values (Pandimadevi and Venkatesh 2014; Patel and Kumar 2016a).

$$n = 2^K + 2K + D_0 \qquad \textbf{Eq. 1}$$

where n is the total number of runs; D_o were the centre points used in the design and K is the independent variables to be optimized (Pandimadevi and Venkatesh 2014).

$$\gamma = \beta_0 + \sum^{k}_{i=1} \beta_i\chi_i + \sum^{k}_{i=1} \beta_{ii}\chi^2_i + \sum_i \sum_j \beta_{ij}\chi_i\chi_j \qquad \textbf{Eq. 2}$$

where Y is the predicted values; k is the number of factors variables; X_i and X_j are the independent variable s that influence the response; β_0 is the constant term; β_i is the ith linear coefficient; β_{ii} is the ith of the quadratic coefficient and β_{ij} is the ij^{th} interaction coefficient (Pandimadevi and Venkatesh 2014; Patel and Kumar 2016).

$$X_i = (U_i - U_o)/ \Delta U \qquad \textbf{Eq. 3}$$

where ΔU is the change value; X_i is the coded level of the independent factors; U_i is the actual level of the independent factors; U_o: uncoded level of the independent factor at the centre point (Pandimadevi and Venkatesh 2014).

Adequacy of the fitted model was examined via analysis of variance (ANOVA) and regression analysis. Robustness and quality of the model were assessed using the coefficient of determination (R^2-value) and p-value. Significance of the regression coefficients was examined by Student's t-test. Interactive effects of the significant model terms were expressed in contours and 3D plots. Equality of means and variances between the predicted and experimental values

were determined using a paired t-test (Pandimadevi and Venkatesh 2014; Patel and Kumar 2016a). The experiment was designed using a Design-Expert software version 11Stat-Ease (trial version). The experimental design was verified experimentally to determine its viability and feasibility of the designed experiment, using target values within the coded levels for each independent variable (Online Resource 2).

4.2.6 Biotransformation of PCP at RSM optimized conditions

Experimental run (12, 30, 34, 37, and 41), that gave high PCP degradation at maximum PCP coded level during optimization were repeated to determine whether they are reproducible (Table 1). Each run was terminated at the end of their specific incubation time. Biodegradation study with the RSM optimized conditions was assayed using 100 $mg \cdot L^{-1}$ of PCP, cell growth and the residual PCP concentration in the assay was monitored spectrophotometrically at wavelength 600 nm and 320 nm respectively, every 24 h.

Table 1: Central composite design of experiment with actual and RSM predicted responses

Run	A (PCP)	B (Glucose)	C ($NaNO_3$)	D (pH)	E (Time)	Responses (%)	
	($mg \cdot L^{-1}$)	(%)	($g \cdot L^{-1}$)		day	Actual	Predicted
1	100	0.090	0.50	7.00	4	51.0	40.7
2	85.0	0.070	0.70	7.20	5	62.6	59.2
3	100	0.050	0.50	7.40	4	77.02	66.3
4	85.0	0.070	0.70	7.20	7	53.3	49.9
5	100	0.050	0.90	7.40	6	81.2	96.5
6	70.0	0.090	0.90	7.00	4	64.0	69.2
7	100	0.050	0.50	7.00	6	41.4	46.0
8	70.0	0.050	0.90	7.40	4	55.9	64.3
9	85.0	0.070	0.70	7.20	5	62.6	58.7
10	100	0.050	0.50	7.00	4	73.1	46.6
11	70.0	0.090	0.50	7.00	6	79.4	77.1
12*	100	0.050	0.90	7.40	4	83.1	77.8
13	70.0	0.050	0.50	7.40	4	49.9	55.8
14	85.0	0.070	0.70	7.68	5	26.1	42.2
15	70.0	0.050	0.50	7.40	6	45.0	41.8
16	70.0	0.090	0.90	7.40	4	48.9	46.4
17	70.0	0.050	0.90	7.40	6	59.6	56.9
18	85.0	0.070	0.70	6.72	5	35.7	31.6
19	70.0	0.090	0.90	7.00	6	63.8	73.4
20	85.0	0.070	0.70	7.20	5	62.6	60.8
21	85.0	0.070	0.70	7.20	5	62.6	61.6
22	85.0	0.070	0.70	7.20	3	45.8	38.5
23	70.0	0.090	0.50	7.00	4	80.5	91.8
24	85.0	0.022	0.70	7.20	5	49.9	70.6
25	100	0.050	0.50	7.40	6	81.9	84.7
26	70.0	0.090	0.90	7.40	6	64.8	55.8
27	100	0.090	0.90	7.40	4	52.0	40.6

* Significant actual responses

Table 1: Central composite design of experiment with actual and RSM predicted responses (Cont'd)

Run	A (PCP)	B (Glucose)	C ($NaNO_3$)	D (pH)	E (Time)	Responses (%)	
	($mg \cdot L^{-1}$)	(%)	($g \cdot L^{-1}$)		day	Actual	Predicted
28	50.0	0.070	0.70	7.20	5	68.02	99.5
29	85.0	0.070	0.70	7.20	5	62.6	59.6
30*	100	0.090	0.50	7.40	6	83.9	68.4
31	85.0	0.118	0.70	7.20	5	57.1	62.7
32	70.0	0.050	0.90	7.00	6	68.0	48.9
33	70.0	0.050	0.50	7.00	4	81.8	51.1
34*	121	0.070	0.70	7.20	5	76.4	97.8
35	70.0	0.050	0.90	7.00	4	69.1	60.4
36	100	0.050	0.90	7.00	4	50.1	61.5
37*	100	0.090	0.90	7.00	6	83.6	52.6
38	85.0	0.070	0.70	7.20	5	62.6	61.5
39	100	0.090	0.90	7.00	4	32.9	34.5
40	70.0	0.050	0.50	7.00	6	51.0	39.7
41*	100	0.090	0.90	7.40	6	83.8	70.7
42	85.0	0.070	1.18	7.20	5	49.9	66.3
43	100	0.090	0.50	7.00	6	57.4	53.0
44	85.0	0.070	0.70	7.20	5	62.6	70.4
45	100	0.050	0.90	7.00	6	60.1	71.2
46	85.0	0.070	0.22	7.20	5	67.9	54.2
47	70.0	0.090	0.50	7.40	6	60.9	61.2
48	100	0.090	0.50	7.40	4	62.7	65.7
49	70.0	0.090	0.50	7.40	4	73.8	65.4
50	85.0	0.070	0.70	7.20	5	62.6	62.0

* Significant actual responses

4.2.7 Biodegradation kinetic studies and the effects of different initial PCP concentrations on the growth of *Bacillus tropicus* AOA-CPS1 and PCP transformation

The effects of different initial PCP concentrations on the growth of *Bacillus tropicus* strain AOA-CPS1 and % PCP removal was evaluated by growing the organism in an optimized MSM, supplemented with different initial concentrations (100, 150, 200, 250, 300, 350, 400, 450 and 500 $mg \cdot L^{-1}$) of PCP and incubated for 6 days. Cell growth and PCP removal were determined spectrophotometrically every 24 h (Ammeri et al. 2017). Kinetics studies on the isolate were evaluated in batch experiments. The specific growth rate (μ) of the organism at each initial PCP concentration tested was extrapolated from the exponential growth phase (Eq. 4) as previously described (Monod, 1949). Since PCP is inhibitory to the growth of the organism, biodegradation kinetic constants were calculated using Haldane/Andrew's substrate inhibition model (Andrews, 1968; Kargi and Eker, 2005) described in Eq. 5. At low PCP concentration ($\leq$ 250 $mg \cdot L^{-1}$), inhibition constant was ignored, Eq. 5 becomes Eq. 6, and in a linear form as Eq. 7.

$$\mu = (\mathrm{Log}_2 X_2 - \log_2 X_1) / (t_2 - t_1) \qquad \text{Eq. 4}$$

$$R_s = (R_m S/(K_s+S))\,(K_{si}/(K_{si}+S)) = \{R_m/(1+(K_s/S))\,(1+(S/K_{si}))\} \qquad \text{Eq. 5}$$

$$R_s = R_m S/(K_s+S) \qquad \text{Eq. 6}$$

$$1/R_s = (1/R_m) + (K_s/R_m)\,(1/S) \qquad \text{Eq. 7}$$

Where μ is the specific growth rate, X_2 and X_1 were the cell growth at time T_2 and T_1 respective, K_s is the half-saturation constant ($mg \cdot L^{-1}$), K_i is the substrate inhibition constant ($mg \cdot L^{-1}$), S is the substrate concentration, R_s is the specific degradation rate; R_m is the maximum biodegradation rate.

4.2.8 Statistics analysis

All statistical analyses were conducted with IBM-SPSS statistics v25 (IBM, NY, USA) and Origin Pro evaluation version 8 (OriginLab Cooperation, MA, USA).

4.3 Results and Discussion

4.3.1 Identification of *Bacillus tropicus* strain AOA-CPS1 (BtAOA) and pentachlorophenol (PCP) biodegradation

Initially, the isolate was identified as *Bacillus cereus* strain AOA-CPS1 based on the 16S rDNA sequence analysis (submitted to NCBI as accession number MH504118.1). However, a quality control test by NCBI for the submitted whole genome sequence of the strain, using an average nucleotide identity (ANI), which compares the submitted genome sequence against the whole genomes of the type strains that are already in GenBank (Ciufo et al., 2018; Federhen et al., 2016), resulted in the renaming of *Bc*AOA as *Bacillus tropicus* strain AOA-CPS1. The ANI analysis indicated that the submitted genome sequences of *Bc*AOA were 96.61% identical to the genome of the type culture of *Bacillus tropicus*, with 89.9% coverage of the genome. Consequently, *Bc*AOA was renamed as *Bacillus tropicus* strain AOA-CPS1 (based on the whole genome data submitted at NCBI under accession number CP049019).

Initial biodegradation studies (before optimization) showed that *Bacillus tropicus* strain AOA-CPS1 degraded 74% of 350 $mg{\cdot}L^{-1}$ of PCP within 9 days (Aregbesola et al. 2020), further incubation did not yield a significant increase in PCP transformation (Aregbesola et al. 2020). Also, the growth of the isolate was very low in PCP-MSM without glucose and PCP transformation was equally very low (Aregbesola et al. 2020). Apart from PCP, the organism also transformed 2,4,6-TCP, 2,4-DCP, 4-CP and 2-CP (Aregbesola et al. 2020). The strain also co-metabolized different concentrations of PCP and 2,4,6-TCP mix (Aregbesola et al. 2020).

The growth of *Bacillus tropicus* strain AOA-CPS1 in the presence of PCP decreased with increase in concentrations (Aregbesola et al. 2020). The overall % residual PCP degradation decreased with increase in PCP concentration (Aregbesola et al. 2020). The maximum removal rates increased with increase in PCP concentration at an initial concentration of $\geq$ 100 $mg{\cdot}L^{-1}$ (Aregbesola et al. 2020). The degradation followed first and zero-order kinetics at low and high PCP concentration, respectively with biokinetic constants: maximum degradation rate (0.0996 $mg{\cdot}L^{-1}{\cdot}h^{-1}$); substrate inhibition constant (723.75 $mg{\cdot}L^{-1}$); *half-saturation constant* (171.198 $mg{\cdot}L^{-1}$) and R^2 (0.98), suggesting the possibility of using this isolate for bulk degradation of PCP (Aregbesola et al. 2020).

4.3.2 Determination of suitable inorganic nitrogen and secondary carbon sources

The effects of different inorganic nitrogen sources on cell growth and PCP degradation are presented in Fig. 1A and 1B, while Fig. 1C and 1D represent the cell growth and PCP removal by *Bacillus tropicus* AOA-CPS1, in the presence of different secondary carbon (glucose, mannitol, citrate, glycerol and glutamate) sources. Glucose and sodium nitrate were found to be the most suitable secondary carbon and inorganic nitrogen sources for the isolate. Thus, glucose and sodium nitrate were used for the optimization study. Being a simple and metabolizable sugar, the use of glucose as a growth substrate for PCP biodegradation has been reported to enhance the proliferation of PCP-degrading microbes, thereby increasing their PCP removal efficiencies (Khessairi et al. 2014; Ammeri et al. 2017). Addition of a secondary carbon source (such as glucose) to the degradation medium is sometimes important to boost the initial cell growth whilst adapting and adjusting to its present environment since the compound is toxic to the organism (Icgen et al. 2002; Patel and Kumar 2016a). Metabolism of glucose for cell growth and transformation of recalcitrant compounds in co-metabolism, as expressed by *Bacillus tropicus* strain AOA-CPS1 in this study, has also been described elsewhere (Khessairi et al. 2014; Tripathi and Garg 2010). It has also been reported that some xenobiotic-degrading bacteria does not obtain energy from the compound being transformed but rather use another source of carbon (mostly simple sugar) to generate energy (Patel and Kumar 2016). However, glucose is required in a concentration-dependent manner, as glucose metabolism will drastically alter the pH of the medium which may be deleterious to the degrading enzymes and physiology of the isolate (Untereiner and Wu 2018). Furthermore, glucose is a known catabolite repressor (Magnus et al. 2017; Kim et al. 2019), that can repress the expression of PCP-degrading enzymes if not optimized in a cometabolic culture. Sodium nitrate is also known to enhance bacterial proliferation and stimulate high amino acid production, but its use is also concentration dependent (Esen and Öztürk Ürek 2014). Enzymes are produced from a combination of different amino acid residues. *Bacillus tropicus* group requires essential amino acids for the synthesis of functional enzymes (Icgen et al. 2002) and the use of inorganic nitrogen (N) sources such as NH_4NO_3, $NaNO_3$, $(NH_4)_2SO_4$, etc., as a sole N source is not sufficient. The combination of inorganic and organic N-sources has been reported to be the best for the growth of most strains of *Bacillus tropicus* group (Bhowmik 2012).

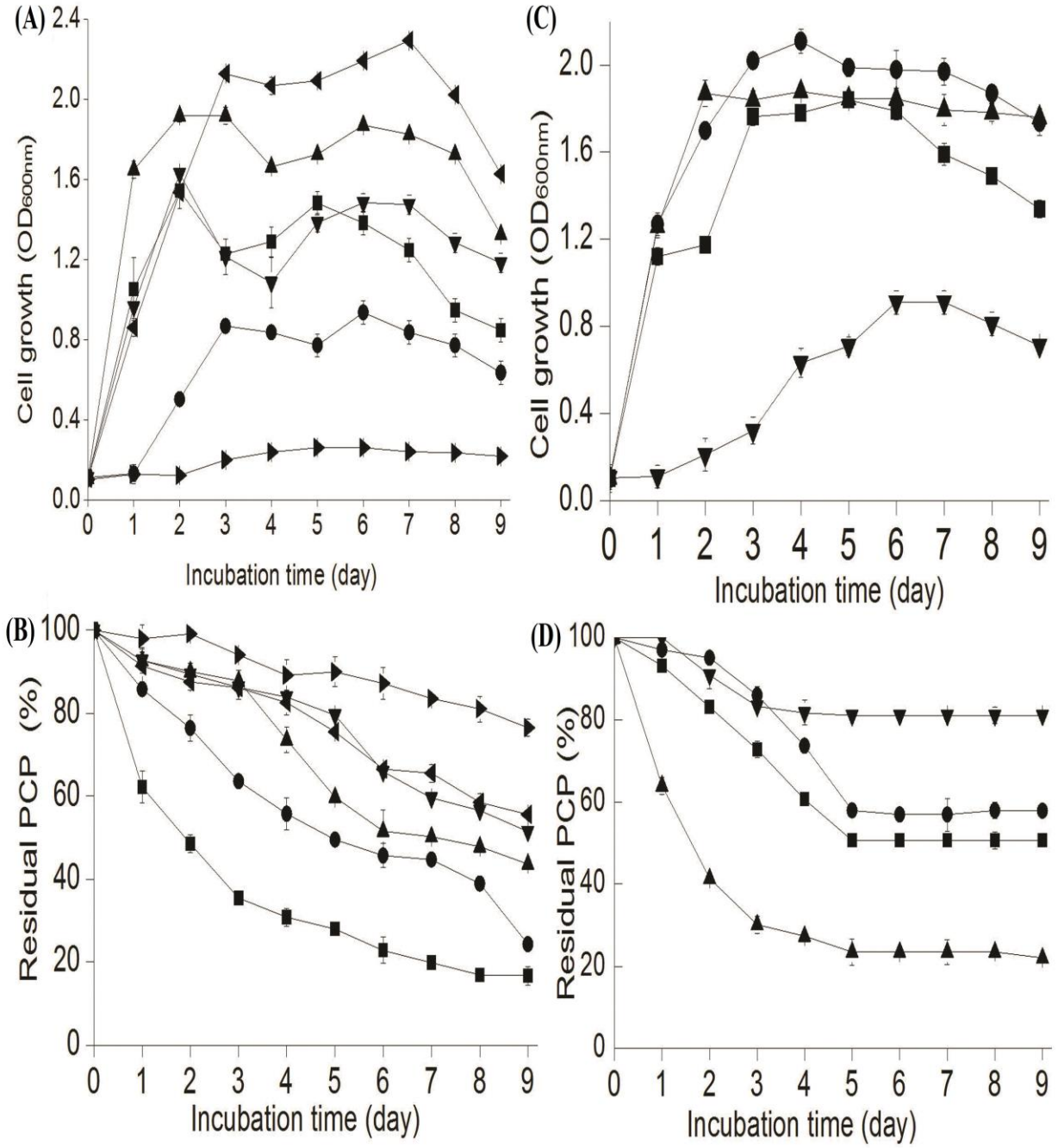


Fig. 1: Cell growth (A) and residual PCP concentration (B) in the presence of different secondary carbon sources; Glucose (■), galactose (●), glutamate (▲), citrate (▼), glycerol (◄), control (►). Cell growth (C) and residual PCP concentration (D) in the presence of different inorganic nitrogen sources; $(NH_4)_2SO_4$ (■), NH_4NO_3 (●), $NaNO_3$ (▲), control (▼).

4.3.3 Optimization of PCP degradation using a rotatable central composite design of experiment.

Viability and the interactive effects of the independent variables predicted by the response surface methodology (RSM) design were evaluated experimentally before the actual optimization run, in order to determine its desirability and efficacy of the designed experiment. The combined effects of the independent variables gave desirability of 0.937 and predicted response of 81.0 % PCP degradation (Online Resource 3), which shows that the designed experiment is experimentally practicable. After 5 days of incubation, an actual PCP transformation of 80.0 % PCP was obtained which is about 98.7 % of the predicted value of 81.0 %. The result of verification of the designed experiment showed that the experimental design is viable and can be used to navigate the design space. Biodegradation process parameters were simultaneously optimized using a rotatable centre composite design of RSM, the predicted and the actual experimental values are presented in Table 1.

Analysis of variance (ANOVA) for the actual experimental values is highlighted in Table 2, the sum of squares was calculated using the sum of squares type III (partial). An F-value of 4.56 obtained from the actual experimental values indicated that the model is highly significant and there is only 0.01 % chance that F-value of this magnitude could occur due to noise. Also, *p*-values of <0.05 obtained in this study showed that the model terms are significant, in this model, PCP and glucose (AB), PCP and pH (AD), PCP and incubation time (AE), glucose and $NaNO_3$ (BC), glucose and incubation time (BE), $NaNO_3$ and incubation time (CE), A^2 and D^2 were the significant model terms.

The regression coefficients for the actual experimental results was estimated via Student's t-test and *p*-value. The higher the t-value, the smaller the *p*-value and the more significant is the corresponding regression coefficient (Table 3). In a design with orthogonal factors such as this model, the variance inflation factors (VIFs) is 1; VIFs greater than 1 indicate multi-collinearity. It is used to diagnose collinearity/multi-collinearity of a model. Higher VIFs values mean that it is difficult and/or impossible to accurately evaluate the effects of a predictor to a model. The coefficient of determination (R^2) for the various model tested is presented in the model fit summary table (Online Resource 3).

Table 2: Analysis of variance for the actual responses

Source	Sum of Squares	Df	Mean Square	F-value	p-value	
Model	663	20	332	4.56	< 0.0001	significant
A - PCP	107	1	107	1.47	0.236	
B - Glucose	39.9	1	39.9	0.548	0.465	
C - $NaNO_3$	36.05	1	36.05	0.496	0.487	
D - pH	209	1	209	2.87	0.101	
E - Time	200	1	200	2.75	0.108	
AB	542	1	542	7.46	0.0106	significant
AC	2.32	1	2.32	0.0319	0.860	
AD	1800	1	1800	24.7	< 0.0001	significant
AE	377	1	377	5.18	0.0304	significant
BC	426	1	426	5.86	0.022	significant
BD	40.48	1	40.48	0.557	0.462	
BE	708	1	708	9.74	0.0041	significant
CD	0.376	1	0.376	0.0052	0.943	
CE	658	1	657.7	9.05	0.0054	significant
DE	15.8	1	15.8	0.217	0.645	
A^2	603	1	603	8.29	0.0074	significant
B^2	0.0196	1	0.0196	0.0003	0.987	
C^2	0.160	1	0.160	0.0022	0.963	
D^2	543	1	543	7.46	0.0106	significant
E^2	141	1	141	1.95	0.174	
Residual	211	29	72.7			
Pure Error	0.000	7	0.000			
Cor Total	8740	49				

A quadratic model was used to fit the experimental data because other models (linear and 2FI) were either of low quality or aliased with other terms. The R^2 for the quadratic model is 0.970, meaning that the sample variation of 97.0 % for PCP degradation is attributed to the experimental variables, which is in concordance with previous report (Alshehria et al. 2016; Patel and Kumar 2016), which justify the high correlation between the experimental variables and their significance in bioremediation. The equations for the quadratic model is represented in (Eq. 5), the negative (regression coefficient) values signify antagonistic effects between some variables while the positive sign represents the synergistic effect of the variables on the overall actual responses (Alshehria et al. 2015; Pandimadevi et al. 2014).

$$Y_{(actual)} = +63.3 + 1.57(A) + 0.959(B) - 0.912(C) + 2.20(D) + 2.15(E) - 4.12(AB) - 0.269(AC) + 7.50(AD) + 3.43(AE) - 3.65(BC) - 1.12(BD) + 4.70(BE) + 0.108(CD) + 4.53(CE) + 0.702(DE) + 3.29(A^2) - 0.0188(B^2) + 0.0537(C^2) - 3.12(D^2) - 1.60(E^2) \quad \textbf{Eq. 5}$$

Where $Y_{(actual)}$ is the actual response; PCP (A), glucose (B), $NaNO_3$ (C), pH (D) and incubation time (E) were the experimental variables optimized.

Equality of variances between the predicted and actual experimental values was estimated via Levene's test for equality of variance, while equality of means between the two responses was analysed by independent sample t-test. A *p*-value of 0.581 obtained from the Levene's test for equality of variances between RSM predicted figures and actual responses are >0.05, therefore, H_0 hypothesis of equality of variances between predicted and experimental values were accepted, variances in responses between predicted and actual response were not significantly different (Table 4). Also, the calculated *t* (t_{cal} = - 0.395) is less than critical *t* ($t_{critical}$ = 0.694), therefore, the null hypothesis (H_0) of equality of means between RSM predicted and experimental values was also satisfied.

Table 3: Regression coefficients for the actual laboratory response

Factor	Coefficient Estimate	df	Standard Error	95 % CI Low	95 % CI High	VIF
Intercept	63.3	1	2.99	57.2	69.46	
A-PCP	1.57	1	1.30	-1.08	4.22	1.000
B-Glucose	0.959	1	1.30	-1.69	3.61	1.000
C-$NaNO_3$	-0.912	1	1.30	-3.56	1.74	1.000
D-pH	2.20	1	1.30	-0.454	4.85	1.000
E- time	2.15	1	1.30	-0.501	4.80	1.000
AB	-4.12	1	1.51	-7.20	-1.03	1.000
AC	-0.269	1	1.51	-3.35	2.81	1.000
AD	7.50	1	1.51	4.41	10.58	1.000
AE	3.43	1	1.51	0.349	6.51	1.000
BC	-3.65	1	1.51	-6.73	-0.567	1.000
BD	-1.12	1	1.51	-4.21	1.96	1.000
BE	4.70	1	1.51	1.62	7.79	1.000
CD	0.108	1	1.51	-2.97	3.19	1.000
CE	4.53	1	1.51	1.45	7.62	1.000
DE	0.702	1	1.51	-2.38	3.78	1.000
A^2	3.29	1	1.14	0.955	5.63	1.05
B^2	-0.0188	1	1.14	-2.36	2.32	1.05
C^2	0.0537	1	1.14	-2.29	2.39	1.05
D^2	-3.12	1	1.14	-5.46	-0.786	1.05
E^2	-1.60	1	1.14	-3.93	0.744	1.05

A, D, C, D, E are the independent variables; VIF: variance inflation factor

Table 4: Equalities of means and variances between RSM predicted and experimental values

Independent sample t test										
		Levene's test		t-test for Equality of Means						
									95 % CI of the difference	
		F	Sig.	t	df	Sig. (2-tailed)	MD	SED	Lower	Upper
predicted & actual responses	EVA	0.307	0.581	-0.395	98	0.694	-1.18	2.99	-7.12	4.76
	EVNA			-0.395	96.6	0.694	-1.18	2.99	-7.12	4.76

MD: mean difference; SED: standard error difference; CI: confidence interval.

EVA: Equal variances assumed; EVNA: Equal variances not assumed

4.3.4 Interactive effects of variables on cell growth and % PCP degradation

The interactive effects of glucose and PCP at different initial concentration on the overall % PCP removal, by *Bacillus tropicus* strain AOA-CPS1 when other factors were kept at their various centre points, are shown in Fig. 2A. At the minimum glucose concentrations, an increase in PCP concentration increased the overall % PCP degradation. Increase in glucose and PCP concentrations concurrently, decrease the overall % degradation, to the centre point. Further increase in the concentration of both experimental factors beyond the centre point increased PCP removal. The interactive effects of both parameters were more prominent at high PCP concentrations. Increase in glucose concentrations above 0.10 % (w/v) greatly decreased % PCP degradation (Table 1). This could be as a result of repression of the PCP-degrading enzymes since glucose is a known catabolite enzymes repressor in some enzyme systems (López *et al*. 2015). This is in accordance with the previous report (Sridevi *et al*. 2011) that excess glucose in a degradation medium interferes with phenol degradation in some organisms. Also, fermentation of excess glucose during cell growth leads to acid production which would in turn makes the pH of the medium to become acidic which may arrest the continuous growth of the organism and degradation of PCP. Therefore, a glucose concentration of between 0.07 to 0.09 % (w/v) is appropriate for *Bacillus tropicus* strain AOA-CPS1.

The interactive effects of pH and substrate concentration on % PCP degradation are presented in Fig. 2B. At low PCP concentrations, degradation increased, with an increase in pH to the centre point. Further increase in pH above the centre point enhanced % PCP degradation. Also, at high PCP concentration and low pH, % degradation was low. However, at high PCP concentrations, an increase in pH increased degradation progressively to a maximum level. Maximum degradation was obtained at the high PCP concentration and high pH, which showed that the interaction is also significant. High PCP degradation at a high pH value is not unexpected; the chlorine contents of PCP is high and chloride ions released into the medium during dechlorination influences the pH of the medium (Pandey et al. 2009). Increase in PCP concentration of a medium with a low pH might lead to incomplete degradation, since chloride release from PCP transformation has the capacity to react with hydrogen in the medium to for hydrogen chloride which would practically reduce the pH of the medium. Once the pH of the medium is reduced to an intolerable level, cell growth might become either static or completely

inhibited. Therefore, a pH of 7.4 is appropriate for the optimum growth of *Bacillus tropicus* AOA-CPS1 and transformation of PCP at different initial concentration.

The interactive effects of incubation time and PCP concentrations on % PCP removal by the isolate is shown in Fig. 2C. At low PCP concentrations, an increase in incubation time does not have any significantly effects on the overall % degradation. This is in accordance with the fact that once the available substrate in the medium is consumed, cell growth would become static and/or decline with no changes in the overall % PCP transformation. However, as PCP concentrations increased from 88 to 100 mg l^{-1}, PCP degradation increased with increase in incubation time. Maximum % PCP degradation was obtained at the maximum incubation time, this shows that incubation time also play a prominent role in the overall PCP removal, which is in concordance with the report of El-Bialy et al. (2018).

The interactive effects of glucose and $NaNO_3$ on PCP degradation are displayed in Fig. 2D. Increase in glucose at low $NaNO_3$ concentrations increased PCP degradation. Maximum PCP degradation was obtained at the centre point for $NaNO_3$. Glucose and nitrate were supposed to be a coupling substrate to enhance biodegradation of substrates (Scherr 2013), but in this study, both appeared to be competitors for PCP degradation in accordance with a previous report (Huang et al. 2015). In another report, biomass were also found to be very sensitive to nitrate and the presence of high concentration of nitrate in the growth medium was found to increase the lag phase and totally inhibited the growth of *Pseudomonas aeruginosa* in the biodegradation of n-Hexadecane and its intermediate (Chayabutra and Ju 2000). Nitrate has also been reported to have adverse effects on the growths of *Acinetobacter baumannii* CA2, *Pseudomonas putida* CA16 and *Klebsiella* sp. CA17 in the biodegradation of 4-chloroaniline (Vangnai and Petchkroh 2007). This also buttress the previous report that the use of inorganic nitrogen (N) sources such as $NaNO_3$, NH_4NO_3 and $(NH_4)_2SO_4$ etc., as a sole nitrogen source is practically insufficient for the growth of *Bacillus tropicus* group because the group requires other essential amino acids or peptides for the synthesis of functional enzymes during degradation or metabolism of substrate (Icgen et al. 2002). Therefore, using appropriate proportions of both secondary carbon (glucose) and nitrogen sources is key for efficient and effective PCP degradation.

The interactive effects of glucose and incubation time on PCP degradation by the isolate, while other process parameters are kept at their various centre points, is shown in Fig. 2E. At a minimum glucose concentration, an increase in the incubation period to the maximum coded level did not significantly increase the overall PCP degradation. The interactive effects between both factors produced a maximum degradation at their respective highest coded levels**.** The effects of $NaNO_3$ and incubation time on PCP degradation by the isolate is represented in Fig. 2F. Increase in $NaNO_3$ concentration decrease in the overall PCP degradation. Increase in $NaNO_3$ concentrations and incubation time concurrently did not have much impact on the overall % PCP degradation. Moreover, maximum degradation was observed by keeping both factors at their minimum coded levels.

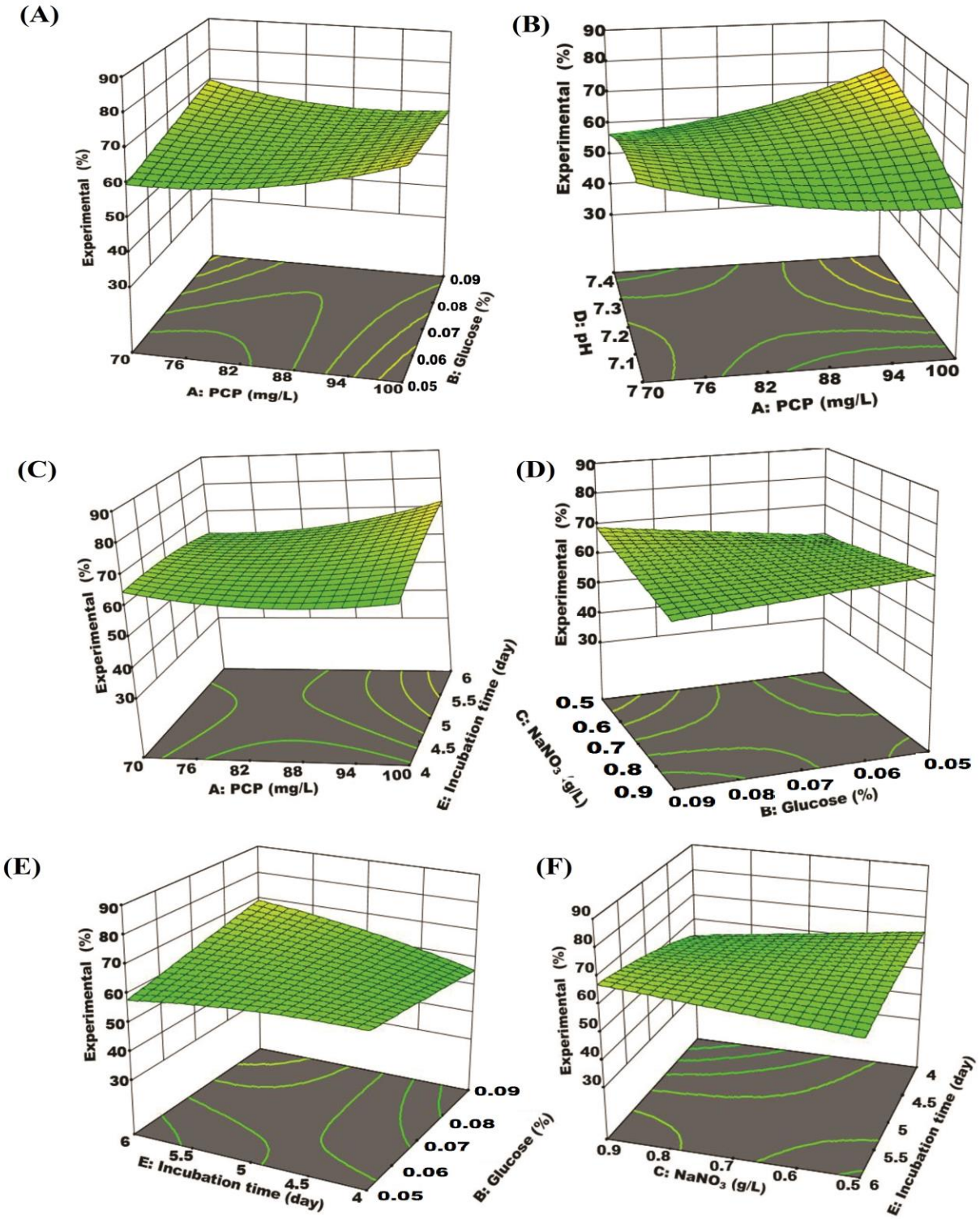


Fig. 2: Interactive effects of biodegradation process parameters on the growth of *Bacillus tropicus* strain AOA-CPS1 and PCP degradation.

4.3.5 Pentachlorophenol biodegradation at optimized RSM conditions

Experimental runs (12, 30, 34, 37, and 41), that gave higher PCP degradations, during optimization process (Table 1), were re-run to determine the best-optimized conditions for the isolate, the results were represented in Fig. 3A and 3B. Thus, optimized conditions for PCP degradation by *Bacillus tropicus* strain AOA-CPS1 were glucose (0.05% w/v), $NaNO_3$ (0.7 $g \cdot L^{-1}$), pH 7.4, incubation time (6 days). These conditions were then used for subsequent PCP biotransformation and biodegradation kinetic studies. Biotransformation of PCP (100 $mg \cdot L^{-1}$) by *Bacillus tropicus* AOA-CPS1 using the best optimized conditions gave 85.9% PCP transformation and peak cell growth of 1.91 at day 6 and 5, respectively (Fig. 3C and 3D). These represent significant 59.2% and 38.07% increase in PCP degradation and cell growth yield compared to 50.9% PCP removal at wavelength 600 nm of 0.726 biomass concentration (Aregbesola et al. 2020) obtained at unoptimized conditions with the same initial PCP concentration after 9 days. The incubation time required to produce maximum PCP degradation was greatly reduced with the RMS optimized conditions. The inhibitory effect of the compound on the growth of the organism can be seen in the growth profile when compared with the positive control (Fig. 3A).

The percentage yield of PCP transformation by *Bacillus tropicus* strain AOA-CPS1 at optimized condition in this study is higher than that of the unoptimized condition by the same organism (Aregbesola et al. 2020) and those of hexavalent chromium and PCP by *Brevibacterium casei* (Verma and Singh 2013), phenol degradation by *Pseudomonas aeruginosa* NCIM 2074 (Chandanalakshmi *et al*. 2010) and *P. aeruginosa* strain MTCC 7814 (Pandimadevi et al. 2014), 4-chlorophenol biodegradation by *Arthrobacter chlorophenolicus* A6 (Sahoo *et al*. 2011), chloroxylenol by free and immobilized *klebsiella pneumoniae* D2 (Ghanem et al. 2017), diesel oil degradation by *Enterobacter cloacae* strain KU923381 (Ramasamy et al. 2017), 2,4-dichlorophenol by *Bacillus endophyticus* and *Kocuria rhizophila* strans individually (Patel and Kumar 2016; Patel and Kumar 2016b), hydroquinone, resorcinol and catechol (Elboughdiri et al. 2015) and in microbial fuel cell (Alshehri 2015).

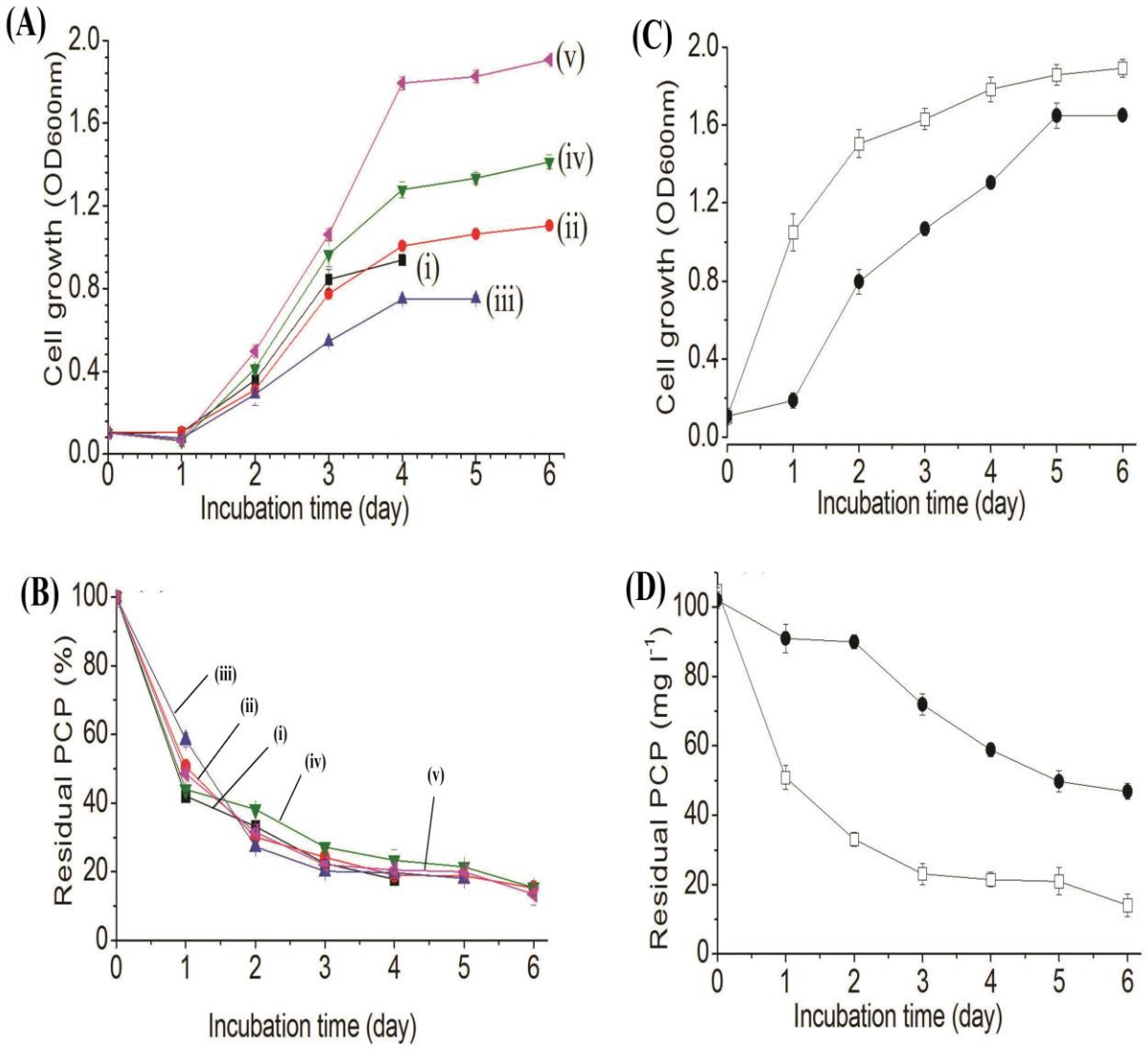


Fig. 3: Cell growth (A) and PCP degradation (B) by *Bacillus tropicus* AOA-CPS1 at all the optimized RSM conditions used for reproducibility determination. Cell growth (C) and PCP degradation (D) at the best RSM optimized conditions. Run 12 (i); run 30 (ii); run 34 (iii); run 37 (iv); run 41 (v); optimized (□); unoptimized (●) conditions.

4.3.6 Effects of different initial PCP concentrations on cell growth and PCP degradation using RSM optimized conditions.

At RSM optimized conditions, *Bacillus tropicus* strain AOA-CPS1 was able to grow and degrade PCP at different initial PCP concentrations up to 500 $mg \cdot L^{-1}$ (Fig. 4A and 4B) compared to the unoptimized conditions in previous report (Aregbesola et al. 2020). There was no significant growth at >500 $mg \cdot L^{-1}$ of PCP (Result not shown), therefore, 500 $mg \cdot L^{-1}$ of PCP is the saturation concentration for the isolate. The inhibitory effects of PCP on the growth of the organism increased with increase in concentration. As the substrate concentration increases, the lag phase of growth by the organism also increased from a minimum of 24 h to a maximum of 72 h before a visible growth could be observed in consonance with the previous report (Aregbesola et al. 2020). Maximum cell growth was recorded with 100 $mg \cdot L^{-1}$ PCP (Fig. 4A). However, as soon as the organism overcame the inhibitory effects of the compound, its growth and overall PCP removal became very rapid.

Contrary to the growth profile of the organism, PCP removal increases with an increase in PCP concentration (Fig. 4B). Thus, the maximum PCP biodegradation (98.2 %) was recorded with 500 $mg \cdot L^{-1}$ of PCP, while the least (84.9 %) was obtained with 100 $mg \cdot L^{-1}$ (Fig. 4C). The organism degraded 74 % of 350 $mg \cdot L^{-1}$ of PCP at the unoptimized condition in 9 days (Aregbesola et al. 2020). Whereas at optimized conditions, 93.3 % of the same initial PCP concentration was degraded in 6 days, which represent a significant 21.6 % increase in degradation at optimized conditions. PCP transformation by *Bacillus tropicus* strain AOA-CPS1 at optimized MSM also increased with increase in PCP concentration in accordance with the degradation pattern obtained with the unoptimized (Aregbesola et al. 2020), which is also in agreement with previous reports, that PCP transformation by some PCP-degrading bacteria increases with increase in the PCP concentration (Khessairi et al. 2014; Yuancai et al. 2014).

Furthermore, maximum PCP transformation rates increased with an increase in PCP concentration in agreement with reports of Ammeri et al. (2017) and El-Bialy et al. (2018), that biotransformation of PCP increases with increase in PCP concentration and that biotransformation is substrate concentration-dependent and not growth-dependent. The

inhibitory effects of PCP on the growth of the isolate probably stimulate the synthesis of the degradation enzymes during the lag phase, which led to rapid PCP-transformation when the bacterium eventually overcame the inhibitory and toxic effects of PCP as previously reported (Yang et al. 2006).

The observed increase in PCP biotransformation by the *Bacillus tropicus* strain AOA-CPS1 with increase in initial PCP concentrations can be explained from the bioenergetics and homeostasis point of view. Microorganisms normally make use of different bioenergetic systems to harness available energy sources based on the toxicity and concentration of the substrate (PCP) been used as an energy source while keeping their physiological states at equilibrium. At low PCP concentration, the organism probably directs most of its energy to growth at the expense of degradation, while at higher PCP concentration, the organism direct energy towards transformation of the substrate to sustain its homoeostatic balance which resulted to higher PCP degradation, which is in accordance to the previous reports (Papazi and Kotzabasis 2013; Patel and Kumar 2016).

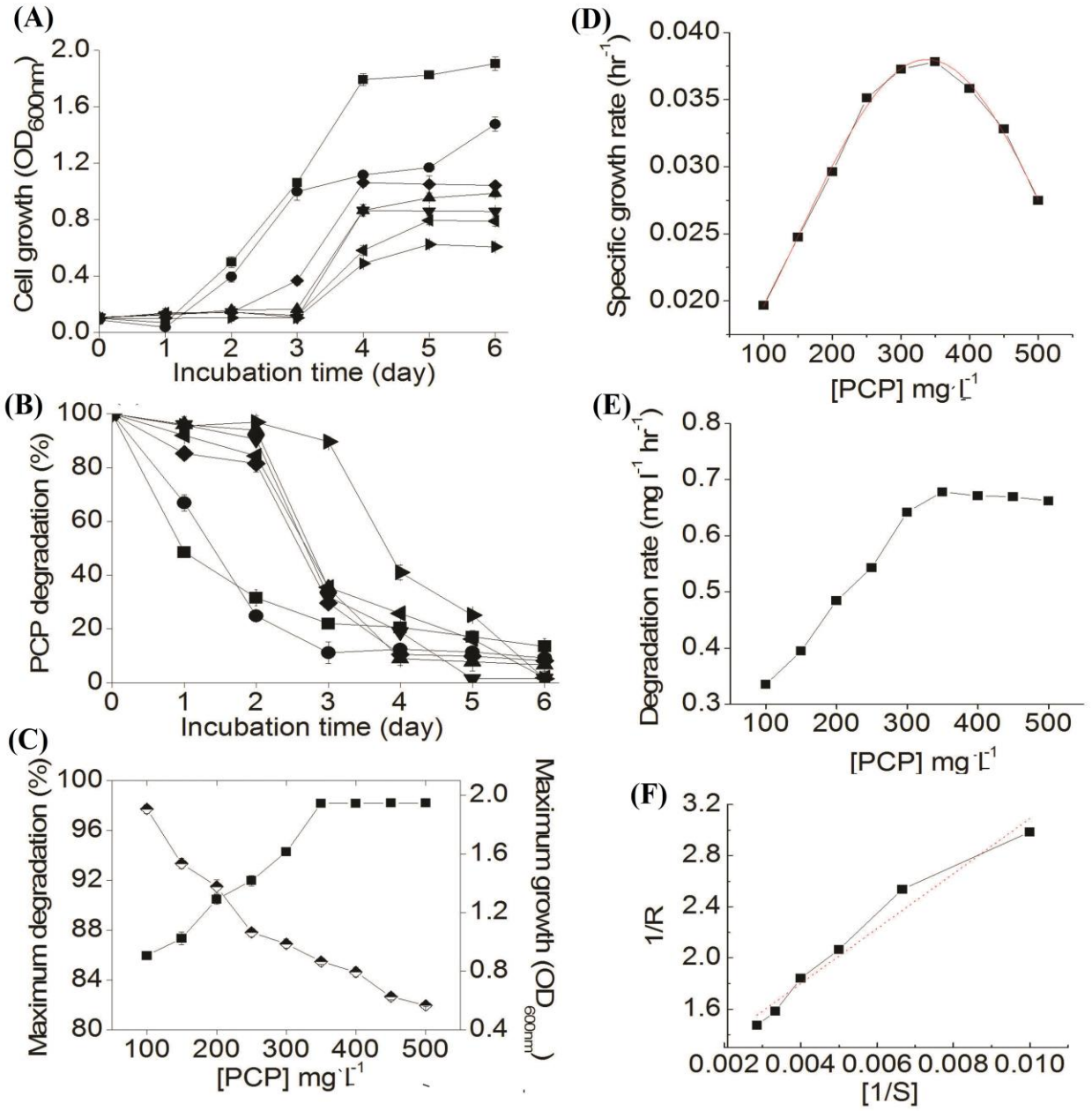


Fig. 4: Effects of different initial PCP concentrations on (A) cell growth; (B) PCP concentrations and (C) PCP biodegradation kinetics (maximum cell growth (◆) and PCP removal (■)) by *Bacillus tropicus* AOA-CPS1 Specific growth rate (D) and specific PCP degradation rate (E) at different initial PCP concentrations; and double reciprocal plot of specific PCP removal rate and PCP concentration (F). PCP concentrations ($mg \cdot L^{-1}$): 100 (■); 200 (●); 250; (◆); 300 (●); 350 (▼); 400 (◀); 500 (▶); control (⬟).

4.3.7 Biodegradation kinetic studies

Specific growth rate (Fig. 4D) and degradation rate for the organism increased with increase in PCP concentrations up to 350 $mg \cdot L^{-1}$, after which a marginal decrease was recorded. The specific PCP degradation rate (Fig. 4E) obtained in this study resembled a typical substrate inhibition curve, therefore, kinetic parameters were evaluated using the Haldane/Andrew's substrate inhibition model (Andrews 1968; Kargi and Eker 2005). At low PCP concentrations (≤ 350 $mg \cdot L^{-1}$), the inhibition constrain can be ignored, then (Eq. 6) can be written as (Eq. 7), and in a linear format as (Eq. 8). A reciprocal plot between removal rate ($1/R_s$) and PCP concentration ($1/S$) at low PCP concentration (≤ 350 $mg \cdot L^{-1}$ PCP), yielded a linear graph with a slope of K_s/R_m and an intercept of $1/R_m$. From a line of the best fit estimation, maximum PCP biotransformation rate (R_m) = 1.064 ± 0.114 (mg PCP $L^{-1} \cdot h^{-1}$); affinity constant (K_s) = 229 ± 19.5 $mg \cdot L^{-1}$, and R^2 = 0.96 were obtained.

Critical substrate concentration (i.e., the concentration at which PCP transformation rates decreased) was obtained from the derivation of equation (5) as S approaches zero (Eq. 8a). From the degradation rate curve (Fig. 7b), the maximum PCP concentration (S^*) at which optimum PCP biodegradation rate was obtained is 350 $mg \cdot L^{-1}$. Substituting this value and the value of K_s (229 $mg \cdot L^{-1}$) into (Eq. 6), PCP inhibition constant (K_{si}) for the organism is 535 $mg \cdot L^{-1}$. Therefore, the PCP biodegradation kinetic model for the isolate can be written as (Eq. 8c). The biokinetic parameters for PCP biodegradation by *Bacillus tropicus* AOA-CPS1 are maximum biodegradation rate (R_m) = 1.064 ± 0.114 mg PCP $L^{-1} \cdot h^{-1}$; half-saturation constant (K_s) = 229 ± 19.5 $mg \cdot L^{-1}$; inhibition constant (K_{si}) = 535 $mg \cdot L^{-1}$ and R^2 = 0.96.

The half-saturation constant is an index of the affinity coefficient of an organism to the substrate (Patel and Kumar 2016a). At high concentrations, degradation kinetics are independent of substrate concentration but depends on maximum specific degradation rate (Jenkins and Wanner 2014; Arnaldos *et al*. 2015). The half-saturation constant (K_S) obtained in this study showed that the organism has a high affinity for the compound. The inhibition constant shows the inhibition effects of the compound on an organism, a high K_{si} indicate lower inhibition effects and vice versa (Patel and Kumar 2016a). The high value of K_{si} obtained in this study showed that the

organism has a high tolerance for PCP, suggesting a high potential for its application in the rapid bioremediation of PCP *in vitro*.

$R_s = \{(R_m S/K_s+S)\ (K_{si}/K_{si}+S)\} = R_m/\ \{(1+(K_s/S))\ (1+(S/K_{si}))\}$ Eq. 6

$R_s = R_m S/(K_s+S)$ Eq. 7

$1/R_s = \{(1/R_m) + (K_s/R_m)\}\ (1/S)$ Eq. 8

$dRs/dS = 0$ Eq. 8a

$S^* = \sqrt{k\ sk_1}$ Eq. 8b

$R_s = \{(R_m S/(K_s+S))\ (K_{si}/(K_{si}+S)\} = 1.064/\{(1+(229/S))\ (1+(S/535))\}$ Eq. 8c

Where K_s is the half-saturation constant ($mg \cdot L^{-1}$), K_i is the substrate inhibition constant ($mg \cdot L^{-1}$), S is the substrate concentration, μ is the specific growth rate (h^{-1}), μ_{max} is the maximum specific growth rate; R_s is the specific degradation rate and R_m is the maximum biodegradation rate (Pandimadevi and Venkatesh 2014; Patel and Kumar 2016).

4.4 Conclusion

In this study, environmental variables affecting bioremediation were successfully optimized using response surface methodology (RSM). The optimized RSM conditions significantly enhanced cell growth and PCP degradation with significant decreased in incubation period, compared to unoptimized conditions. Specific growth and degradation rates were also significantly increased at RSM optimized conditions. The interactive effects of the various environmental factors highlighted the actual and limits of the concentrations of each factor to be incorporated into the degradation medium, to achieve maximum degradation efficiency. This showed that *RSM* is very efficient at optimizing various environmental variables affecting biodegradation, at the same time predict an anticipated effect of each recipe in the medium. This will guide in designing a model medium for large-scale remediation of PCP and other environmental pollutants with respects to biokinetic parameters put into consideration while designing the model. Further, apart from PCP, *Bacillus tropicus* strain AOA-CPS1 also transformed 2,4,6-trichlorophenol (2,4,6-TCP), 2,4-dichlorophenol, 4-chlorophenol and 2-chlorophenol. The strain also co-metabolized different concentrations of PCP and 2,4,6-TCP mix. *Bacillus tropicus* strain AOA-CPS1 transformed PCP and its congeners, singly and in co-metabolism. The high K_{si} value obtained in this study showed that the organism has a high

tolerance for the substrate, and therefore has high potential for fast and effective bioremediation of PCP.

4.5 Acknowledgments: University of KwaZulu-Natal and National Research Foundation, South Africa (Grant No: 94036 and 92803).

4.6 Author contributions: O.A. and A.O conceived and designed the project. O.A. performed the experiments. M.P.M. contributed reagents and materials. O.A., M.P.M. and A.O. wrote the manuscript. All the authors have read and approved the manuscript.

4.7 Funding: National Research Foundation, South Africa (Grant No: 94036 and 92803).

4.8 Conflict of interest: All the authors declare he has no conflict of interest.

4.9 Ethical statement: This article does not contain any studies with human participants or animals performed by any of the authors.

4.10 References

- Alshehria, A.N.Z., Ghanem, K.M., Al-Garni, S. M. 2016. Application of a five level central composite design to optimize operating conditions for electricity generation in a microbial fuel cell. *J. Taibah Univ. Sci.* 10, 797–804. https://doi.org/10.1016/j.jtusci.2015.01.004
- Ammeri, R.W., Mehri, I., Badi, S., Hassen, W., Hassen, A. 2017. Pentachlorophenol degradation by *Pseudomonas fluorescens*. *Water Qual. Res. J. Canada* 52, 99–108. https://doi.org/10.2166/wqrj.2017.003
- Andrews, J.F. 1968. A Mathematical model for the continuous culture of microorganisms utilizing inhibitory substrates. *Biotechnol. Bioeng. X,* 707–723.
- Aregbesola A.Oladipupo, Mduduzi P. Mokoena and Ademola O. Olaniran (2020). Biotransformation of pentachlorophenol by an indigenous *Bacillus cereus* AOA-CPS_1

isolated from wastewater effluent in Durban, South Africa. *Biodegradation* (In Press). https://doi.org/10.1007/s10532-020-09915-w.

- Arnaldos, M., Amerlinck, Y., Rehman, U., Maere, T., Van Hoey, S., Naessens, W., Nopens, I. 2015. From the affinity constant to the half-saturation index: Understanding conventional modeling concepts in novel wastewater treatment processes. *Water Res.* 70, 458–470. https://doi.org/10.1016/j.watres.2014.11.046

- Atashgahi, S., Liebensteiner, M.G., Janssen, D.B., Smidt, H., Stams, A.J.M., Sipkema, D. 2018. Microbial synthesis and transformation of inorganic and organic chlorine compounds. *Front. Microbiol.* 9, 3079. https://doi.org/10.3389/fmicb.2018.03079

- ATSDR, 2001. Toxicological profile for pentachlorophenol. Atlanta, GA: U.S. Department of Health and Human Services, Public Health Service, ATSDR, http://www.atsdr.cdc.gov/toxprofiles/tp51.pdf.

- ATSDR, 2017. Agency for Toxic Substances and Disease Registry. Substance priority list (candidates for toxicological profiles). 190.

- Bisaillon, A., Beaudet, R., Le´pine, F., Deziel, E., Villemu, R. 2010. Identification and characterization of a novel CprA reductive dehalogenase specific to highly chlorinated phenols from Desulfitobacterium hafniense strain PCP-1. *Appl. Environ. Microbio*, *1*(76), 7536–7540. https://doi.org/doi:10.1128/AEM.01362-10

- Boyer, A., Pagé-BéLanger, R., Saucier, M., Villemur, R., Lépine, F., Juteau, P., and Beaudet, R. 2003. Purification, cloning and sequencing of an enzyme mediating the reductive dechlorination of 2,4,6-trichlorophenol from Desulfitobacterium frappieri PCP-1. *The Biochemical Journal*, *373*(Pt 1), 297–303. https://doi.org/10.1042/BJ20021837

- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T. L. 2009. BLAST+: architecture and applications. *BMC bioinformatics*, 10, 421. DOI: 10.1186/1471-2105-10-421

- Chandanalakshmi, M.V.V., Sridevi, V., Narasimha, R.M., Swamy, A.V.N. 2010.

Optimization of phenol degradation from *Pseudomonas aeruginosa* (NCIM 2074) using response surface methodology. *Inter J Res. pharm chem* 1, 925–935.

- Chayabutra, C., Ju, L.K. 2000. Degradation of n-hexadecane and its metabolites by *Pseudomonas aeruginosa* under microaerobic and anaerobic denitrifying conditions. *Applied and Environmental Microbiology,* 66(2):493-498. doi: 10.1128/aem.66.2.493-498.2000.

- Ciufo, S., Kannan, S., Sharma, S., Badretdin, A., Clark, K., Turner, S., Brover, S., et al. 2018. Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *International Journal of Systematic and Evolutionary Microbiology*, 68(7), 2386–2392. DOI: 10.1099/ijsem.0.002809

- Crosby, D.G., Beynon, K.I., Greve, P.A., Korte, F., Still, G.G., Vonk, J.W. 1981. Environmental chemistry of pentachlorophenol. *Pure Appl. Chem* 53, 1051 – 1080.

- Czaplicka, M. 2004. Sources and transformations of chlorophenols in the natural environment. *Sci. Total Environ.* 322, 21–39. https://doi.org/10.1016/j.scitotenv.2003.09.015

- Darko, G., Akoto, O., Oppong, C. 2008. Persistent organochlorine pesticide residues in fish, sediment and water from Lake Bosomtwi, Ghana. *Chemosphere* 72, 21 – 24.

- DOH, 2005. Pesticides registered in South Africa, including information on chemical classification, application use and relevant crops as indicated by South African MRL levels (South African Department of Health, 2005).

- El-Bialy, H.A., Khalil, O.A.A., Gomaa, O.M. 2018. Bacterial-mediated biodegradation of pentachlorophenol via electron shuttling. *Environ. Technol.* 1–9. https://doi.org/10.1080/09593330.2018.1442501

- Elboughdiri, N., Mahjoubi, A., Shawabkeh, A., Khasawneh, H.E., Jamoussi, B., Elboughdiri, N. 2015. Optimization of the degradation of hydroquinone, resorcinol and catechol using response surface methodology. *Adv. Chem. Eng. Sci.* 5, 111–120.

https://doi.org/10.4236/aces.2015.52012

- Engst, R., Macholz, R. M., Kujawa, M., Jochen, H. L., and Plass, R. 1976. The metabolism of lindane and its metabolites gamma-2,3,4,5,6-pentachlorocyclohexene, pentachlorobenzene, and pentachlorophenol in rats and the pathways of lindane metabolism. *Journal of Environmental Science and Health, Part B*, 11(2), 95–117. DOI: 10.1080/03601237609372028

- EPA, 2008. Reregistration eligibility decision for pentachlorophenol. U.S. EPA report, september 2008, available via http://www.epa.gov/oppsrrd1/REDs/pentachlorophenol_red.pdf.

- EPA, 2010. Toxicological review of pentachlorophenol (CAS No. 87-86-5). In support of summary information on the Integrated Risk Information System (IRIS). Washington (DC), USA: United States Environmental Protection Agency.

- Esen, M., Öztürk Ürek, R. 2014. Nitrate and iron nutrition effects on some nitrate assimilation enzymes and metabolites in *Spirulina platensis*. *Turkish J. Biol.* 38, 690–700. https://doi.org/10.3906/biy-1405-37

- Federhen, S., Rossello-Mora, R., Klenk, H. P., Tindall, B. J., Konstantinidis, K. T., Whitman, W. B., Brown, D., et al. 2016. Meeting report: GenBank microbial genomic taxonomy workshop (12-13 May, 2015).

- Gakuba, E., Moodley, B., Ndungu, P., & Birungi, G. 2018. Partition distribution of selected organochlorine pesticides in water, sediment pore water and surface sediment from umngeni river, kwazulu-natal, south africa. *Water SA*, *44*(2), 232–249. https://doi.org/10.4314/wsa.v44i2.09

- Ghanem, K.M., El-Ahwany, A.M.D., Farag, A.M., Ghanem, D.A.M. 2017. Optimization the degradation of chloroxylenol by free and immobilized *klebsiella pneumoniae* D2. *International Journal of Current Advanced Research*, *6*(10), 7082–7091. https://doi.org/10.24327/ijcar.2017.7091.1077

- Herrera, Y., Okoh, A. I., Alvarez, L., Robledo, N., Trejo-Hernández, M.R. 2008.

Biodegradation of 2,4-dichlorophenol by a Bacillus consortium. *World Journal of Microbiology and Biotechnology*, *24*(1), 55–60. https://doi.org/10.1007/s11274-007-9437-0

- Huang, J.-S., Yang, P., Li, C.-M., Guo, Y., Lai, B., Wang, Y., Feng, L., Zhang, Y. 2015. Effect of Nitrite and Nitrate Concentrations on the Performance of AFB-MFC Enriched with High-Strength Synthetic Wastewater. *Biotechnology Research International*, *2015*, 798397. https://doi.org/10.1155/2015/798397

- IARC. 2019a. International Agency for Research on Cancer monographs on the identification of carcinogenic hazards to humans. In: Agents Classified by the IARC Monographs, Volumes 1–124. Retrieved from https://monographs.iarc.fr/agents-classified-by-the-iarc/

- IARC. 2019b. Pentachlorophenol and some related compounds. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. (Vol. 117). Retrieved from http://publications.iarc.fr/Book-And-Report-Series/Iarc-Monographs-On-The-Identification-Of-Carcinogenic-Hazards-To Humans/Pentachlorophenol-And-Some-Related-Compounds-2019

- Igbinosa, E.O., Odjadjare, E.E., Chigor, V.N., Igbinosa, I.H., Emoghene, A.O., Ekhaise, F.O., Igiehon, N.O., Idemudia, O.G. 2013. Toxicological profile of chlorophenols and their derivatives in the environment: the public health perspective. *Scientific World Journal*. 2013, 460215. https://doi.org/10.1155/2013/460215

- IPEN POPRC, 2013. Pentachlorophenol is one of the world's worst chemicals - Alaska Community Action On Toxics. Retrieved from https://www.akaction.org/pentachlorophenol_2013_10_18/

- Jenkins, D., Wanner, J. 2014. Activated sludge - 100 Years and counting. IWA Publishing.

- Joshi, V.V., Prewitt, M.L., Ma, D.-P., Borazjani, H. 2015. Enhanced Remediation of Pentachlorophenol (PCP)-Contaminated Groundwater by Bioaugmentation with Known

PCP-Degrading Bacteria. *Bioremediation Journal*, *19*(2), 160–170. https://doi.org/10.1080/10889868.2014.995369

- Kargi, F., Eker, S. 2005. Kinetics of 2,4-dichlorophenol degradation by *Pseudomonas putida* CP1 in batch culture. *Int. Biodeterior. Biodegrad*. 55, 25–28. https://doi.org/10.1016/j.ibiod.2004.06.010

- Karn, SK., Chakrabarty, S.K., Reddy, M.S. (2010). Pentachlorophenol degradation by *Pseudomonas stutzeri* CL7 in the secondary sludge of pulp and paper mill. *J. Environ. Sc*, *22*, 1608–1612.

- Khessairi, A., Fhoula, I., Jaouani, A., Turki, Y., Cherif, A., Boudabous, A., Hassen, A., Ouzari, H. 2014. Pentachlorophenol degradation by *Janibacter* sp., a new *Actinobacterium* isolated from saline sediment of arid land. *Biomed Res. Int.* 2014, 1–9. https://doi.org/http://dx.doi.org/10.1155/2014/296472

- Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., et al. 2019. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* 2019 Jan 8; 47(D1):D1102-1109. doi:10.1093/nar/gky1033.

- Kim, S.-B., Kwon, D.-H., Park, J.-B., Ha, S.J. 2019. Alleviation of catabolite repression in Kluyveromyces marxianus: the thermotolerant SBK1 mutant simultaneously coferments glucose and xylose. *Biotechnol. Biofuels* 12, 90. https://doi.org/10.1186/s13068-019-1431-x

- Krasotkina, J., Walters, T., Maruya, KA, Ragsdale, S. W. 2001. Characterization of the B12- and ironsulfur-containing reductive dehalogenase from *Desulfitobacterium chlororespirans*. *J. Biol. Chem*, *276*, 40991– 40997. https://doi.org/doi:10.1074/jbc.M106217200

- Loffler, F.E., Sanford, R.A., Tiedje, J. 1996. Initial characterization of a reductive dehalogenase from *Desulfitobacterium chlororespirans* Co23. *Appl. Environ. Microbiol*, *62*, 3809 – 3813.

- López, N. I., Pettinari, M. J., Nikel, P. I., Méndez, B. S. 2015. Polyhydroxyalkanoates. 73–106. https://doi.org/10.1016/bs.aambs.2015.06.001

- Lopez-Echartea, E., Macek, T., Demnerova, K., Uhlik, O. 2016. Bacterial biotransformation of pentachlorophenol and micropollutants formed during its production process. *International Journal of Environmental Research and Public Health*, *13*(1146), 1–21. https://doi.org/10.3390/ijerph13111146

- Magnus, N., Weise, T., Piechulla, B. 2017. Carbon catabolite repression regulates the production of the unique volatile sodorifen of *Serratia plymuthica* 4Rx13. Front. Microbiol. 8, 2522. https://doi.org/10.3389/fmicb.2017.02522

- Marchesi, J. R., Sato, T., Weightman, A. J., Martin, T. A., Fry, J. C., Hiom, S. J., and Wade, W. G. 1998. Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. *Applied and Environmental Microbiology*, 64(2), 795–799.

- Nthunya, L. N., Khumalo, N. P., Verliefde, A. R., Mamba, B. B. Mhlanga, S. D. 2019. Quantitative analysis of phenols and PAHs in the Nandoni Dam in Limpopo Province, South Africa: A preliminary study for dam water quality management. *Physics and Chemistry of the Earth*, *112*, 228–236. https://doi.org/10.1016/j.pce.2019.02.003

- Olaniran, A. O., Igbinosa, E. O. 2011. Chlorophenols and other related derivatives of environmental concern: Properties, distribution and microbial degradation processes. *Chemosphere*, *83*(10), 1297–1306. https://doi.org/10.1016/j.chemosphere.2011.04.009

- Olatunji, O. S. 2019. Evaluation of selected polychlorinated biphenyls (PCBs) congeners and dichlorodiphenyltrichloroethane (DDT) in fresh root and leafy vegetables using GC-MS. *Scientific Reports*, *9*(538), 1–10. https://doi.org/10.1038/s41598-018-36996-8

- Osibanjo, O., Bouwman, H., Bashir, N. H. H., Okond'Ahoka, J., Choong Kwet Yve, R., Onyoyo, H. A. 2002. Regionally based assessment of persistent toxic substances. Sub-Saharan Africa regional report. UNEP Chemicals, Geneva.

- Owolabi, R. U., Usman, M. A., Kehinde, A. J. 2018. Modelling and optimization of process variables for the solution polymerization of styrene using response surface methodology. *J. King Saud Univ. - Eng. Sci.* 30, 22–30. https://doi.org/10.1016/J.JKSUES.2015.12.005.
- PAN, 2010. PAN Pesticide database, 20.03.2011, Available from http://www.pesticideinfo.org [WWW Document].
- Pandey, J., Chauhan, A., Jain, R. K. 2009. Integrative approaches for assessing the ecological sustainability of in situ bioremediation. *FEMS Microbiol. Rev*. 33, 324–375. https://doi.org/10.1111/j.1574-6976.2008.00133.x
- Pandimadevi, M., Venkatesh, P. M., Kumar, V. 2014. Optimization of phenol degradation using *Pseudomonas aeruginosa* (MTCC 7814) by Plackett-Burman design and response surface methodology. *Journal of Bioremediation & Biodegradation*, *5*(7), 261. https://doi.org/10.4172/2155-6199.1000261
- Papazi, A., Kotzabasis, K., 2013. Rational management of dichlorophenols biodegradation by the microalga *Scenedesmus obliquus*. *PLoS One* 8, e61682. https://doi.org/10.1371/journal.pone.0061682
- Patachia, S., Croitoru, C. 2016. Biopolymers for wood preservation, in: Biopolymers and Biotech Admixtures for Eco-Efficient Construction Materials. Elsevier, pp. 305–332. https://doi.org/10.1016/B978-0-08-100214-8.00014-2
- Patel, B. P., Kumar, A. 2016. Biodegradation of 2,4-dichlorophenol by *Bacillus endophyticus* strain: optimization of experimental parameters using response surface methodology and kinetic study. *Desalination and Water Treatment*, 57, 5932-15940. https://doi.org/10.1080/19443994.2015.1076351
- Patel, B. P., Kumar, A. (2016a). Multi-substrate biodegradation of chlorophenols by defined microbial consortium. *3 Biotech*, *6*(2), 191. https://doi.org/10.1007/s13205-016-0511-x
- Patel, B.P., Kumar, A. 2016b. Optimization study for maximizing 2,4-dichlorophenol

degradation by *Kocuria rhizophila* strain using response surface methodology and kinetic study. *Desalin. Water Treat.* 57, 18314–18325. https://doi.org/10.1080/19443994.2015.1091988

- Pera-Titus, M., García-Molina, V., Baños, MA., Giménez, J., Esplugas, S. 2004 . Degradation of chlorophenols by means of advanced oxidation processes: a general review. *Appl Catal B Environ*, *47*(4), 219–256.

- Quinn, L. P., de Vos, B. J., Fernandes-Whaley, M., Roos, C., Bouwman, H., Van den Berg, J. 2011. Pesticide use in South Africa: One of the largest importers of pesticides in Africa, Pesticides in the Modern World - Pesticides Use and Management. Margarita Stoytcheva, IntechOpen. https://doi.org/10.5772/16995

- Radehaus, P. M., Schmidt, S. K. 1992. Characterization of a novel *Pseudomonas* sp. that mineralizes high concentrations of pentachlorophenol. *Applied and Environmental Microbiology*, *58*(9), 2879–2885. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1444401

- Ramasamy, S., Arumugam, A., Chandran, P. 2017. Optimization of *Enterobacter cloacae* (KU923381) for diesel oil degradation using response surface methodology (RSM). *Journal of Microbiology*, *55*(2), 104–111. https://doi.org/10.1007/s12275-017-6265-2

- Ravikumar, K., Ramalingam, S., Krishnan, S., Balu, K. 2006. Application of response surface methodology to optimize the process variables for reactive red and acid brown dye removal using a novel adsorbent. *Dyes and Pigments*, *70*, 18–26.

- Ren, H., Li, Q., Zhan, Y., Fang, X., Yu, D. 2016. 2,4-Dichlorophenol hydroxylase for chlorophenol removal: Substrate specificity and catalytic activity. *Enzyme and Microbial Technology*, *82*, 74–81. https://doi.org/10.1016/j.enzmictec.2015.08.008

- Saber, D.L., Crawford, R.L. 1985. Isolation and characterization of *Flavobacterium* strains that degrade pentachlorophenol. *Appl. Environ. Microbiol*. 50, 1512–1518.

- Sahoo, N.K., Pakshirajan, K., Ghosh, P.K., Ghosh, A. 2011. Biodegradation of 4-

chlorophenol by *Arthrobacter chlorophenolicus* A6: effect of culture conditions and degradation kinetics. *Biodegradation* 22, 275–286. https://doi.org/10.1007/s10532-010-9396-2

- Sahoo, P., Barman, T.K. 2012. ANN modelling of fractal dimension in machining. In *Mechatronics and Manufacturing Engineering* (pp. 159–226). Elsevier. https://doi.org/10.1533/9780857095893.159

- Scherr, K.E. 2013. hydrocarbon In: Vladimir Kutcherov, Anton Kolesnikov editors. Extracelular electron transfer in in situ petroleum hydrocarbon bioremediation., in: Chapter 8. https://doi.org/10.5772/53290

- Sharma, A., Thakur, I. S., and Dureja, P. 2009. . Enrichment, isolation and characterization of pentachlorophenol degrading bacterium *Acinetobacter* sp. ISTPCP-3 from effluent discharge site. *Biodegradation*, *20*, 643–650.

- Silva, V., Mol, H.G.J., Zomer, P., Tienstra, M., Ritsema, C.J., Geissen, V. 2019. Pesticide residues in European agricultural soils – A hidden reality unfolded. *Science of the Total Environment*, *653*, 1532–1545. https://doi.org/10.1016/j.scitotenv.2018.10.441

- Singh, S.K., Singh, S.K., Tripathi, V.R., Khare, S.K., Garg, S.K. 2011. Comparative one-factor-at-a-time, response surface (statistical) and bench-scale bioreactor level optimization of thermoalkaline protease production from a psychrotrophic *Pseudomonas putida* SKG-1 isolate. *Microbial Cell Factories*, *10*(1), 114. https://doi.org/10.1186/1475-2859-10-114

- Singh, V., Haque, S., Niwas, R., Srivastava, A., Pasupuleti, M., Tripathi, C.K.M. 2016. Strategies for fermentation medium optimization: An in-depth review. *Front. Microbiol.* 7, 2087. https://doi.org/10.3389/fmicb.2016.02087

- Sridevi, V., Lakshmi, M.V.V.C., Swamy, A.V.N., 2011. Implementation of response surface methodology for phenol degradation using *Pseudomonas putida* (NCIM 2102). *J. Bioremediation Biodegrad*. 02, 1–7. https://doi.org/10.4172/2155-6199.1000121

- Stockholm convention 2019. Stockholm convention on persistent organic pollutants. In *Draft guidance on best available techniques and best environmental practices for the production and use of pentachlorophenol listed with specific exemptions under the Stockholm Convention. (UNEP/POPS/COP.9/INF/16).* Geneva. Retrieved from http://chm.pops.int/TheConvention/ConferenceoftheParties/Meetings/COP9/tabid/7521/Default.aspx

- Takeuchi, M., Hamana, K., Hiraishi, A. 2001. Proposal of the genus Sphingomonas sensu stricto and three new genera, *Sphingobium*, *Novosphingobium* and *Sphingopyxis*, on the basis of phylogenetic and chemotaxonomic analyses. *Int. J. Syst. Evol. Microbiol*, *51*, 1405–1417.

- Thibodeau, J., Gauthier, A., Duguay, M., Villemur, R., Le´pine, F., Juteau, P, Beaudet, R. 2004. Purification, cloning, and sequencing of a 3,5-dichlorophenol reductive dehalogenase from *Desulfitobacterium frappieri* PCP-1. *Appl. Environ. Microbiol*, *70*, 4532–4537. https://doi.org/doi:10.1128/AEM.70.8.4532-4537.2004

- Thompson, L. A., Darwish, W. S., Ikenaka, Y., Nakayama M. M. S, Mizukawa, H., Ishizuka, M. 2017. Organochlorine pesticide contamination of foods in Africa: Incidence and public health significance. *Journal of Veterinary Medical Science*. Japanese Society of Veterinary Science. https://doi.org/10.1292/jvms.16-0214

- UNEP, 2013. United Nations Stockholm Convention on Persistent Organic Pollutants: Report of the persistent organic pollutants review committee on the work of its ninth meeting, Addendum: Risk profile on pentachlorophenol and its salts and esters.

- Untereiner, A., Wu, L. 2018. Hydrogen sulfide and glucose homeostasis: A tale of sweet and the stink. Antioxid. Redox Signal. 28, 1463–1482. https://doi.org/10.1089/ars.2017.7046

- Uotila, J.S., Kitunen, V.H., Saastamoinen, T., Coote, T., Haggblom, M.M., Salkinoja-Salonen, M. 1992. Characterization of aromatic dehalogenases of *Mycobacterium fortuitum* CG-2. *J Bacteriol*, *174*, 5669–5675.

- Uotila, J.S., Salkinoja-Salonen, M.S., Apajalahti, J.H.A. 1991. Dechlorination of pentachlorophenol by membrane boundenzymes of *Rhodococcus chlorophenolicus* PCP-I. *Biodegradation*, *2*, 25–31.

- van de Pas, B. A., Smidt, H., Hagen, W. R., van der Oost, J., Schraa, G., Stams, A. J. M., de, V. W. 1999. Purification and molecular characterization of ortho-chlorophenol reductive dehalogenase, a key enzyme of halorespiration in *Desulfitobacterium dehalogenans*. *J. Biol. Chem*, *274*, 20 287–20 292. https://doi.org/doi:10.1074/jbc.274.29.20287

- Vangnai, A. S., Petchkroh, W. 2007. Biodegradation of 4-chloroaniline by bacteria enriched from soil. *FEMS Microbiology Letters*, *268*(2), 209–216. https://doi.org/10.1111/j.1574-6968.2006.00579.x

- Verma, T., Singh, N. 2013. Isolation and process parameter optimization of *Brevibacterium casei* for simultaneous bioremediation of hexavalent chromium and pentachlorophenol. *Journal of Basic Microbiology*, *53*(3), 277–290. https://doi.org/10.1002/jobm.201100542

- Villemur, R. 2013. The pentachlorophenol-dehalogenating *Desulfitobacterium hafniense* strain PCP-1. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 368, 20120319. https://doi.org/10.1098/rstb.2012.0319

- WRC, 2015. Persistent organic pollutants Threats: Threats posed by the organic pollutants to a tourist destination. [WWW Document]. Water Res. Comm. Repub. South Africa. URL http://www.wrc.org.za/wp-content/uploads/mdocs/PB_1977_PoPs in DBN estuaries.pdf (accessed 10.19.19).

- Yahaya, A., Okoh, O.O., Agunbiade, F.O., Okoh, A.I. 2019. Occurrence of phenolic derivatives in Buffalo River of Eastern Cape South Africa: Exposure risk evaluation. Ecotoxicol. *Environ. Saf.* 171, 887–893. https://doi.org/10.1016/j.ecoenv.2019.01.037

- Yu, C.W.F., Kim, J.T. 2012. Long-term impact of formaldehyde and VOC emissions from wood-based products on indoor environments; and issues with recycled products.

Indoor Built Env. 21, 137–149.

- Zhang, J., Liu, X., Xu, Z., Chen, H., Yang, Y. 2008. Degradation of chlorophenols catalyzed by laccase. *International Biodeterioration and Biodegradation*, *61*(4), 351–356. https://doi.org/10.1016/J.IBIOD.2007.06.015

4.11 ONLINE RECOURSES

4.11.1 Online Recourses 1

Table S1: Central composite design (CCD) of independent variables and their corresponding actual and coded levels used for the optimization experiment.

			Actual levels			Coded levels			
Variable	Name	Units		Min	Max	Coded low	Coded high	Mean	Std. Dev.
A	PCP	mg l^{-1}	70 - 100	49.3	120.7	-1 ↔ 70.0	+1 ↔ 100	85.0	13.7
B	Glucose	%	0.5 – 00.9	0.22	1.18	-1 ↔ 0.50	+1 ↔ 0.90	0.70	0.183
C	$NaNO_3$	g l^{-1}	0.5 – 0.9	0.22	1.18	-1 ↔ 0.50	+1 ↔ 0.90	0.70	0.183
D	pH		7.0 – 7.4	6.72	7.68	-1 ↔ 7.00	+1 ↔ 7.40	7.20	0.183
E	Time	day	4 – 6	2.62	7.38	-1 ↔ 4.00	+1 ↔ 6.00	5.00	0.913

Min: minimum; Max: maximum;

4.11.2 Online Recourses 2

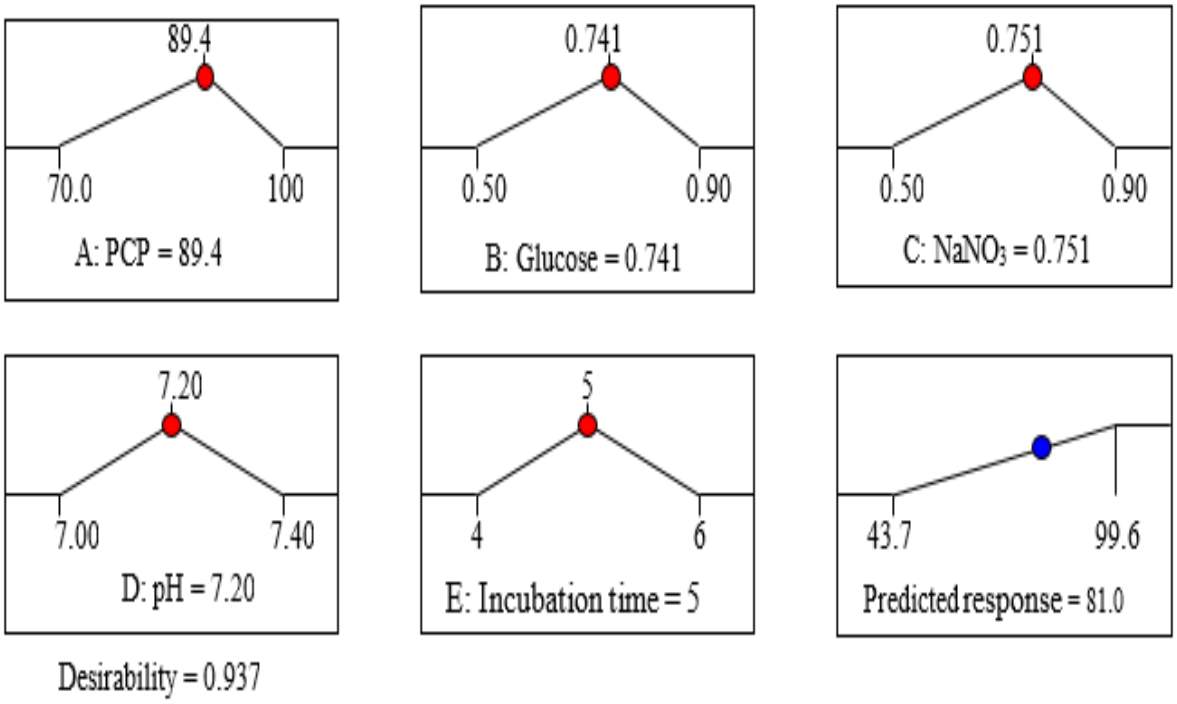


Figure S1: Verification of the designed experimental

4.11.3 Online Recourses 3

Table S2: Response surface methodology design Fit Summary for the quadratic model

Source	Sequential p-value	Lack of Fit p-value	Adjusted R^2	Predicted R^2	
Linear	0.0140	0.0006	0.198	-0.165	
2FI	0.0005	0.0030	0.571	0.493	
Quadratic*	**< 0.0001**	**0.861**	**0.970**	**0.939**	**Suggested**
Cubic	0.874	0.617	0.959	0.654	Aliased

CHAPTER FIVE

This chapter is being finalised for submission

Classic Pentachlorophenol Hydroxylating Phenylalanine 4-Monooxygenase from Indigenous *Bacillus tropicus* strain AOA-CPS1: Cloning, Overexpression, Purification, Characterization and Structural Homology Modelling

Oladipupo A. Aregbesola, Ajit Kumar, Mduduzi P. Mokoena and Ademola O. Olaniran*

Abstract

Phenylalanine hydroxylating system (PheOHS) is an efficient metabolic operon comprising phenylalanine 4-monooxygenase (Phe4MO), pterin-4α-hydroxytetrahydrobiopterin dehydratase (PCD) and dihydropteridine reductase acting cooperatively as a self-sufficient, metabolic, and regulatory system. The metabolically promiscuous pentachlorophenol (PCP) hydroxylating Phe4MO (represented as CpsB in the present study) was detected, amplified (from the genome of *Bacillus tropicus* strain AOA-CPS1), cloned, overexpressed, purified, and characterised. The 1.755 kb gene cloned in pET15b expression vector expressed a ≅64 kDa monomeric protein which was purified to homogeneity by single step affinity chromatography, with a total yield of 82.1%. The optimum temperature and pH of the enzyme were found to be 30ºC and 7.0, respectively. CpsB showed functional stability between pH 6.0-7.5 and temperature 25ºC-30ºC. CpsB activity was stimulated by Fe^{2+}, Ca^{2+}, EDTA (0.5-1.5 mmol) and dithiothreitol (0.5-1.0 mmol) but inhibited by Na-azide and SDS (>0.5 mmol). The enzyme substrate reaction kinetics studies showed allosteric nature of the enzyme and followed pre-steady state using NADH as a co-substrate with apparent v_{max}, K_m, k_{cat} and k_{cat}/K_m values of 0.465 $\mu mol \cdot s^{-1}$, 140 µmol, 0.099 s^{-1} and 7.07 x 10^{-4} $\mu mol^{-1} \cdot s^{-1}$, respectively, for the substrate PCP and 0.259 $\mu mol \cdot s^{-1}$, 224 µmol, 0.055 s^{-1} and 2.47 x 10^{-4} $\mu mol^{-1} \cdot s^{-1}$, respectively, for co-substrate NADH. The Hill plots and M-W-C model reveal CpsB allosteric nature and belong to K-System. The in-gel trypsin digestion experiments and bioinformatics tools confirmed that the reported enzyme is a Phe4MO. CpsB exhibited multiple putative conserved domains and metal ion binding site, confirming its allosteric nature. Though, Phe4MO has been reported to have a diverse catalytic function, here we report, for the first time, that it functions as a PCP dehalogenase or PCP-4-monooxygenase

by hydroxylating PCP. Hence, the use of this enzyme may be further explored in the bioremediation of PCP and other related xenobiotics.

Keywords: Pentachlorophenol; Phenylalanine 4-monooxygenase; Phenylalanine hydroxylase

5.1 Introduction

Pentachlorophenol (PCP) is a priority pollutant [1], previously used as fungicides, pesticides, and wood preservative [2]. The current use of PCP has been restricted to certified industries for wood preservation of railroad ties, wharf pilings and utility [3], due to its toxicity [4]. PCP is toxic to all forms of life [2], carcinogenic hazard to humans [5–7], and resistant to microbial degradation [2]. However, PCP degradation have been reported in some microorganisms [2,8–16], including an indigenous *Bacillus tropicus* strain AOA-CPS1 recently isolated [101]. Furthermore, PCP degradation pathway and enzymes involved have been studied extensively in *Sphingobium chlorophenolicum* L-1 [17].

The first step in the PCP degradation pathway (Fig. S1a) as reported in *S. chlorophenolicum* is the rate-limiting step involving the hydroxylation of PCP to 1,4-tetrachlorobenzoquinone (Tet-CBQ) [18]. The reaction is catalysed by enzyme PCP 4-monooxygenase (PcpB) [18] and cytochrome p450 monooxygenase [19]. The PCP degradation pathway of the model organism (*S. chlorophenolicum* L-1) was assembled in a patchwork manner using catabolic enzymes acquired via different horizontal genes transfers [18,20]. Hence, the pathway is still at the development stage and the first three enzymes in the pathway (PcpB, PcpD and PcpC) are shown to be metabolically inefficient [18,20,21].

Enzymes associated with intermediary metabolism such as phenylalanine hydroxylation system (PheOHS) may have some yet unrecognised alternative catalytic function(s) of protecting the organism from potential toxic compounds [22]. The PheOHS catalyses the irreversible hydroxylation of phenylalanine (Phe) to tyrosine (Tyr) and recycling of tetrahydrobiopterin [23]. A typical mammalian PheOHS (Fig. S1b) comprised of phenylalanine 4-monooxygenase (Phe4MO), pterin-4α-carbinolamine dehydratase/dimerization of hepatocyte nuclear factor-1

(PCD/DCoH), dihydropteridine reductase, tetrahydrobiopterin (BH4) and molecular oxygen [24]. Phe4MO hydroxylates Phe to Tyr while PCD/DCoH and a reductase convert the intermediate back to BH4 [23] in a rate-limiting step [25]. Unlike the mammalian and other higher animals, PheOHS in most bacteria do not produce BH4 where the primary role of PheOHS in bacteria is to synthesize folate for one-carbon transfer reactions [26].

It has been previously reported that Phe4MO also play a critical role in xenobiotic compound degradation [27] and lipid metabolism [23]. The mammalian Phe4MO catalyses the rate-limiting step in the Phe catabolism, transforming about 75% of Phe in diet and protein catabolism under physiological states [28,29]. Mutations in Phe4MO gene causes phenylketonuria (PKU) in humans [29]. Phe4MO's have a high mutability index, with at least >500 mutations detected in the Phe4MO's genes that cause PKU and they are inherited in an autosomal recessive manner [30]. In addition, most mutations are mainly associated with Phe4MO instability and misfolding [29].

To the best of our knowledge, the catalytic promiscuity of Phe4MO in the biotransformation of chlorophenols like PCP and other organochlorine pesticides has not been explored. Studies on the PCP biotransformation by this recently isolated indigenous strain *Bacillus tropicus* strain AOA-CPS1 (*Bt*AOA), revealed efficient degradation of PCP. The project was further extended to investigate the genes and enzymes involved in the PCP degradation pathway in *Bt*AOA. Indirect experimental approaches and whole genome sequence of *Bt*AOA revealed the presence of genes involved in PCP degradation pathway (Unpublished data).

The whole genome annotation data (GeneBank accession no: CP049019.1) did not indicate the presence of any PCP dehalogenase or PCP-4-monooxygense, therefore, the database of *Bt*AOA genome was searched for the presence of Phe4MO. The search results indicated the presence of an aromatic amino acid hydrolase and was considered as Phe4MO in *Bt*AOA genome (GeneBank accession no: QIE38732.1). To explore the possible role of Phe4MO in PCP degradation pathway in *Bt*AOA, the gene was cloned; the recombinant protein was overexpressed, purified, and characterized. Further, the proper identity of the purified enzyme was confirmed using the bioinformatics tools and homology modelling. Hence, this study reports on the cloning, overexpression, characterization, and structural homology modelling of PCP hydroxylating Phe4MO from an indigenous strain *Bt*AOA.

5.2 Materials and methods

5.2.1 Materials

Pure pentachlorophenol (PCP, 98%), FAD (≥95%), Na-azide (≥99%), IPTG (≥99%), NADH (≥97%), Luria Bertani (LB) agar and Broth (Vegitone), imidazole ACS (≥99%), TMS (≥98.5%), SDS, PMSF (≥98.5%), Na-ampicillin salt, MES (≥98%) and 2-Mercaptoethanol (≥99%) were purchased from Merck (NJ, USA). Molecular biology grade DTT (≥99.5%), EDTA (99.4%), TEMED (≥99%), restriction endonuclease (*XhoI* and *BamHI*) FastDigest T4 DNA ligase, DNA markers, protein marker (10-250 kDa), Phusion high-fidelity DNA polymerase and PCR reaction mix were purchased from ThermoFisher Scientific (MA, USA). Bradford reagent and 40% acrylamides/bis-acrylamide solutions were purchased from Bio-Rad (Bio-Rad Laboratories, CA, USA). *Escherichia coli* strains DH5α and BL21 (DE3) (Invitrogen, ThermoFisher Scientific, MA, USA) were used as cloning and expression hosts, respectively. pET15b DNA (Novagen, Merck, NJ, USA) was used as a cloning and expression vector. Chemically competent cells of *E. coli* strains DH5α and BL21(DE3) were prepared as previously described [31]. All chemicals and reagents used in this study were of analytical grade standards. Stock solutions (10 mmol) of PCP was prepared in 0.5N NaOH and diluted with 50 mmol sodium phosphate (Na_2HPO_4-NaH_2PO_4) buffer, pH 7.0.

5.2.2 Isolation and identification of *Bacillus tropicus* strain AOA-CPS1 (*Bt*AOA)

*Bt*AOA was isolated from wastewater effluent, via culture enrichment and purified through successive sub-culturing on sterile nutrient agar plates until distinct colonies were obtained. Briefly, after three successive sub-culturing, 0.1 mL of the enriched culture was spread inoculated on minimal salt agar plates supplemented with 50 $mg \cdot L^{-1}$ of PCP. The plates were incubated at 30 °C until visible growths were observed. The isolates were purified via successive sub-culturing on sterile nutrient agar (NA) plates until distinct colonies were obtained. The isolated bacteria were individually screened in PCP-MSM, and strains with potential for PCP transformation were selected for further studies. The 16S rDNA fragment was amplified from the genomic DNA of the isolate via PCR using 63F- 5'-CAGGCCTAACACATGCAAGTC-3' and 1387R- 5'-GGGCGGTGTGTACAAGGC-3' universal primers pair [32], sequenced (Inqaba Biotech., Pretoria, South Africa) and submitted at NCBI BLASTn server [33] for the

identification of the pure culture. Further, the whole genome data was generated (Inqaba Biotech, Pretoria, South Africa) using a combination of Sequel II System, PacBio Single-Molecule Real-Time (SMRT Link Version 7.0.1.66975) sequencing technology (Moine -Scientist and Applications Support, 2019), FALCON assembler and Hierarchical Genome Assembly Process 4 (HGAP4) de novo assembly analysis application. This Whole Genome Shotgun project of *Bt*AOA has been deposited in GenBank under accession number CP049019 (version CP049019.1) (unpublished data).

5.2.3 Cloning of *cps*B

To amplify *cps*B from *Bt*AOA, PCR was performed using genomic DNA (purified using Quick-DNA Fungal/Bacteria Kit, Zymo Research Corporation, USA) of the isolate as a template and the primer pair 1 (forward: 5'-AGAATGACAAAGAAAACAGAAATT-3'), and (Reverse: 5'-AAATCAGTTAATCTTAGCATCATT-3'). To design the primer pair 1, the *Bacillus* species strains entries (from the NCBI BLASTn for 16S rDNA) with whole-genome sequencing projects (Fig. S2) were selected and genome sequences searched for the presence of phenylalanine 4-monooxygenase (Phe4MO). Whole-genome entries of *B. thuringiensis* strain ATCC 10792, *B. thuringiensis* strain L-7601, and *Bacillus* sp. FDAARGOS_235 that showed the presence of putative Phe4MO were retrieved, aligned using software DNAMAN (trail version, Lynnon corporation, CA, USA and used to design the primer pair 1. The PCR successfully amplified ≅1755 bp fragment and was sequenced. The DNA sequence was again submitted at NCBI BLASTn and showed a 100% homology with a 1755 bp nucleotide sequence with Phe4MO from many *Bacillus* spp.

The sequence search in GenBank accession number CP049019 showed 100% homology with gene sequence annotated as aromatic amino acid hydroxylase with an accession no: QIE38732.1 and locus tag: GM610_18120. The later phylogenetic analysis of the gene revealed that QIE38732.1 is a Phe4MO. Subsequently, the primer pair 2, forward: 5'-AGA<u>CTCGAG</u>ATGACAAAGAAAACAGAAATT-3' and reverse: 5'-AAA<u>GGATCC</u>TCAGTTAATCTTAGCATCATT-3', targeting expression vector pET15b for cloning purpose were synthesized to amplify the full-length gene. Underlined are the restriction site sequences for restriction enzymes *Xho*I and *Bam*HI. The gene was amplified in a 10 µL PCR

mix containing: 1 µL of 10X buffer, 1.5 mmol of $MgCl_2$, 20 µmol of dNTPs, 1.0 µmol of each primer (primer pair 2), 1 µL of genomic DNA, 1.25 U of high-fidelity DNA polymerase and 5.15 µL of ddH_2O using the following amplification conditions: initial denaturation at 95°C for 5 min, followed by denaturation at 95°C for 1 min, annealing at 58°C for 1 min, 25 cycles; extension at 72°C for 1 min, and a final extension at 72°C for 10 min.

The PCR product of the amplified *cps*B was extracted from agarose gel using a gene gel purification kit. The purified gene was sequenced and confirmed to be the expected product. Thereafter, the pET15b plasmid DNA and the gel-purified *cps*B gene were double digested using *Xho*I and *Bam*HI FastDigest restriction enzymes. The digested pET15b plasmid DNA and *cps*B fragment were gel purified and ligated using T4-DNA Ligase. The recombinant plasmid (pET15b-*cps*B) was transformed into chemically competent *E. coli* DH5α by heat shock technique [31], positive clones were selected on LB agar ampicillin (100 $\mu g \cdot mL^{-1}$) plates followed by colony PCR to select for positive transformants. The recombinant pET15b-*cps*B plasmid was then purified using a plasmid miniprep kit and again transformed into *E. coli* BL21(DE3) expression host followed by colony PCR to confirm the transformation. The presence of *cps*B in the recombinant pET15b-*cps*B plasmid was confirmed by restriction digestion using the corresponding restriction enzymes. The in-frame cloning of *cps*B and the correct sequence was confirmed by sequencing, using the universal T7 promoter and T7 terminator primers.

5.2.4 Overexpression and purification of CpsB

The transformed *E. coli* BL21(DE3) harbouring pET15b-*cps*B was pre-grown overnight at 37°C in 100 mL of LB broth containing ampicillin (100 $\mu g \cdot mL^{-1}$) followed by inoculation in 2 L of LB broth containing ampicillin (100 $\mu g \cdot mL^{-1}$) and incubated at 37°C to an optical density (OD) of 0.6 at 600 nm. The culture broth was induced with 1 mmol IPTG and incubated at 20°C for 24 h, with shaking at 200 rpm. The cell pellets were harvested by centrifugation at 8,000 rpm for 10 min and the supernatant was decanted. The pellets were re-suspended in 50 mmol NaH_2PO_4-Na_2HPO_4 buffer (pH 7.0) containing 1 mmol each of DTT and PMSF, washed twice and re-suspended in 100 mL of the same buffer. The re-suspended cells were sonicated at Psi 40 and 40 V for 5 min with 10 seconds pulse (Omni Sonic Ruptor 400, Omni International, Kennesaw,

GA). The lysate was centrifuged for 30 min at 15000 rpm and 4°C. The supernatant was used for purification of CpsB. The total protein concentration in the supernatant was determined by the Bradford method [34].

To purify the recombinant 6 × His-tagged CpsB in single step, 5 mL (≅25 mg total protein) of supernatant was loaded in 5 mL HisPur Cobalt Chromatography Cartridge (#90094, Thermo Scientific, IL, USA) connected to ÄKTA purifier system (GE Healthcare Life Sciences, IL, USA) and equilibrated with buffer 'A' (50 mmol NaH_2PO_4-Na_2HPO_4, 300 mmol NaCl, 5 mmol imidazole, pH 7.4) followed by elution with buffer 'B' (50 mmol NaH_2PO_4-Na_2HPO_4, 300 mmol NaCl, 150 mmol imidazole, pH 7.4). The elution fractions showing a single protein peak of interest at wavelength 280 nm were pulled together and desalted using 30 kDa cut-off Amicon Ultra-15 centrifugal filter unit (Merck, NJ, USA). The desalted protein was treated with thrombin to remove 6 × His-tag using Thrombin Cleavage Capture Kit (Cat. No. #69022, Merck Millipore, NJ, USA). The expression of CpsB, it's level in cell lysate, homogeneity of the purified protein and molecular weight was determined by loading the samples in 12% SDS-PAGE [35]. Enzyme activity of CpsB at each purification steps were determined as described below.

5.2.5 Enzyme activity assay

Enzyme activity for CpsB was evaluated in a 1 mL reaction mixture by continuous spectrophotometric method (Hlouchova et al., 2012), with some modifications. The 1 mL reaction mix contained 100 µmol PCP, 160 µmol NADH, 6.0 mmol 2-Mercaptoethanol and 300 µg (4.7 µmol final concentration) CpsB in 50 mmol NaH_2PO_4-Na_2HPO_4 buffer. 2-Mercaptoethanol was added to the reaction of PCP with CpsB to trap 1,4-tetrachlorobenzoquinone (Tet-CBQ) formed as 2,3,5,6-tetrakis[(2-hydroxyethyl)thio]-1,4-hydroquinone (THTH). Transformation of PCP, oxidation of NADH and formation of THTH were monitored spectrophotometrically at 320 nm, 340 nm and 350 nm, respectively. The NADH was included in the blank reactions to prevent interference with the absorption peak of PCP. The absorbance was converted to molar concentrations of PCP, NADH and THTH by using the molar extinction coefficients of 935 $mol^{-1}\cdot cm^{-1}$ [36], 6200 $mol^{-1}\cdot cm^{-1}$ and 2175 $mol^{-1}\cdot cm^{-1}$ [18], respectively.

5.2.6 Determination of optimum pH and pH stability

To determine the optimum pH for the CpsB activity, 146 µL (300 µg, 4.7 µmol final concentration) of the purified enzyme was assayed as described above except that the reactions were performed using different pH buffers i.e., sodium acetate-acetic acid (pH 4.0-5.0), NaH_2PO_4-Na_2HPO_4 (pH 6.0-8.0), glycine-NaOH (pH 9.0-10.0), Na_2HPO_4-NaOH (pH 11-12) and KCl-NaOH (pH 13.0). To determine the stability of purified CpsB at different pH, the enzyme (300 µg, 4.7 µmol final concentration) was pre-incubated in buffers ranging from pH 7-13, aliquots were withdrawn at 15 min intervals and the % residual activity was determined as described above. The activity obtained at optimum pH 7.0 was considered as 100% residual activity.

5.2.7 Determination of optimum temperature and temperature stability

To determine the optimum temperature for the CpsB activity, 146 µL (300 µg, 4.7 µmol final concentration) of the purified enzyme was assayed as described above except that the reactions were performed at different temperatures ranges from 25-60°C at the optimal pH 7.0. The thermal stability of the enzyme was investigated by incubating the enzyme (300 µg, 4.7 µmol final concentration) at different temperatures (25, 30, 40, 50 and 60ºC) for 180 min (aliquots was taken out after every 15 min) and the % residual activity was determined at optimum conditions as described above. The residual CpsB activity obtained at optimum conditions was considered as 100% residual activity.

5.2.8 Determination of kinetic parameters

To determine the kinetic parameters for the enzymatic reaction of CpsB with substrate PCP, the purified CpsB (300 µg, 4.7 µmol final concentration) was incubated with initial concentrations of PCP (20, 40, 60, 80, 100, 120, 140, 160, 180 and 200 µmol) in the presence of 160 µmol NADH and the activity was determined at optimum conditions as described above. The purified CpsB (300 µg, 4.7 µmol final concentration) enzyme activity was also determined by incubating with varying concentrations of NADH (50, 100, 150, 200 and 250 µmol) in the presence of 150 µmol PCP and the activity was determined at optimum conditions as described above. The K_m and v_{max} values were obtained by applying the Michaelis-Menten equation in Line-weaver Burk

plots. The catalytic constant of the enzyme-substrate reaction (k_{cat}), also referred to as the turnover number, represents the number of reactions catalysed per unit time by each active site; was determined using the equation, $k_{cat} = v_{max}$ /[E]t, where v_{max} is the maximum velocity and [E]t is the total enzyme concentration. Catalytic efficiency of the enzyme was calculated by the equation, k_{cat}/K_m. The Hill coefficient n_H, as a measure of homotropic cooperative effects, was estimated by applying Hill approximation equation [37–39]. The allosteric properties of the enzyme were analysed by M-W-C modal as described previously [40,41]. The kinetics properties, the micro-constants such as dissociation constant (K_S), apparent first order rate constant (k_{obs}) and specific rate constants (k_1, k_2, k_3, k_4, k^*_1, k^*_2) for the enzyme-substrate interaction were determined as described previously [42].

5.2.9 Effects of metal ions and inhibitors on CpsB activity

To determine the effect of different concentrations of metal ions and inhibitors, (300 µg, 4.7 µmol final concentration) of purified CpsB was pre-incubated with the metal ions and different concentrations of inhibitors for 10 min and assays were performed as described above. The enzyme activity determined without the addition of any metal ion or inhibitor was considered as 100% residual activity [43].

5.2.10 In-gel trypsin digestion and identification of the purified CpsB

The pure protein (50 µg) was loaded onto 12% SDS-PAGE and stained with Coomassie blue R250. The protein band was excised carefully and digested with trypsin and fragments analysed by electrospray mass spectrometry (ES-MS) (Central Analytical Facility, Stellenbosch University, Stellenbosch, South Africa). Data analysis and protein identification were done as described previously [44] and detailed in the supplementary material (Fig. S3).

5.2.11 Template-based structure prediction and homology modelling

Three-dimensional structure and homology modelling of CpsB were predicted by submitting the amino acid sequence at the SWISS-MODEL tool at *ExPASy* bioinformatics resource portal workspace (https://swissmodel.expasy.org/interactive). The default parameters used for performing the automated SWISS-MODEL were as explained previously [45,46] and elaborated

at (http://swissmodel.expasy.org/docs/help) webpage. The modelled PDB files were submitted to online tool PDBsum to determine structural summary [47].

5.2.12 Evolutionary relationships of CpsB with other monooxygenases

To study the evolutionary relatedness of CpsB with other Phe4MO and PCP 4–monooxygenases, a phylogenetic based evolutionary analysis was performed using the Neighbour-Joining method [48]. The evolutionary history, bootstrap consensus tree, percentage of replicate trees [49] and the evolutionary distance tree [50] were performed using MEGA7 software [51].

5.2.13 The nucleotide sequence submission:

This Whole Genome Shotgun project of *Bacillus tropicus* strain AOA-CPS1 has been deposited in GenBank under accession number CP049019 (version CP049019.1) (unpublished data). The putative aromatic amino acid hydroxylase gene sequence was submitted with an accession no: QIE38732.1 and locus tag: GM610_18120.

5.3 Results and discussion

5.3.1 Identification of *Bacillus tropicus* strain AOA-CPS1 (*Bt*AOA)

Initially, the isolate was identified as *Bacillus cereus* strain AOA-CPS1 (*Bc*AOA) based on the 16S rDNA sequence analysis (submitted to NCBI as accession number MH504118.1). However, a quality control test by NCBI for the submitted whole genome sequence of the strain, using an average nucleotide identity (ANI), which compares the submitted genome sequence against the whole genomes of the type strains that are already in GenBank [52,53], resulted in the renaming of *Bc*AOA as *Bacillus tropicus* strain AOA-CPS1 (*Bt*AOA). The ANI analysis indicated that the genome sequences of *Bc*AOA are 96.61% identical to the genome of the type strain of *Bacillus tropicus*, with 89.9% coverage of the genome. Consequently, *Bc*AOA was renamed as *Bt*AOA (based on the whole genome data submitted at NCBI under accession number CP049019).

5.3.2 Cloning, overexpression, and purification of 6 × His-tagged CpsB

A 1.755 kb gene fragment (*cps*B) encoding aromatic amino acid hydroxylase Phe4MO was detected by PCR (Fig. 1A) and was successfully cloned into an expression vector (pET15b) and

confirmed by restriction digestion (Fig. 1B, Fig. S4). The recombinant C-terminal 6 × His-tagged CpsB was overexpressed in *E. coli* BL21(DE3), purified to homogeneity and showed a single band of ≅64 kDa (Fig. 1C). The molecular weight of CpsB match with that of previously reported Phe4MO from other bacteria and fungi [23,54] and in variant to that of mammalian [24]. The mammalian Phe4MO's is mostly homotetramer due to the complex regulatory and activation states of human Phe4MO [55]. Contrary to the mammalian Phe4MO, bacteria Phe4MO are mostly monomeric protein, consisting of only one domain corresponding to the catalytic domains of aromatic amino acid hydroxylase [56].

Furthermore, genomic organization of Phe4MO from animals, insects, fungi, birds, mosquitoes, bee, and other groups of organisms showed a high variability in Phe4MO exon-intron structure [57]. The molecular weight of CpsB is also similar to Phe4MO [58], PcpB [17,18,59], 2,4,6-trichlorophenol 4-monooxygenases (TCP4MO) [60,61], chlorophenol 4-monooxygenase (PhCl4MO) [62] and phenol hydroxylases (PhOH) [44,63–65], but higher than that of para-nitrophenol 4-monooxygenase [66]. A single-step purification scheme purified CpsB 1.03-fold with a total yield of 82.1% (Table 1). 1.03-fold purification is low since over 80% of the total protein expressed in crude extract is CpsB. The purification table also indicated the total and specific activity of CpsB at each purification step.

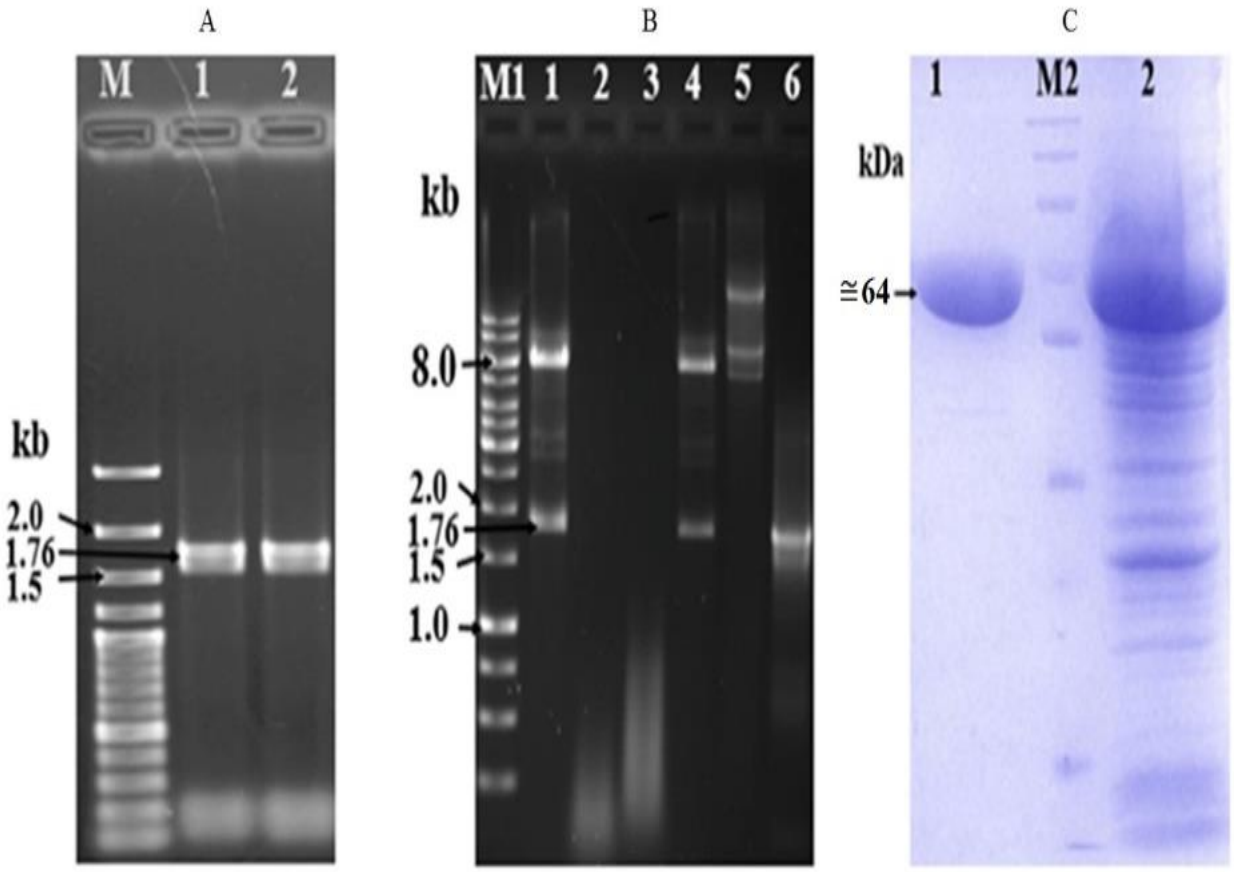


Fig. 1: PCR amplification, restriction digestion, overexpression, and purification of CpsB. (A), Amplification of cpsB gene from *Bt*AOA; Fermentas 100 bp plus DNA marker (Lane M), amplified cpsB gene (Lane 1 and 2); **(B),** Agarose gel of double digested recombinant *pET15*-cpsB plasmid vector extracted from *E. coli* BL21(DE3) transformants; Fermentas 1 kb DNA marker (Lane M1), positive *pET15b*-cpsB vector with the excised 1.76 kb cpsB fragment of cpsB after double restriction digestion with *XhoI* and *BamHI* (Lane 1 and 4), 7.464 kb (Fig. S4) undigested pET15b-cpsB plasmid vector (Lane 5), control experiment showing a 1.76 kb fragment of cpsB amplified using *AOA-CPS1* DNA as template; **(C)**, 12% SDS-PAGE showing overexpressed CpsB; 10–250 kDa protein marker (Lane M2), crude CpsB (Lane 2), ÄKTA fractions of purified recombinant 6xHis-tagged CpsB (Lanes 1). ≅64 kDa molecular weight indicated is based on the biophysical properties calculated at ProtPram tool at ExPASy for CpsB amino acid sequence.

Table 1: Purification scheme for His-tagged phenylalanine 4-monooxygenase (CpsB) from *E. coli* BL21(DE3).

Purification step	Total Protein (mg)	Protein ($mg \cdot mL^{-1}$)	Total activity (U*)	Total Activity ($U \cdot mL^{-1}$)	Specific activity ($U \cdot mg^{-1}$)	Purification (fold)	Yield (%)
Crude(cell lysate)	496.8 [a]	4.968	12.09[c]	201.58	40.57	1.0	100
HisPur Cobalt column	20.53[b]	2.053	12.61[d]	86.39	42.08	1.03	82.1[e]

[a]From 100 mL of crude *E. coli* cell lysate. [b]Total protein in 10 mL of eluted fractions pooled together after loading 25 mg of total protein in the column. [c] in ≅60 µL (300 µg total protein) cell lysate. [d] in ≅146 µL fraction (300 µg total protein) showing single band on SDS-PAGE (Fig 1c, Lane 1). [e]Yield (%) = (20.53 mg / 25 mg) × 100. * One unit of CpsB activity was defined as the 1 µmole of THTH produced per min under standard assay conditions.

5.3.3 Optimum pH and pH stability of CpsB

CpsB showed an optimum pH at 7.0 (Fig. 2A) similar to those of *S. chlorophenolicum* PcpB [67], 2,4,6-trichlorophenol 4-monooxygenase (TCP4MO) [60,61], phenol hydroxylase (PhOH) [65] and PhCl4MO [62] but different from that of fungal Phe4MO [54,68], and PhOH from *Bacillus* sp. [64], with the optimum pH of 7.5 and 7.8, respectively. CpsB activity at pH 7.5 was slightly lower than that at pH 7.0 and drastically decreases at pH 8.0 in accordance to previous reports [23,54,61]. CpsB retained about 10% of its activity at pH 11.5 and no activity at pH 12 and 13 (Fig. 2A). CpsB was stable at slightly acidic and neutral pH, retaining >95% of its activity between pH 6.0-7.0 and over 60% activity at pH 9.0 relative to activity at pH 7.0 (Fig. 2B). This corroborates earlier report that more than 80% of PCP was used as wood preservation in combination with sodium hydroxide, which can turn most PCP-polluted sites into alkaline environments [69]. In addition, PCP is mostly dissolved in sodium hydroxide for microbial degradation analysis since the compound is sparingly soluble in water. This also strengthens the fact that bacteria PCP transformation occurred more rapidly at neutral to alkaline conditions than acidic conditions [70].

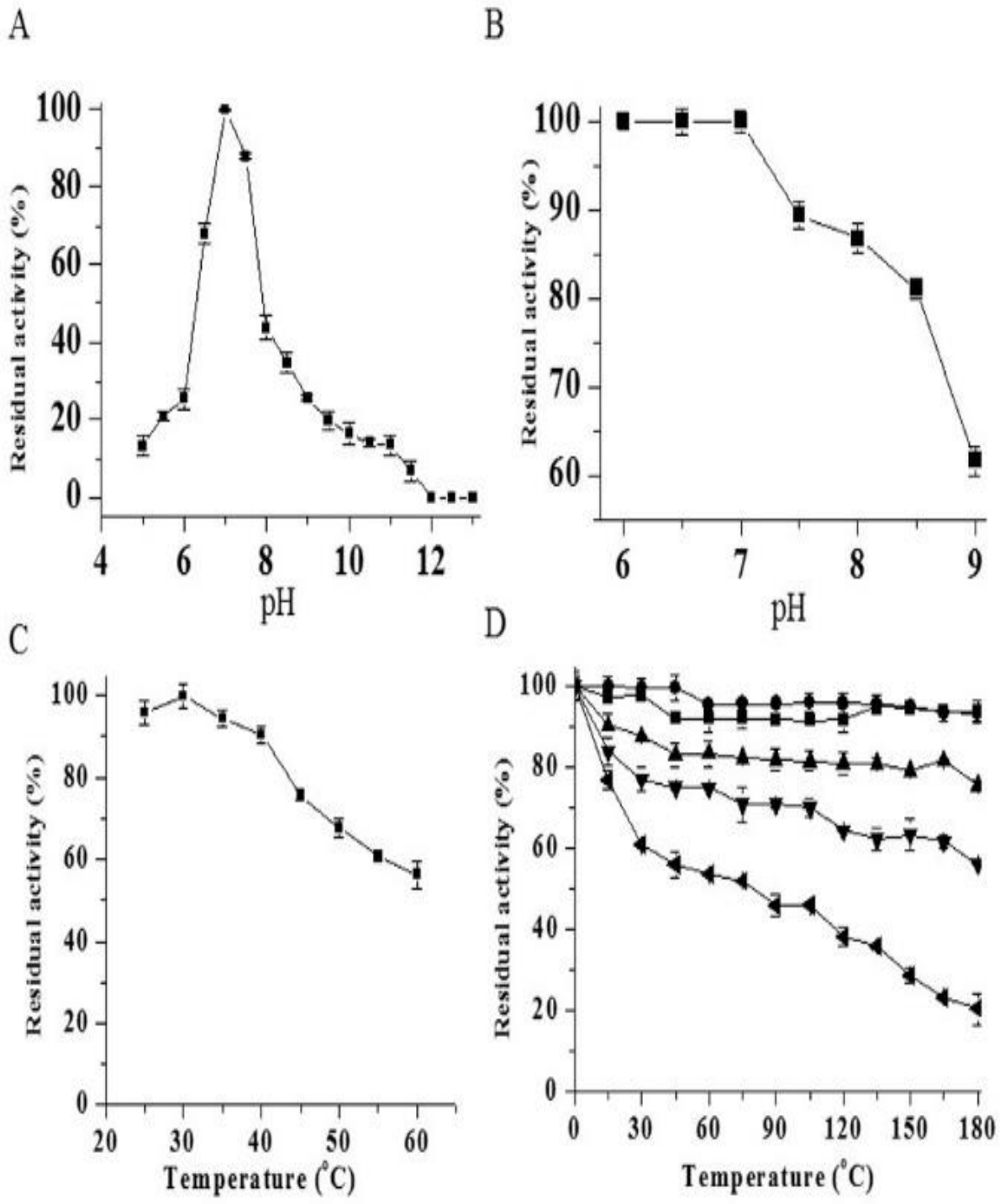


Fig. 2: Characterization of CpsB. (A), Optimum pH; **(B)**, pH stability at different pH range for 180 min; **(C),** optimum temperature; **(D),** temperature stability at 25 ºC (■), 30 ºC (●), 40 ºC (▲), 50 ºC (▼) and 60 ºC (◄) over time.

5.3.4 Optimum temperature and thermal stability of CpsB

CpsB showed optimum activity at 30°C (Fig. 2C) similar to chlorophenol 4-monooxygenases [71], PhOH [65] and TCP4MO [61], but, contrary to the optimum temperature of 37°C and 40°C reported for the fungal Phe4MO [23,54] and 35°C reported for PcpB [67], respectively. CpsB activity decreased sharply at >40°C with about 60% residual activity at 60°C, also in concordance to Phe4MO from *Mortierella alpina* ATCC 32222 [23]. Moreover, CpsB showed good stability between 25°C-30°C, retaining >90% activity and about 80% activity at 40°C, after 180 min of incubation (Fig 2D). CpsB activity decreased to about 65% and 20% at 50°C and 60°C, respectively after 180 min and show no activity at 70°C (data not shown). Fungal Phe4MO's from *Mortierella alpina* ATCC 32222 [23] and *Dictyostelium discoideum* [54] also exhibited similar property in their temperature stability. Based on the optimum temperature of CpsB, the enzyme can perform maximally if used in wastewater remediation without incurring additional cost on mimicking environmental temperature to suit the enzyme. In addition, CpsB maintained about 50% of its activity at 50°C similar to that of phenol hydroxylase from *B. thermoglucosidasius* strain A7 [64]. The complete loss of activity at high temperature is not unexpected based on the source of the isolate. Inactivation of phenol hydroxylase and other chlorophenol monooxygenases isolated from water samples at greater than environmental temperature (25°C) has also been reported [65,72].

5.3.5 Effect of metal ions and inhibitors on CpsB activity

CpsB exhibited 63.3% and 55.6% increased activity in the presence of Fe^{2+} and Ca^{2+} ions, respectively compared to activity in the absence of metal ion (Table 2), suggesting that CpsB is Fe^{2+} and Ca^{2+} dependent. Previous studies have also reported an increase in Phe4MO activity in the presence of Fe^{2+} [73,74]. Phe4MO has also been reported to be Fe^{2+} dependent [55,75], with Fe^{2+} reported to stimulate PcpB activity [60,76]. The other metal ions (Zn^{2+}, Mn^{2+}, Fe^{3+}, Cu^{2+}, Mg^{2+}, Pb^{2+}, Hg^{2+} and Ni^{2+}) tested could not enhance CpsB activity. Metal ions are present at the active sites of some enzymes to coordinate and reduce dioxygen (O_2) to electrophilic activated oxygen species to attack substrate(s) carbon-hydrogen bond [77]. In a similar study [55], Phe4MO was found to be active only in the presence of Fe^{2+} while other metal ions could not sustain hydroxylation of substrate. Structural analysis of Phe4MO has shown the presence of iron and some other metal ions in the catalytic active site of Phe4MO and it is crucial for the

functionality of the enzyme [78,79]. Comparison of human and bacteria iron dependent Phe4MO's revealed similar conformation but different reactions rates and stability [55].

Table 2: Effects of metal ions on CpsB activity

Metal ions	Residual activity (%)
*Control	100
Fe^{2+}	163.32 ± 2.85
Ca^{2+}	155.55 ± 2.13
Cu^{2+}	0
Mg^{2+}	36.62 ± 1.39
Pb^{2+}	0
Ni^{2+}	0
Co^{2+}	83.23 ± 2.94
Mn^{2+}	60.28 ± 3.38
Na^{2+}	23.48 ± 3.03
Hg^{2+}	0
Fe^{3+}	37.08 ± 2.36

*Reaction carried out without metal ion

CpsB exhibited about 3-fold, 2.4-fold, and 1.3-fold increased activity in the presence of 0.5 mmol, 1.0 mmol and 1.5 mmol of EDTA, respectively (Table 3). The maximum activity observed in the presence of 0.5 mmol EDTA corroborates previous reports [65]. However, about 48.6% reduction in CpsB activity was obtained in the presence of 2.0 mmol EDTA (Table 3). Furthermore, 2.7-fold and 2.1-fold increase in CpsB activity was obtained in the presence of 0.5 mmol DTT and 0.5 mmol SDS, respectively, relative to activity in the absence of inhibitors. About 60% residual activity of CpsB was inhibited in the presence of 1.0 mmol DTT and complete inhibition in the presence of 1.5 mmol and 2.0 mmol DTT as well as 2 mmol SDS. Moreover, about 28% activity of CpsB was inhibited in the presence of 0.5 mmol Na-azide while activity was completely inhibited in the presence of 1.0 mmol, 1.5 mmol and 2.0 mmol Na-azide. Complete inhibition of CpsB activity was observed in the presence of Tet-CBQ at all concentrations tested similar to the previous report [80].

Table 3: Effects of inhibitors at different concentrations on CpsB activity.

Inhibitor (mmol)	*Control	EDTA	DTT	Na-Azide	Tet-CBQ	SDS
0.5	100	297.42±3.45	270.62±3.02	72.49±3.01	0	211.33
1.0	100	239.3±3.82	40.34±1.92	0	0	29.67
1.5	100	131.82±2.94	0	0	0	7.95
2.0	100	51.38±4.02	0	0	0	0

*Reaction carried out without inhibitor

5.3.6 Cofactor requirement and reaction stoichiometry

Effects of cofactors (NADH and FAD) on CpsB activity in the presence of PCP was evaluated to check if CpsB prefers NADH or FAD as a cofactor. CpsB did not show any activity in the presence of FAD (data not shown). The enzyme transformed 100 µmol PCP to THTH in the presence of 160 µmol NADH within 30 min (Fig. 3A, 3B). The initial velocity of 0.06 s^{-1} PCP transformation by CpsB to THTH in the presence of NADH indicated the reaction was very fast. The rate of increase in absorbance at wavelength 350 nm (formation of product THTH) was 0.054 s^{-1} while the rate of increase in absorbance at wavelength 340 nm (representing NADH oxidation) was 0.064 s^{-1} (Fig. 3B).

The 100 µmol of PCP transformed by CpsB yielded 96.736 µmol of the product with the consumption of 108.186 µmol NADH. The 100 µmol PCP was completely transformed within 28 min, however, 96.736 µmol of the product was obtained at the end of 30 min with the graph depicting that NADH reduction remained at a plateau after 28 min. Based on PCP and NADH consumption rates and product formation, there was a delay in the rate of the release of the product. Also, there might be excess of NADH at the cofactor binding site of the enzyme. This indicated that the rate of PCP and NADH consumption and product formation were stoichiometrically equal at the ratio of 1:1 for PCP: NADH.

Previously, flavin-dependent, non-heme iron, heme-containing cytochrome p450 and pterin-dependent enzymes were reported to catalyse aromatic compounds monooxygenation/hydroxylation [81]. These categories of enzymes insert one oxygen atom from molecular oxygen (O_2) into the aromatic organic compounds (to overcome their spin-forbidden reactions with the substrates) and most of these enzymes require cofactors (such as NADH/NADPH and FAD) as an electron donor and to transfer electrons to molecular oxygen for its activation [82]. However, unlike flavin-dependent PcpB that require NADH or NADPH to generate a reactive flavin intermediate C4α-hydroperoxyflavin (which often resulted into an unproductive consumption of NADH and/or futile cycle) require for PCP hydroxylation [56], but CpsB does not require FAD as a cofactor as confirmed by no absorption peak was observed at wavelength 450 nm when CpsB was scanned over a wide range of wavelength (Fig 3A).

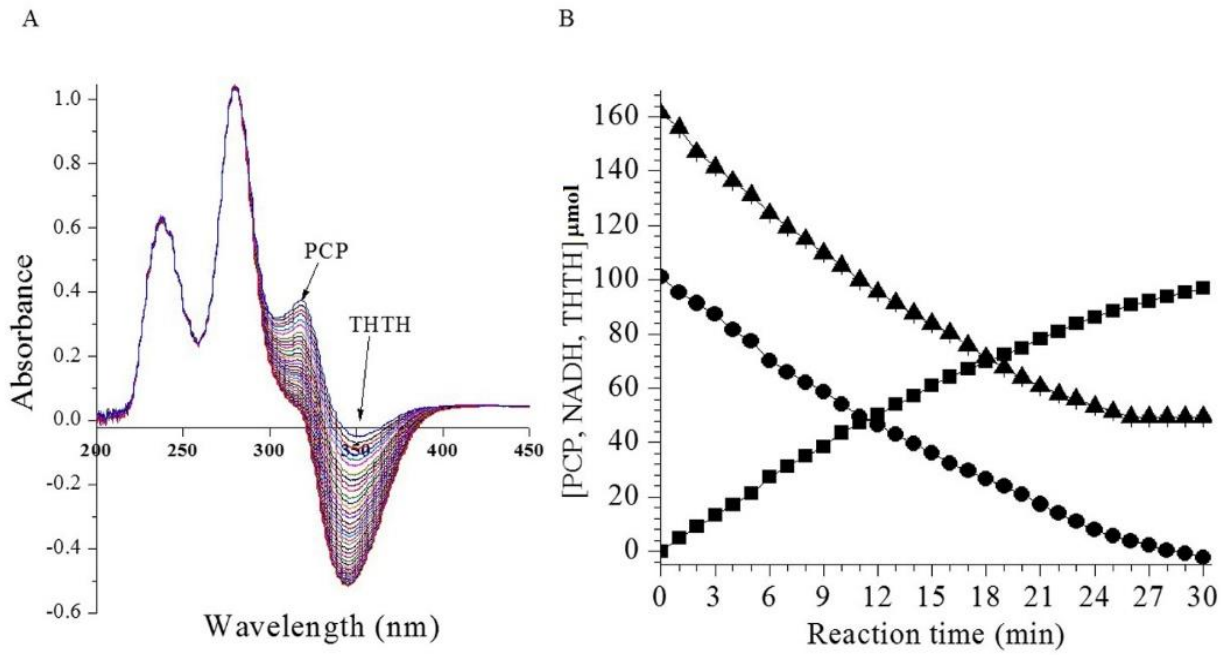


Fig. 3: Continuous spectrophotometric reaction of CpsB (300 µg) with PCP (100 µM) in the presence of NADH (160 µM) and the reaction stoichiometry. (A), continuous spectrophotometric reaction of CpsB (300 µg) with PCP (100 µmol) and NADH (160 µmol; **(B)**, stoichiometry of PCP (100 µmol) transformation at 320 nm (●), product accumulation at 350 nm (■) and reduction of 160 µmol NADH at 340 nm (▲) by CpsB (300 µg) in a continuous spectrophotometric reaction.

5.3.7 Steady-states enzyme kinetic properties of CpsB

The enzyme substrate reaction exhibited pre-steady state at low PCP concentrations (20-100 µmol) but showed a steady state at high PCP concentrations (100-200 µmol) exhibiting the S-shaped curve (Fig. 4A). The double reciprocal diagram according to Lineweaver Burk plot is not linear but fits a parabola. It becomes linear, however, if $1/v$ is plotted against $1/[PCP]^2$ in a specific substrate concentration range (80-200 µmol) resulting in calculated v_{max}, K_m, k_{cat} and k_{cat}/K_m values of 0.465 µmol·s^{-1}, 140 µmol, 0.099 s^{-1} and 7.07 x 10^{-4} µmol^{-1} s^{-1}, respectively (Fig. 4B). The double reciprocal diagram according to Lineweaver and Burk for $1/v$ plotted against 1/[NADH] was fit linear exhibiting the v_{max}, K_m, k_{cat} and k_{cat}/K_m values of 0.259 µmol·s^{-1}, 224 µmol, 0.055 s^{-1} and 2.47 x 10^{-4} µmol^{-1}·s^{-1}, respectively (Fig. 4C). The results indicate that optimum PCP and NADH concentrations for steady state kinetics of PcpB substrate reaction must be above 200 µmol for each. The reported K_m and k_{cat}/K_m for PCP oxidation by the wild type PcpB [83], were lower than those of CpsB. However, the conditions used were not steady state and the concentrations of the substrate was not in excess of the PcpB, which could raise suspicion over the reliability of the kinetic parameters obtained [67]. Furthermore, the K_m, k_{cat} and k_{cat}/K_m of the recombinant CpsB were higher than those of a recombinant PcpB-catalysed oxidation of PCP [84]. Also, the k_{cat} and k_{cat}/K_m of CpsB were higher than those of the wild-type and recombinant PcpB [67], indicating that CpsB is catalytically more efficient than both the wild-type and recombinant PcpB. Plotting the enzyme substrate and enzyme co-substrate data as Hill plots provided additional useful information. The Hill plot (Fig. 4D) was found to be nonlinear over the range of substrate concentrations studied, showing two prominent straight lines, one from the substrate range of 20-100 µmol (Fig. 4D) and second from 100-200 µmol (Fig. 4E). The plots were used to calculate $K_{0\,5}$, considered as K_m [85] and showed that the enzyme has two different K_m values i.e. 6.33 µmol (4D) at low substrate concentrations (20-100 µmol) and 114.55 µmol (4E) at high substrate concentrations (100-200 µmol). The results indicate that CpsB has low binding affinity at low substrate range, but suddenly show burst in product formation due to the high binding affinity at high substrate concentrations, indicating primarily its allosteric nature and has homotropic cooperative interaction effect. The Hill plot (Fig. 4F) for NADH interaction with CpsB suggests equal affinity at different reaction conditions and possibility of heterotropic cooperative interaction effects.

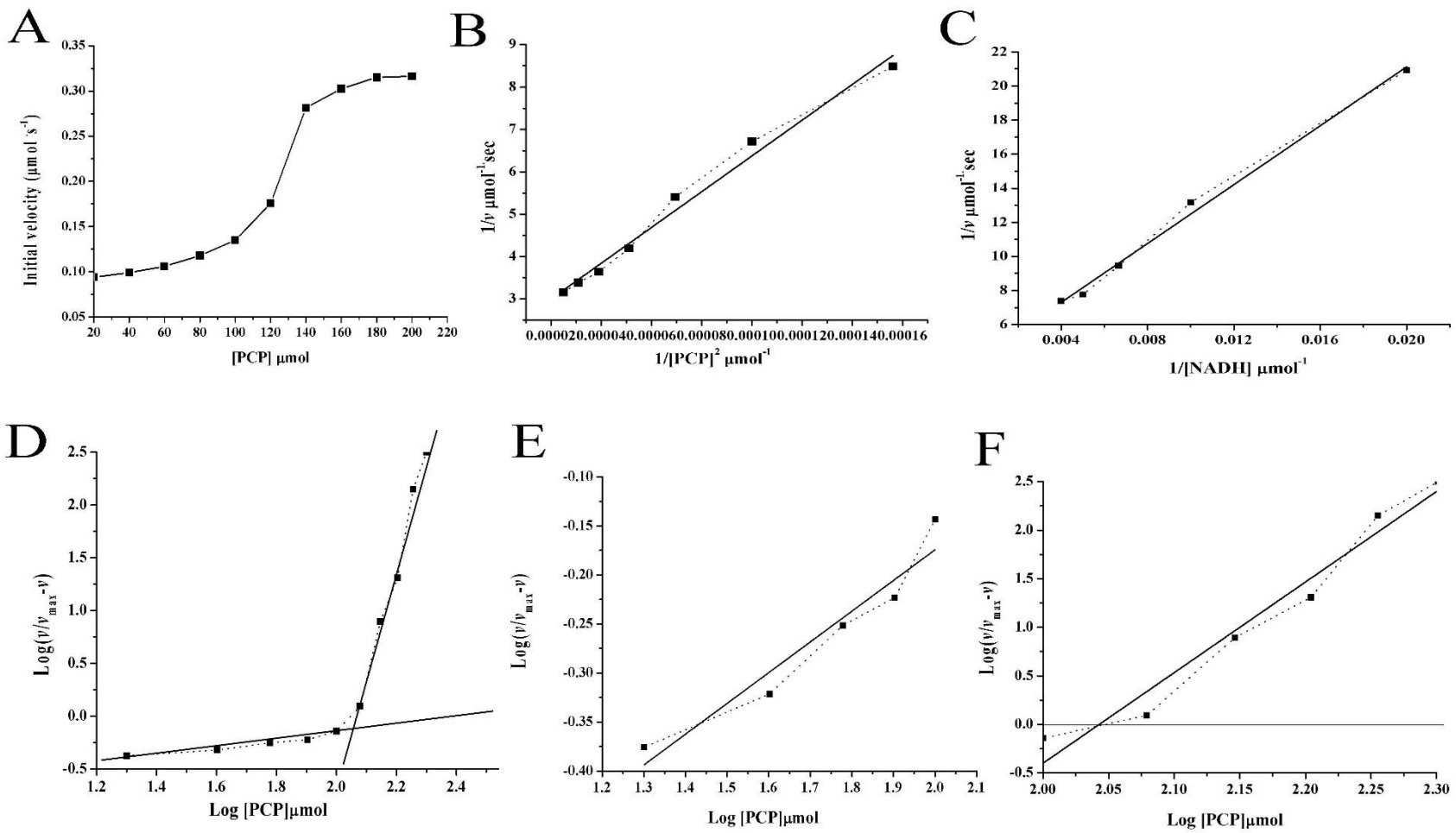


Fig. 4: Enzyme reaction kinetics: (A), Initial velocity of CpsB (150 µg) reaction with PCP at different initial concentration (20-200 µmol) in the presence of 160 µmol NADH**. (B)**, The double reciprocal plot of data shown in 4A, except $[PCP]^2$. **(C)** the double reciprocal plot of initial velocity of CpsB (150 µg) reaction with PCP (200 µmol) in presence of different NADH initial concentration (50, 100, 150, 200 and 250 µmol). **(D)**, Hill plot for the data shown in 4A. **(E)**, Hill plot for the data shown in 4A, but only for the PCP at low [PCP] (20-100 µmol). **(F)**, Hill plot for the data shown in 4A, but only at high [PCP] 100-200 µmol.

Two reaction mechanisms (mechanism I and mechanism II) may be used to further discuss the nature of the steady state kinetics of CpsB and PCP interaction. Analysis of both mechanisms by the steady state method shows that, in general, simple behaviour is not to be found. However, in both cases, the assumption that small k_4 value, but >0, leads to the observed hyperbolic dependence of k_{obs} versus [S] (Fig. 5A) or l/k_{obs} versus l/[S] (Fig. 5B).

However, initial velocity vs [S] plot (Fig. 4A) indicated the possibility of mechanism I, where the limiting rate at high [S] observed is k_1, the isomerization of E to E*, as this process is independent of the nature of substrate. The isomerization criteria in Mechanism I is supported by the reduction of cytochrome c which requires the initial cleavage of the Fe^{3+}-methionine bond [86], and also in the determination of the rate constant for the conversion of $S_2O_4^-$ to SO_2^- [87]. The second, if the equilibrium concentration of E* in Mechanism 1 is >l0% of E, then addition of a saturating [S] agent will lead to a burst in the production of P followed by the remainder of the reaction proceeding at the rate of k_1.

The data (Fig. 4A) shows a lag phase where E may be isomerizing in E* at low substrate concentrations (20-100 µmol), then the reaction led to the burst in the production of P i.e. at [S] above 100 µmol. Further, the experiments involving the spectrophotometric monitoring of the protein absorption in the presence of saturating substrate concentration (200 µmol) did not show any instantaneous change (till 30 s after adding the substrate) in spectrophotometric property of the protein (data not shown), indicating that E is isomerizing to E* before the conversion of E*S followed by slow conversion of E*S to P.

Although the possibility of mechanism I was established, it was necessary to test the mechanism II supported by experimental data (Fig. 5A-D). The steady state reaction kinetics in this case is evident by the hyperbolic dependence of apparent first order rate constant (k_{obs}) on substrate concentration [S]. The plot of k_{obs} versus [S] (Fig. 5A) or l/k_{obs} versus l/[S] (Fig. 5B) was found to be hyperbolic with a readily intercept on the ordinates. The l/k_{obs} versus l/[S] plot (Fig. 5C) at high [S] (100-200 µmol) show a linear behaviour and was used to calculate value of k_1 (9.5×10^4 $mol^{-1} \cdot s^{-1}$), k_2 (42.2 s^{-1}), k_3 (0.48 s^{-1}) and K_d (k_2/k_1=442 µmol) (the dissociation constant for the intermediate ES). The value of k_3 (1.14 s^{-1}), k_4 (0.91 s^{-1}) and K_d (473 µmol) were also calculated

from the plot k_{obs} versus [S] and $1/(k_{obs}-k_4)$ vs 1/[S] (Fig. 5D) plotted at high [S] (100-200 µmol) where the data showed linear behaviour.

Consequently, it is necessary to find out if the steady state behaviour of the reaction is simply a one-step reversible equilibrium (Mechanism III) or two-step reversal equilibrium (Mechanism II) (Fig 5B). In the case of a simple reversible equilibrium, k_{obs} is a linear function of [S] with slope k_1 and intercept at zero [S]= k_2 [88] (theoretical slop 1 in Fig. 5A). In contrast, such plots are nonlinear for the two-step Mechanism II as can be seen from steady state equation [42]: at high [S], the limiting rate is the sum of the rate constants of the second equilibrium, k_3+k_4, and at zero [S], the limiting rate is k_4 (experimentally determined slop 2 in Fig. 5A). This plot is very useful to differentiate between one- and two-step equilibrium processes, and when extrapolated, provides numerical values for k_2* or k_4, respectively.

It is established [42] that when equilibrium conditions apply, i.e. $k_2 >> k_3$, the semilogarithmic plots used to determine k_{obs} are perfectly linear from zero time of reaction, and reciprocal plots are linear over a wide range of [S]. On the other hand, when the steady state conditions apply, i.e. $k_3 >> k_2$, the primary semilogarithmic plots show significant lags before becoming linear. Contrarily, data from this study showed that the primary semilogarithmic plots are perfectly linear from zero time of reaction and showed no significant lags (data not shown due to complexity) but, the reciprocal plot (Fig. 5B) showed hyperbolic nature i.e. nonlinear.

Finally, for the experimental data shown in Fig 4A-C and 5A-D, mechanism IV is proposed for enzyme substrate reaction. The enzyme isomerized first upon binding with the co-substrate NADH and then lead to the formation of intermediate complex ES which further dissociated to form the product. It is adjudged that isomerized form E* of the enzyme is present in the complex and the nature of intermediate is E*S and not ES.

I $$E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} E^* \qquad E^*+S \underset{k_4}{\overset{k_3}{\rightleftharpoons}} P$$

II $$E+S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_4}{\overset{k_3}{\rightleftharpoons}} P$$

III $$E+S \underset{k_2^*}{\overset{k_1^*}{\rightleftharpoons}} P$$

IV $$E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} E^* \qquad E^*+S \underset{k_2^*}{\overset{k_1^*}{\rightleftharpoons}} E^*S \underset{k_4}{\overset{k_3}{\rightleftharpoons}} P$$

Fig. Mechanisms : The possible mechanisms to discuss the nature of steady state kinetics of CpsB and PCP interaction. E, free enzyme; S, substrate; ES, enzyme-substrate complex (intermediate); E*, isomerization for of enzyme; E*S, isomerization for of enzyme complex with substrate (intermediate); P, product that a formation rate (k_{obs}) that is experimentally measured. k_1, k_2, k_3, k_4, k^*_1, k^*_2 are the specific rate constants.

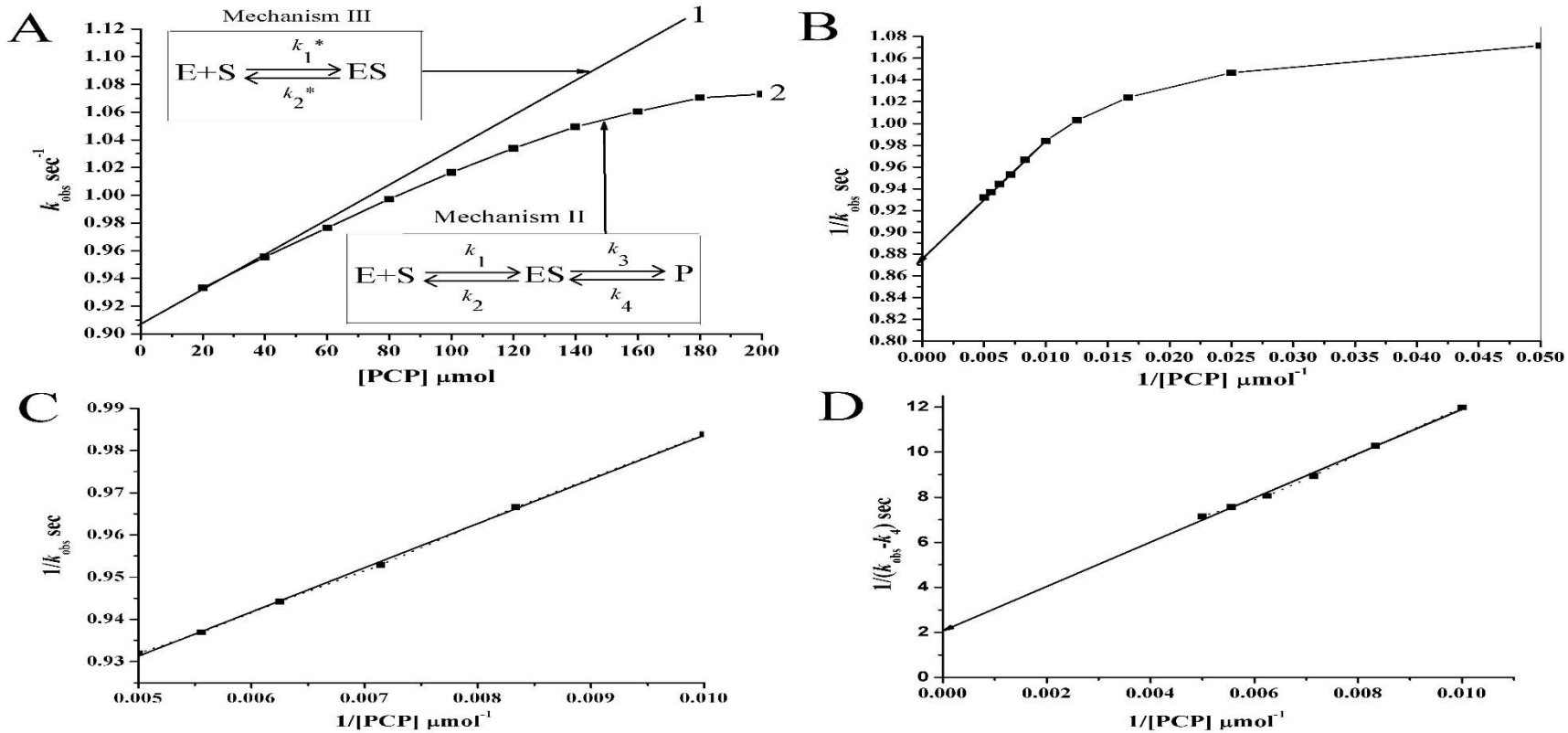


Fig. 5: Steady state enzyme kinetics: (A) the plot of k_{obs} versus [PCP] (20-200 µmol). The value of k_{obs} was calculated from semilogarithmic plots of CpsB (150 µg) reaction with PCP at different initial concentration (20-200 µmol) versus time obtained from continuous spectrophotometric methods over the time range of initial 10 min. The strait plot represents the theoretical Line (1) when the reaction is in a simple one step reversible equilibrium conditions (mechanism III). The curve (2) represent when the reaction is in a two-step reversible equilibrium condition (mechanism II). **(B),** Reciprocal plot of the observed first order rate constant (k_{obs}) as a function of [PCP], $1/k_{obs}=1/(k_3+k_4)$ when extrapolated at high [PCP], $1/k_{obs} =1/k_4$ when extrapolated at low [PCP]. **(C)** The $1/k_{obs}$ versus 1/[S] plot at high [PCP] (100-200 µmol) show a linear behave and was used to calculate value of k_1, k_2, k_3 and K_d (the dissociation constant for the intermediate ES). (D) $1/(k_{obs}-k_4)$ vs 1/[PCP] plotted at high [PCP] (100-200 µmol), used to calculate value of k_3 and K_d.

5.3.8 Allosteric properties of CpsB

The enzyme-substrate interaction plots (Fig. 4A and Hill plot form 4D) confirmed the allosteric nature of CpsB. The curve was significantly found to be sigmoidal shape indicating slow formation of product at first followed by a sudden increase before reaching saturation. The Hill plot for the co-substrate NADH was found to be linear over a wide range of concentration indicating that CpsB has equal affinity for NADH all the time. As previously reported, the Hill plot [37–39] is a useful tool for analysing some allosteric effects, and the values of the slopes obtained are considered to be a measure of the number of interacting sites as well as the strength of their interaction. When the interaction is very strong, the value of the slope (n_H) of the Hill plot approaches the number of binding sites for the ligand [85]. The results shown in Fig. 4E shows n_H=0.31, indicating all the binding sites are not occupied at low substrate concentrations, but n_H=9.33 indicating most of the binding sites are occupied and has homotropic substrate-substrate interaction. This also shows that CpsB may exists in multimeric form while reacting with the substrate. The value of n_H=1 from Fig. 4F indicate that CpsB has one binding site for NADH. Although the results shown in Fig 1C show the monomeric form of CpsB, multimeric forms still need to be confirmed by further studies.

The plots obtained using the M-W-C model [40,85] at low and high substrate concentrations show that the curve was more sigmoidal at low substrate range and shifted towards hyperbola shape at high substrate range (Fig. 6) shifting the L (allosteric constant) value from 0 to1000. Bigger shift in L value shows higher allosteric effect. The n_H shift from 0.31 to 9.33 also conclude the homotropic allosteric effect of increased substrate concentrations. The change in NADH concentrations from 160 to 200 µmol showed a shift in L value from 0 to 100, indicating the heterotropic allosteric effect, but there was no shift observed when the NADH concentration was above 200 µmol (Fig. 6). The results clearly indicate, that not only is NADH necessary for the formation of E*S complex, but it also acts as an allosteric activator, binding at a different site. The results clearly show that CpsB belongs to the allosteric enzymes. It possesses the properties of a K-system with homotropic effects of the substrate (PCP), and heterotropic effects of the activator (NADH). The physiological significance of these results with respect to metabolic control functions remains to be elucidated.

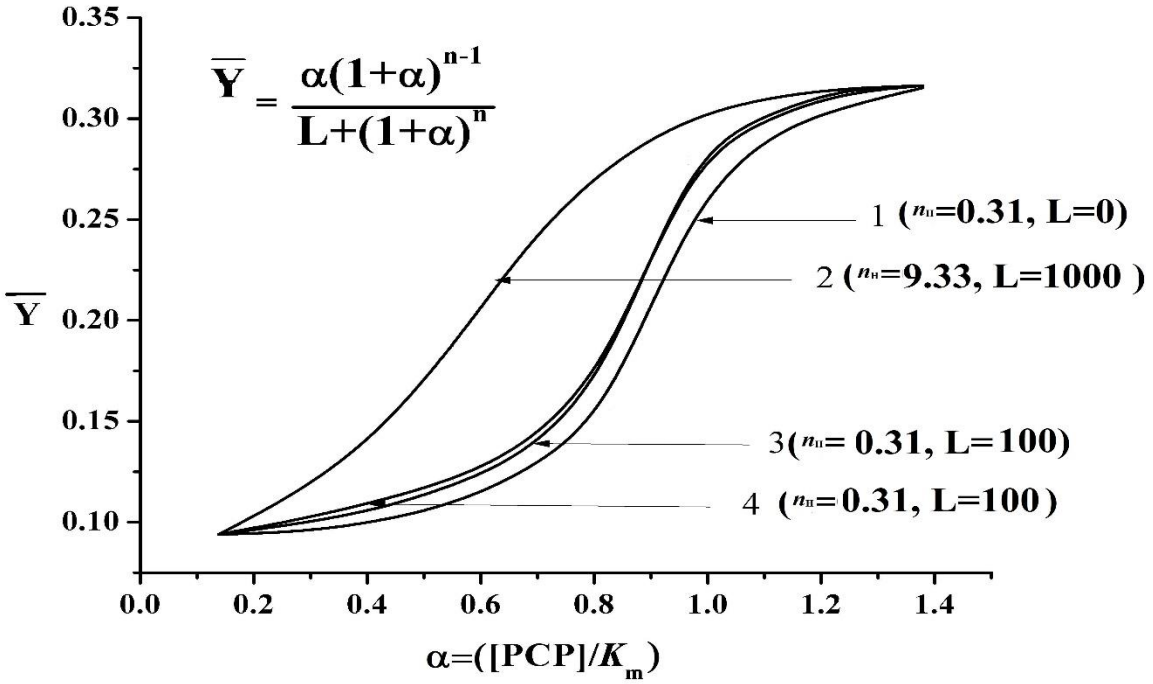


Fig. 6: M-W-C model at low (1, from data shown in fig. 4E) and high (2, data shown in Fig. 4F) [PCP], in the presence of 160 µmol NADH, (3, from data shown in fig. 4E) in the presence of 200 µmol NADH, (4, from data shown in fig. 4E) in the presence of 250 µmol NADH; n_H = Hill coefficient calculated from the respective graphs, L=allosteric constant.

5.3.9 Confirmation of the identity of purified protein as CpsB

The pure SDS-PAGE band excision followed by tryptic digestion resulted in the generation of peptides of CpsB. The LC-MS analysis of the peptides and resolution with a multiple search engine (SearchGui v3.3.15) followed by virtualisation on PeptideShaker (version 1.16.45), revealed that the peptide structure matched with the multi-species phenylalanine-4-hydroxylase protein (accession number: Q818B4) from *Bacillus cereus* (strains ATCC 14579 / DSM 31 / JCM 2152 / NBRC 15305 / NCIMB 9373 / NRRL B-3711), with 78.77% coverage and a 100% peptide matching confidence. The spectrum overview; protein, peptides and PSMs identification summary; peptides structure matches and spectrum identification results, as well as protein, peptides and PSMs validation and quality control plots, are shown in (Fig. S5a-k).

Furthermore, NCBI protein BLAST (protein to protein) search indicated that CpsB shared 100% sequence homology with aromatic amino acid hydroxylase (AAH) from *Bacillus* sp. (WP_149238481) and 99.32% homology with multispecies aromatic amino acid hydroxylase from *Bacillus cereus* group (WP_029439974) (Fig. S6a). This confirmed that CpsB expressed by the cloned cpsB gene fragment from *Bt*AOA is an aromatic amino acid hydroxylase of interest. The accession numbers of the organisms are shown in parenthesis. Putative conserved domains (Fig. S6b) were detected in the theoretical CpsB structural model, using the CDD/SPARCLE functional classification of proteins via subfamily domain architectures [89] in NCBI's conserved domain database, conserved domain database for the functional annotation of protein [90] and protein domain annotation on the fly [91].

CpsB is classified as aromatic amino acid hydroxylase superfamily (accession: cl01244; domain architecture: ID 10014562). The putative specific domains found on CpsB structure include (1) phenylalanine 4-monooxygenase, provisional (PRK14056), with an interval of 3-575 and E-value of 0e+00; (2) aromatic amino acid hydroxylase (cd00361), a biopterin-dependent aromatic amino acid hydroxylase, a family of non-heme, iron(II)-dependent, interval (19-282) and E-value (1.37e-97); (3) Phenylalanine-4-hydroxylase PhhA (COG3186), amino acid transport and metabolism, interval (8-282), E-value (2.47e-46); (4) Phe4hydrox_mono (TIGR01267) phenylalanine-4-hydroxylase, monomeric form; interval (14-282), E-value (7.62e-44) (5) biopterin_H (pfam00351), biopterin-dependent aromatic amino acid hydroxylase, interval (22-290), E-value (5.70e-31). The metal ion binding site is similar to pterin bound human phenylalanine hydroxylase, while cofactor binding site was similar to *Chromobacterium violaceum* phenylalanine hydroxylase [55].

5.3.10 Sequence homology comparison and structural modelling of CpsB

Inquest for genuine sequence homologues and a comprehensive structural characterization of CpsB, the protein was compared with the available structures in the protein data bank (PDB) via the Dali server [92]. The Dali search indicated that CpsB is most similar to Phe4MO (PDB: 1LTU-A) from *Chromobacterium violaceum* with a Z-score of 36.1 and percentage sequence identity of 31% (Fig. S6a), followed by *Legionella pneumophila* (PDB: 4BPT) and C*olwellia psychrerythraea* 34H (PDB: 2v27) with Z-scores of 31.2 and 30.2 and sequences identities of

29% and 29%, respectively (Fig. S6c). The Z-scores for subsequent proteins structures were lower than 30 and were not considered.

Two CpsB structural models were built on the Swiss-Model server [39] using PDB entries 4Q3W and 4JPY individually as templates (Fig. S7a-b). The PDB files of the CpsB models built were submitted at the PDBsum database [47], to determine and generate the secondary structure and the theoretical overview of the model. The overview of the model can be found in the supplementary material (Fig. S8a-i). Ramachandran Plot statistics indicated that 86.6% of CpsB residues were at the most favoured regions, with 11.8% additional allowed regions, 0.8% generously allowed regions and only 0.8% disallowed regions (Fig. S8a). The predicted CpsB secondary structure comprises of 290 residues of βαβ motifs, 2 β-sheets, 1 β-hairpins, 6 strands, 15 α-helixes, 24 helix-helix interactions, 25 β-turns (Fig. 7A, B). Helix-helix interaction regulate stability of enzymes and play a critical role in the structure assembly and function of membrane proteins [93,94]. Likewise, βαβ motif is an important complex super-secondary structure in many proteins because many functional and active sites are often located in polypeptides of βαβ motifs [95].

The 3D structure of CpsB (Fig. 8A) and 3D superimpositions of PDB entries 1LTU, 4BPT, 2V27, 5JK5, 5TPG, 4V06, 2X5N, 5FGJ, 3E2T, 5J6D and CpsB sequences showed that the sequences (Fig. 8B) and structures (Fig. 8C) are conserved in the selected structures. The clefts, pores, tunnels, and the active site in CpsB model structure are shown in Fig 8D-L. There were 10 clefts (Fig 8D, E, F) and 1 each of pore (Fig. 8G, H, I) and tunnel (Fig. 8J, K, L) found in the structure. Most of the clefts were large enough to enhance the effectiveness of the protein catalytic function (Fig. S9a). The hydropathy index and average of normalised hydrophobicity index of the pore (Fig. S9b) showed that the amino acids in the pore are mostly hydrophilic. The average of lining amino acid polarity index also showed that the amino acid in the pore is polar (Fig. S9b). Furthermore, the hydropathy, average of normalised hydrophobicity index and polarity index of the tunnel found in CpsB showed that the amino acids that make up the tunnel are hydrophilic, polar, and charged (Fig. S9c).

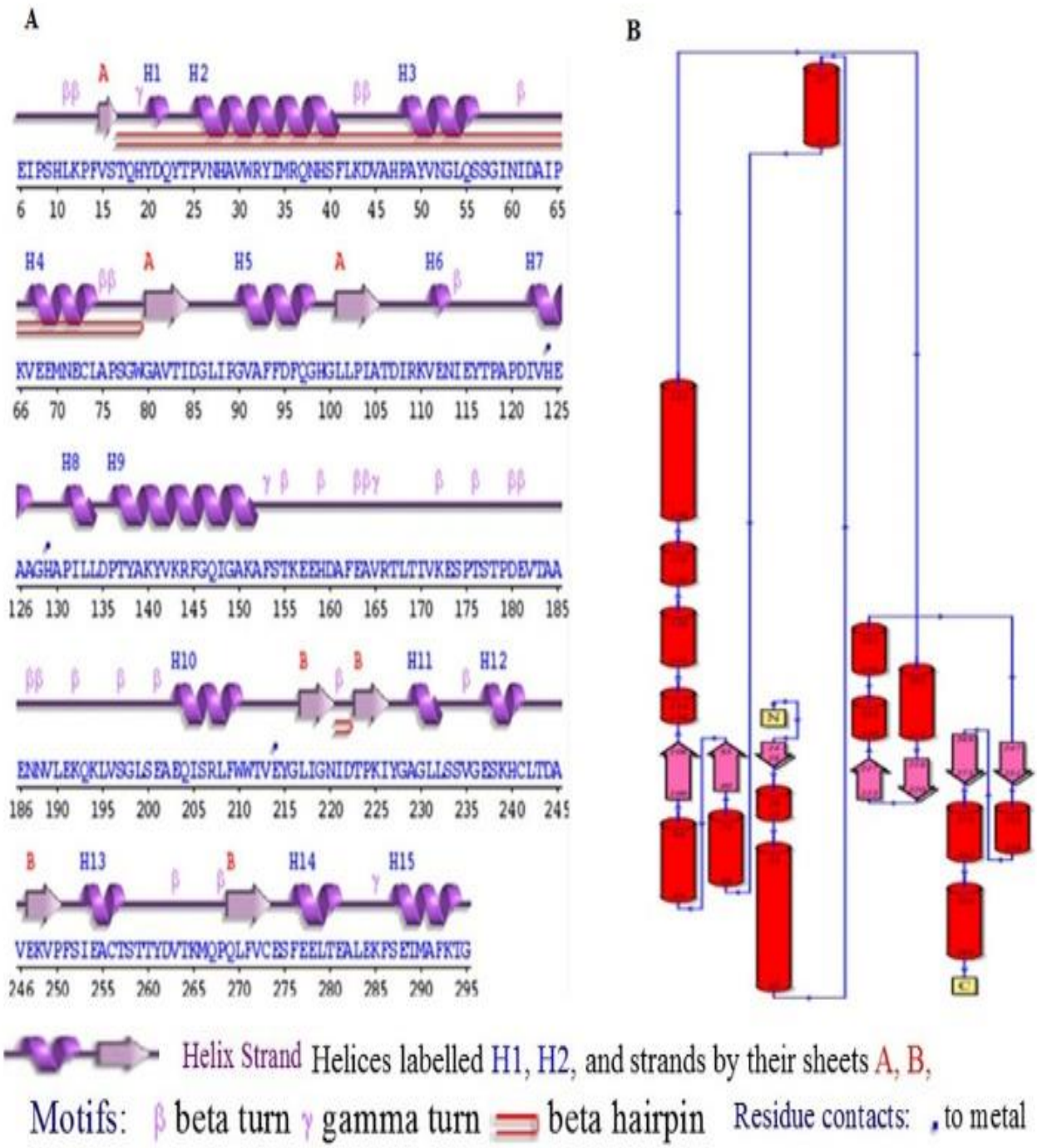


Fig. 7: The predicted 3D and secondary structural framework of CpsB. (A), secondary structural elements of CpsB obtained using PDBsum indicating βαβ motifs, 2 β-sheets, 2 β-hairpins, 7 strands, 15 α-helixes, 21 helix-helix interacts, 25 β-turns and 4 γ-turns; **(B),** topology of CpsB obtained using PDBsum showing the βαβ motifs repeat.

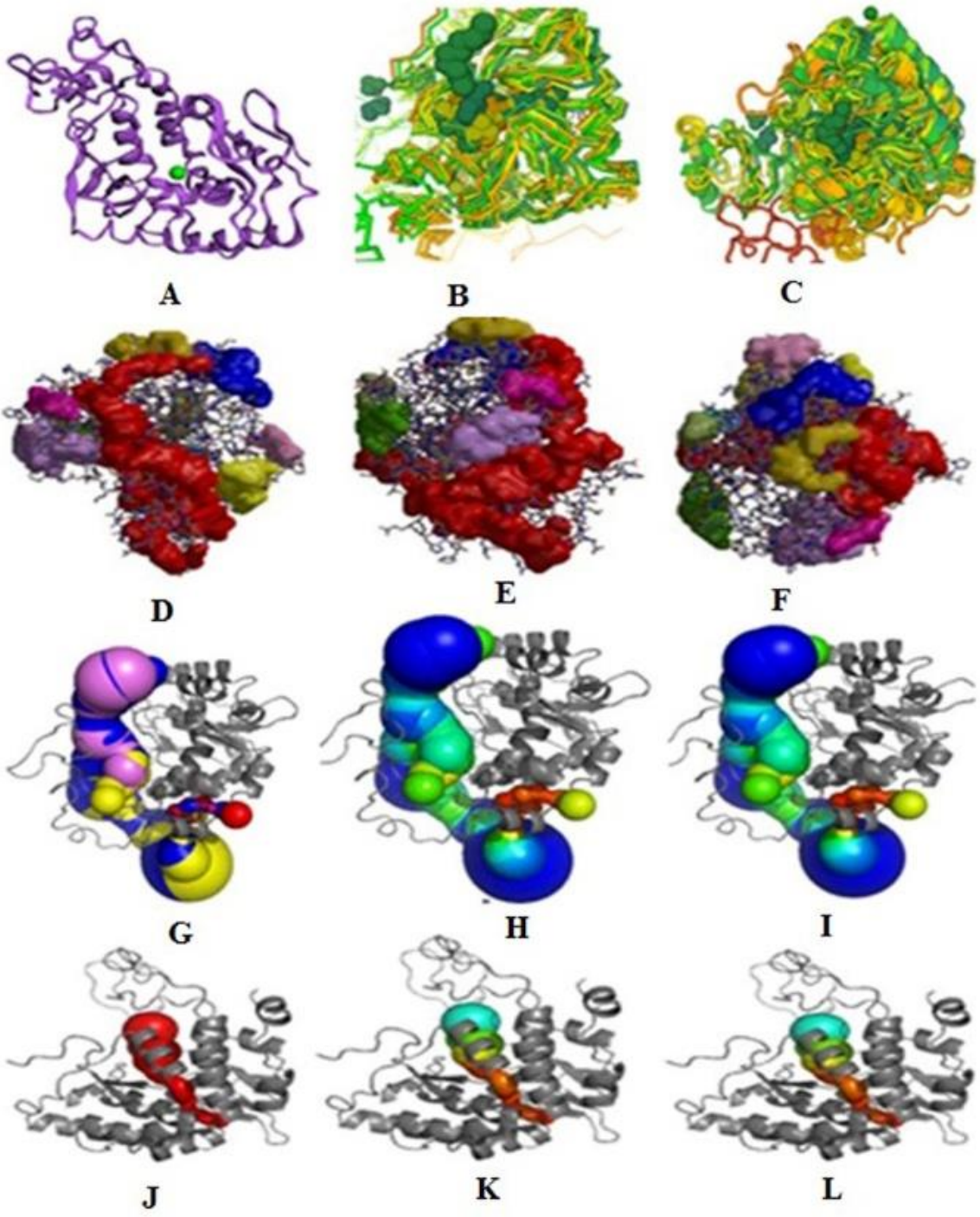


Fig. 8: Graphical representation of the 3D Superimposition of 1LTU, 4BPT, 2V27, 5JK5, 5TPG, 4V06, 2X5N, 5FGJ, 3E2T, 5J6D and CpsB sequences showing: (A), the conserved; **(B)**, conserved structures; **(C)**, 3D structure; **(D,E,F),** cleft; **(G,H,I)**, pore and **(J,K,L),** tunnels found in CpsB structure.

5.3.11 Multiple sequences alignments and evolutionary relatedness

Evolutionary relationship of CpsB was inferred via a phylogenetic based evolutionary study. A phylogenetic tree was constructed using Phe4MO's from all members of Phe4MO superfamily. The phylogenetic tree (Fig. 9) indicated that CpsB is most related to Phe4MO's from *B. cereus* strains E33L (YP085681), AH187 (YP_002340416), and Q1 (YP002531858) with 60% amino acid sequence homology. Also, CpsB shared a common ancestor with Phe4MO's from all members of *B. cereus* group (i.e., from 12 O'clock (YP030501) to 6 O'clock (YP002369167) in the evolutionary tree) in the Phe4MO superfamily, with *B. anthracis* strain Sterne (YP030501) been the first to evolve Phe4MO in the group.

Unexpectedly, 51% (19/37) of the Phe4MO's superfamily were members of *B. cereus* group (i.e. *B. cereus, B. anthracis, B. thuringiensis* and *B. subtilis*)**,** and they are distantly related to Phe4MO from other *Bacillus* species {*B. weihenstephanensis* KBAB4, (YP001646992) and *B. cytotoxicus* NVH 391-98 (YP001376295)} that were on the same node with *Bdellovibrio bacteriovorus* HD100 (NP970269) and *Exiguobacterium* sp. AT1b (YP002886378). This shows that Phe4MO encoding genes might have been spread among members of the *B. cereus* group via horizontal gene transfer (HGT).

Interestingly, CpsB is not evolutionarily related to Phe4MO's from *A. dehalogenans* strains 2CP-C (YP466007) and 2CP-1 (YP002493389), however, they are functionally related in their xenobiotic degradation ability. *Anaeromyxobacter dehalogenans* 2CP-C and 2CP-1 utilized 2-chlorophenol, 2,6-dichlorophenol, 2,5-dichlorophenol, and 2-bromophenol as growth-supporting electron acceptors [96,97]. However, *A. dehalogenans* strains can only utilise ortho-substituted halophenols compared to CpsB that utilise both ortho- and papa-substituted halophenols (unpublished data).

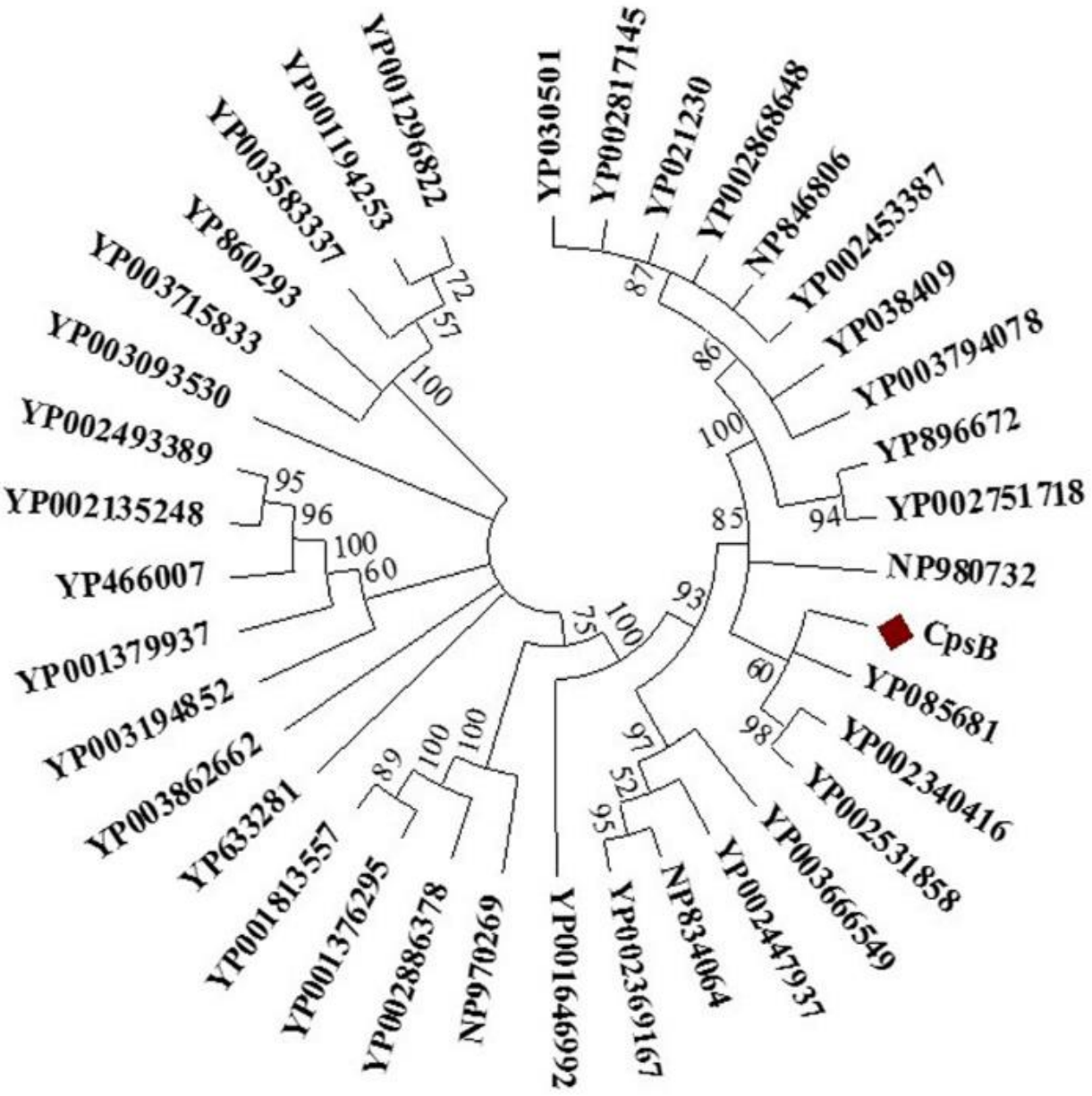


Fig. 9: Evolutionary relatedness of CpsB. The phylogenetic tree was constructed using phenylalanine 4-monooxygenases (Phe4MO) from all the 36 members of the Phe4MO superfamily. *Bdellovibrio bacteriovorus* HD100 (NP970269), *Bacillus anthracis* str. Ames (NP846806), *B. cereus* ATCC 14579 (NP834064), *B. cereus* ATCC 10987 putative (NP980732), *B. thuringiensis* serovar konkukian str. 97-27 (YP038409), *B. anthracis* str. Sterne (YP030501), *B. anthracis* str. Ames ancestor (YP021230), *B. cereus* E33L (YP085681), *Myxococcus xanthus*

DK 1622 (YP633281), *Anaeromyxobacter dehalogenans* 2CP-C (YP466007), *B. thuringiensis* str. Al Hakam (YP896672), *Gramella forsetii* KT0803 (YP860293), *Flavobacterium johnsoniae* UW101 (YP001194253), *F. psychrophilum* JIP02/86 (YP001296822), *B. cytotoxicus* NVH 391-98 (YP001376295), *Anaeromyxobacter* sp. Fw109-5 (YP001379937), *Anaeromyxobacter* sp. K (YP002135248), *Exiguobacterium sibiricum* 255-15 (YP001813557), *B. weihenstephanensis* KBAB4 (YP001646992), *B. cereus* B4264 (YP002369167), *B. cereus* AH187 (YP002340416), *B. cereus* G9842 (YP002447937), *B. cereus* AH820 (YP002453387), *B. cereus* Q1 (YP002531858), *A. dehalogenans* 2CP-1 (YP002493389), *B. cereus* 03BB102 putative (YP002751718), *B. anthracis* str. CDC 684 (YP002817145), *Exiguobacterium* sp. AT1b (YP002886378), *B. anthracis* str. A0248 (YP002868648), *Pedobacter heparinus* DSM 2366 (YP003093530), *Robiginitalea biformata* HTCC2501 (YP003194852), *Zunongwangia profunda* SM-A87 (YP003583337), *B. thuringiensis* BMB171 (YP003666549), *Croceibacter atlanticus* HTCC2559 (YP003715833), *B. cereus* biovar anthracis str. CI (YP003794078) and *Maribacter* sp. HTCC2170 (YP003862662). The bootstrap consensus tree inferred from 500 replicates was taken to represent the evolutionary history of the taxa analysed. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved 37 amino acid sequences. All positions containing gaps and missing data were eliminated and there was a total of 519 positions in the final dataset.

Cluster analysis of CpsB with chlorophenol hydroxylating Phe4MO's from *A. dehalogenans* showed that the residues in contact with the metal ions are conserved, but the sequences are not similar (Fig. 10). The cluster indicated gene deletion at alternating points between CpsB and dehalogenans Phe4MO's. Furthermore, multiple sequences alignments [98], of Phe4MO's from all members of *B. cereus* group in the Phe4MO superfamily showed that the protein structure is conserved among *B. cereus* group with only two members (*B. thuringiensis* BMB171 (YP003666549) and *B. cereus* biovar *anthracis* str. CI (YP003794078)) different in sizes. The alignments (Fig. S10a) also showed that the residues (His124, Glu125, His129 and Glu214, shaded in green) that were theoretically in contact with the Fe^{2+} metal ions were conserved in Phe4MO from all members of the *B. cereus* group in the Phe4MO superfamily. While cluster analysis between Phe4MO's from human, nematodes, rat, zebrafish, fruit fly, amoeba, purple urchin and CpsB showed (Fig. S10b), that the residue (Glu214) conserved in Phe4MO from Phe4MO superfamily is not conserved in human and other mammalians Phe4MO's in accordance with the earlier report [99].

5.3.12 Metal ion interaction and active site of CpsB

The theoretical structural homology modelling of CpsB suggested that the catalytic active site of CpsB has a Fe^{3+} metal ion-binding centre in contact with residues His124, Glu125, His129 and Glu214 (Fig. 11). However, experimental evidence showed that CpsB activity is enhanced in the presence of Fe^{2+} and Ca^{2+} (Table 2). Increased enzyme activity was obtained in the presence of Fe^{2+} as compared to the Fe^{3+}. There were 4 H-bond that might stabilise the connection between the metal ions and the residues at the active centre and H-bond lengths ranged between 1.8-3.3 Å (Fig. S11a). There were other 10 non-bonded contacts (Fig. S11b) within the structure which might play a role in stabilizing the temporary binding of the substrate at the active site of the enzyme.

```
CpsB          ----MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQS  56
YP466007      MGPTERAIAELPPHLRRFVVAQDHAAYTPRDHAVWRHVLRRLTAHLATRAHPRYLAGLAA  60
YP002493389   MGPTERAIAELPPHLRRFVVAQDHAAYTPRDHAVWRHVLRRLTAHLATRAHPRYLAGLAA  60
                      :*:* **: ** :*.:  *** :*****:::*:  :.*   *** *: ** :

CpsB          SGINIDAIPKVEEMNECLAPSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEY  116
YP466007      TGIEVERIPSLDEMNRKLARVGWSAVAVRGFIPPAVFTELQSRRVLAIAADIRTHEHIEY  120
YP002493389   TGIEVERIPSLDEMNRKLARVGWSAVAVRGFIPPAVFTELQSRRVLAIAADIRTHEHIEY  120
              :**::: **.::***. **  **.**:: *:** ..* ::*.: :* **:***. *:***

CpsB          TPAPDIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPT  176
YP466007      TPAPDIIHESAGHAPFIADPTYAEYLRRAGEVGFRAIASAEDQAVFEAIRNLSVVKEDPE  180
YP002493389   TPAPDIIHESAGHAPFIADPTYAEYLQRAGEVGFRAIASAEDQAVFEAIRNLSVVKEDPE  180
              ******:**:*****:: *****:*::* *::* :*::: *:: .***:*.*::***.*

CpsB          STPDEVTAAENNVLEKQKLVSGLSEAEQISRLFWWTVEYGLIGNIDTPKIYGAGLLSSVG  236
YP466007      ASEEEVALSEARLRAASASVRYASESTRASRLYWWTAEYGLVGTLDDPRIYGAGLLSSIG  240
YP002493389   ASDEEVALSEARLRAASASVRYASESTRASRLYWWTAEYGLVGTLDDPRIYGAGLLSSIG  240
              :: :**: :* .:   .  *   **: : ***:***.****:*.:* *:*********:*

CpsB          ESKHCLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCESFEELTEALEKFSETMAFKTGG  296
YP466007      EAVHCLTPAVRKLPLD-PGCADVAYDITRMQPQLFVARDFDQLFEVLDAFDAGLSWRRGG  299
YP002493389   EAVHCLTPAVRKLPLD-PGCADVAYDITRMQPQLFVARDFDQLFEVLDAFTAGLSWRRGG  299
              *: **** **.*:*:.  .*:..:**:*:*******...*::* *.*: *   :::: **

CpsB          KEGLEKAIRSENHATAELNSGLQITGTFTETIENDAGELI-----YMRTSSPTALAIHNK  351
YP466007      DRGLEEARRARTVNHLALSGGRELTGKVAERIPAATEIAPGLSTALVRLDGPVLVS----  355
YP002493389   DHGLEEARRSRSVNHLALSGGRELTGKVAERIPAAAELAPGLSTALVRLDGPVLVS----  355
              ..***:* *:..     *..* ::**..:* *   :          :* ..*. ::

CpsB          QLANHSTSVHSDGFGTPIGLLTGNIALENCTDEQLQSLGITIGNKAAFTFASGIHVKG--  409
YP466007      ----RDGRAEGKPWPGEVVVAFGDAAVPER-------------GPFDLALPGGLALRGFA  398
YP002493389   ----REGRGEGKPWPGEVVVAFGDAEVPDR-------------GPFDLALPGGLALRGFA  398
                   :.   ... :   : :  *:  : :              .   ::: .*: ::*

CpsB          ----TVTDIVKNDKKIALISFINCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAA  465
YP466007      VGGGEVVDLRA---------------TRDGRPLELPTWALLFV--SRDLRSVAGGPADPG  441
YP002493389   VGGGEVVDLRA---------------SREGRPLDLPTWALLFV--SRDLRSVAGGPADPG  441
                   *.*:                 : :.* *   :*. : :  .  : **  * ** .

CpsB          S----------FFPMEEE--IQEIPAPLVLNELERMYQTVRDIRNEGILHDAHIEQLVAI  513
YP466007      AWDRWFGEHGTFTAGEGEARARARKAKALPPALAALYDEVRRLRETGRAT---RERLLAI  498
YP002493389   AWDRWFGEHGTFTAGEGEARARARKAKALPPALAALYDEVRRLREAGRGT---RERLLAI  498
              :          *   * *   :   *  :   *  :*: ** :*: *       *:*:**

CpsB          QEVLNKFYTKEWLLRLEILELLLEHNKGHETSAALLQQLSTFATDEAVTRLINNGLTLLP  573
YP466007      REA-AAAFPGDWLLRAEVDELLAPGAEAGAHP----------------------------  529
YP002493389   REA-AAAFPGDWLLRTEVDELLAPGADAPAHA---------------------------  529
              :*.    :  :**** *: ***    ..

CpsB          VKDVKNDATIN 584
YP466007      ----------- 529
YP002493389   ----------- 529
```

Fig. 10: Multiple sequences alignments of CpsB and Phe4MO's from chlorophenol dehalogenating *Anaeromyxobacter dehalogenans* strains 2CP-C (YP466007) and 2CP-1 (YP002493389) from Phe4MO's superfamily. Cluster analysis constructed with Cluster Omega (1.2.4) multiple sequence alignment tool [98]. Residues in contact with the metal ion is were shaded in green.

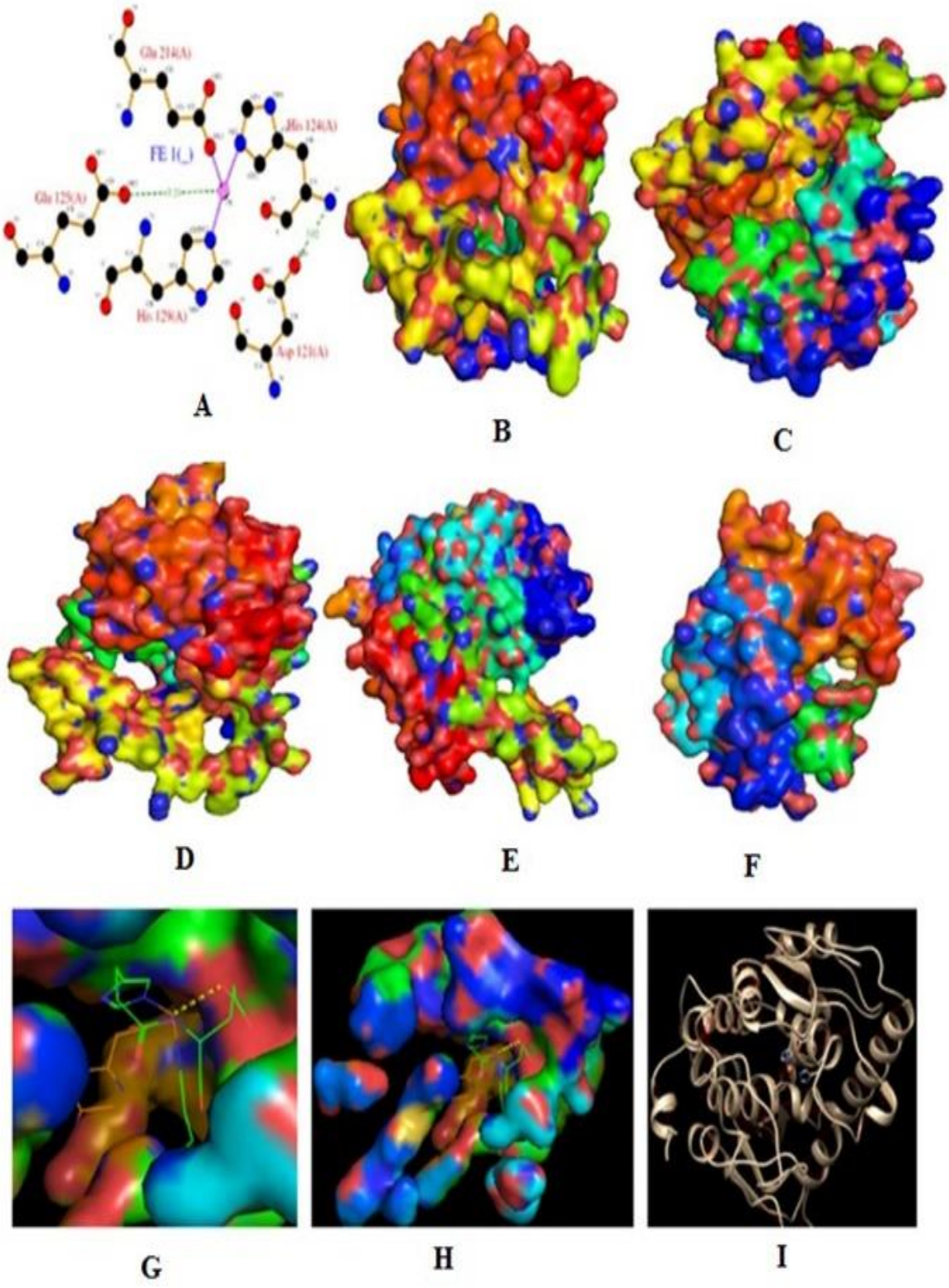


Fig. 11: Metal ion interaction and active site of CpsB based on structural and protein ligand modelling

5.4 Conclusion

In conclusion, Phenylalanine hydroxylating system is a metabolically efficient system that has been implicated in the degradation of many xenobiotics. Although, Phe4MO has been reported to have a diverse catalytic function, to the best of our knowledge, this is the first report of Phe4MO functioning as a PCP dehalogenase or PCP-4-monooxygenase by hydroxylating PCP. Further, this study provides experimental evidence that PheOHS in *Bacillus tropicus* strain AOA-CPS1 can efficiently transform PCP with high affinity. The low K_m of the enzyme showed that the enzyme has a high affinity for PCP and can be of great biotechnological value. The enzyme can operate maximally at ambient temperature, and thus can be used to detoxify PCP in a reactor without incurring additional financial cost to optimize the reaction temperature.

5.5 Acknowledgments:

Authors thank Dr Mare Vlok of the CAF, Stellenbosch University, South Africa for the Liquid-Chromatography Mass-Spectrometry proteomics analysis of the protein. O.A. thank the Tertiary Education Fund (TETFund) Nigeria for the Scholarship.

5.6 Author contributions:

O.A. and A.O. conceived and designed the project; O.A. and A.K. designed the experiments; O.A. performed the experiments; M.P. contributed reagents and materials; O.A., A.K., M.P. and A.O. wrote the manuscript; all the authors have read and approved the manuscript.

5.7 Funding:

National Research Foundation, South Africa (Grant No: 94036 and 92803).

5.8 Conflict of interest:

All the authors declare no conflict of interest.

5.9 Ethical statement:

This article does not contain any studies with human participants or animals performed by any of the authors.

5.10 Reference:

1. ATSDR. Agency for Toxic Substances and Disease Registry. Substance priority list (candidates for toxicological profiles). 2017; 190. accessed on 31st December 2019. https://www.atsdr.cdc.gov/SPL/resources/.
2. Lopez-Echartea E, Macek T, Demnerova K, Uhlik O. Bacterial biotransformation of pentachlorophenol and micropollutants formed during its production process. Int J Environ Res Public Health. 2016;13: 1–21. doi:10.3390/ijerph13111146.
3. Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2019 update: improved access to chemical data. Nucleic Acids Res. 2019 Jan 8; 47(D1):D1102-1109. doi:10.1093/nar/gky1033.
4. Igbinosa EO, Odjadjare EE, Chigor VN, Igbinosa IH, Emoghene AO, Ekhaise FO, et al. Toxicological profile of chlorophenols and their derivatives in the environment: the public health perspective. Sci World J. Hindawi; 2013;2013.
5. IARC. International Agency for Research on Cancer monographs on the identification of carcinogenic hazards to humans. In: Agents Classified by the IARC Monographs, v 1–124., 2019. https://monographs.iarc.fr/agents-classified-by-the-iarc/ (accessed May 28, 2019).
6. IARC. Pentachlorophenol and some related compounds. IARC monographs on the evaluation of carcinogenic risks to humans., 2019. http://publications.iarc.fr/Book-And-Report-Series/Iarc-Monographs-On-The-Identification-Of-Carcinogenic-Hazards-To-Humans/Pentachlorophenol-And-Some-Related-Compounds-2019.
7. Stockholm convention, Stockholm convention on persistent organic pollutants. In: Draft guidance on best available techniques and best environmental practices for the production and use of pentachlorophenol listed with specific exemptions under the Stockholm Convention (UNEP/POPS/COP9/INF/16). Geneva; 2019. Available: http://chm.pops.int/TheConvention/ConferenceoftheParties/Meetings/COP9/tabid/7521/Default.aspx.
8. Ammeri, RW, Tlili SM, Mehri I, Badi S, Hassen A. Pentachlorophenol biodegradation by *Citrobacter freundii* isolated from forest contaminated soil. Water, Air, Soil Pollut. 2016;227: 367. doi:10.1007/s11270-016-2959-z.

9. Alshehri, ANZ. Statistical optimization of pentachlorophenol biodegradation and electricity generation simultaneously in mediator – less air cathode microbial fuel cell. J Environ Appl Bioresearch. 2015;03: 6–15.

10. Karn SK, Chakrabarti SK, Reddy MS. Degradation of pentachlorophenol by *Kocuria* sp. CL2 isolated from secondary sludge of pulp and paper mill. Biodegradation. 2011;22: 63–69. doi:10.1007/s10532-010-9376-6.

11. Joshi VV, Prewitt ML, Ma D-P, Borazjani H. Enhanced remediation of pentachlorophenol (PCP)-contaminated groundwater by bioaugmentation with known PCP-degrading bacteria. Bioremediat J. 2015;19: 160–170. doi:10.1080/10889868.2014.995369.

12. Sharma, A, Thakur, IS, Dureja, P. Enrichment, isolation and characterization of pentachlorophenol degrading bacterium *Acinetobacter* sp. ISTPCP-3 from effluent discharge site. Biodegradation. 2009;20: 643–650.

13. Yang S, Shibata A, Yoshida N, Katayama A. Anaerobic mineralization of pentachlorophenol (PCP) by combining PCP-dechlorinating and phenol-degrading cultures. Biotechnol Bioeng. 2009;102: 81–90.

14. Ayude MA, Okada E, González JF, Haure PM, Murialdo SE. *Bacillus subtilis* as a bioindicator for estimating pentachlorophenol toxicity and concentration. J Ind Microbiol Biotechnol. 2009;36: 765–768. doi:10.1007/s10295-009-0550-y.

15. Singh S, Singh ABB, Chandra AR, Patel ADK, Rai AV. Synergistic biodegradation of pentachlorophenol by *Bacillus cereus* (DQ002384), *Serratia marcescens* (AY927692) and *Serratia marcescens* (DQ002385). 2009; doi:10.1007/s11274-009-0083-6.

16. Cassidy MB, Lee H, Trevors JT, Zablotowicz RB. Chlorophenol and nitrophenol metabolism by *Sphingomonas* sp UG30. J Ind Microbiol Biotechnol. 1999;23: 232–241. doi:10.1038/sj.jim.2900749.

17. Chanama M, Chanama S. Expression of pentachlorophenol-degradative genes of *Sphingobium chlorophenolica* ATCC39723 in *Escherichia coli*. Asian J Public Heal. 2011;2: 78–83.

18. Hlouchova K, Rudolph J, Pietari JMH, Behlen LS, Copley SD. Pentachlorophenol hydroxylase, a poorly functioning enzyme required for degradation of pentachlorophenol by *Sphingobium chlorophenolicum*. Biochemistry. 2012;51: 3848–3860. doi:10.1021/bi300261p.

19. Ning D, Wang H. Involvement of cytochrome P450 in pentachlorophenol transformation in a white rot fungus *Phanerochaete chrysosporium*. PLoS One. 2012;7:e45887. doi:10.1371/journal.pone.0045887.

20. Dai M, Copley SD. Genome shuffling improves degradation of the anthropogenic pesticide pentachlorophenol by *Sphingobium chlorophenolicum* ATCC 39723. Appl Environ Microbiol. 2004;70: 2391–2397. doi:10.1128/AEM.70.4.2391-2397.2004.

21. Tiirola MA, Wang H, Paulin L, Kulomaa MS. Evidence for natural horizontal transfer of the pcpb gene in the evolution of polychlorophenol-degrading *Sphingomonads*. Appl Environ Microbiol. 2002;68: 4495–4501. doi:10.1128/AEM.68.9.4495-4501.2002.

22. Mitchell SC, Steventon GB. Phenylalanine 4-monooxygenase: the "sulfoxidation polymorphism." Xenobiotica. 2020;50: 51–63. doi:10.1080/00498254.2019.1636419.

23. Wang H, Chen H, Hao G, Yang B, Feng Y, Wang Y, et al. Role of the phenylalanine-hydroxylating system in aromatic substance degradation and lipid metabolism in the oleaginous fungus *Mortierella alpina*. Appl Environ Microbiol. 2013;79: 3225–3233. doi:10.1128/AEM.00238-13.

24. Ding Z, Harding CO, Rebuffat A, Elzaouk L, Wolff JA, Thöny B. Correction of murine PKU following AAV-mediated intramuscular expression of a complete phenylalanine hydroxylating system. Mol Ther. 2008;16: 673–681. doi:10.1038/mt.2008.17.

25. Flydal MI, Alcorlo-Pagés M, Johannessen FG, Martínez-Caballero S, Skjaerven L, Fernandez-Leiro R, et al. Structure of full-length human phenylalanine hydroxylase in complex with tetrahydrobiopterin. PNAS USA. 2019;166: 11229–11234. doi:10.1073/pnas.1902639116.

26. Wang H, Yang B, Hao G, Feng Y, Chen H, Feng L, et al. Biochemical characterization of the tetrahydrobiopterin synthesis pathway in the oleaginous fungus *Mortierella alpina*. Microbiology. 2011;157: 3059–3070. doi:10.1099/mic.0.051847-0.

27. Steventon GB, Mitchell SC. Phenylalanine 4-monooxygenase and the role of endobiotic metabolism enzymes in xenobiotic biotransformation. Expert Opinion on Drug Metabolism and Toxicology. 2009. 1213–1221. doi:10.1517/17425250903179318.

28. Greule A, Stok JE, De Voss JJ, Cryle MJ. Unrivalled diversity: the many roles and reactions of bacterial cytochromes P450 in secondary metabolism. 2018; Nat. Prod. Rep. 35757–35791. doi:10.1039/c7np00063d.

29. Flydal MI, Martinez A. Phenylalanine hydroxylase: Function, structure, and regulation. IUBMB Life. 2013;65: 341–349. doi:10.1002/iub.1150.

30. Garg, U., Smith L. D. Chapter 3 - Organic acid disorders. In: Garg U, Smith LD, editors. Biomarkers in inborn errors of metabolism. Elsevier; 2017. pp. 65–85. doi:https://doi.org/10.1016/B978-0-12-802896-4.00002-X.

31. Chang AY, Chau VWY, Landas JA, Pang Y. Preparation of calcium competent *Escherichia coli* and heat-shock transformation. JEMI Methods. 2017;1: 22–25.

32. Marchesi JR, Sato T, Weightman AJ, Martin TA, Fry JC, Hiom SJ, et al. Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. Appl Environ Microbiol. 1998; 64: 795–799.

33. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics; 2009;10: 421. doi:10.1186/1471-2105-10-421.

34. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. 1976;72: 248–254.

35. Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 1970; 227: 680–685. doi:http://dx.doi.org/10.1038/227680a0.

36. Silva MI, Burrows HD. Formosinho SJ, Ferreira L, Alves A, Godinho MJ, Antunes D. Photocatalytic degradation of chlorophenols using $Ru(bpy)_3^{2+}/S2O_8^{2-}$. Env Chem Lett. 2007; 5: 143–149. doi:10.1007/s10311-007-0096-z.

37. Brown WEL, Hill AV. The oxygen-dissociation curve of blood, and its thermodynamical basis. Proc R Soc London Ser B, Contain Pap a Biol Character. The Royal Society London; 1923;94: 297–334.

38. Endrenyi L, Kwong FHF, Fajszi C. Evaluation of Hill slopes and Hill coefficients when the saturation binding or velocity is not known. Eur J Biochem. 1975; 51:317–328.

39. Horn A, Börnig H, Thiele G. Allosteric properties of the Mg^{++}-dependent inorganic Pyrophosphatase in mouse liver cytoplasm. Eur J Biochem. 1967; 2: 243–249.

40. Monod J, Wyman J, Changeux J-P. On the nature of allosteric transitions: a plausible model. J Mol Biol. 1965;12: 88–118.

41. Horn A, Börnig H. Analysis of kinetic data of allosteric enzymes by a linear plot. FEBS

Lett. 1969; 3: 325–329.

42. Strickland S, Palmer G, Massey V. Determination of dissociation constants and specific rate constants of enzyme-substrate (or protein-ligand) interactions from rapid reaction kinetic data. J Biol Chem. 1975; 250: 4048–4052.

43. Liu S, Su T, Zhang C, Zhang W, Zhu D, Su J, et al. Crystal structure of PnpCD, a two-subunit Hydroquinone 1, 2-Dioxygenase, reveals a novel structural class of Fe^{2+}-dependent Dioxygenases. J Biol Chem. 2015; 290: 24547–24560. doi:10.1074/jbc.M115.673558.

44. Setlhare B, Kumar A, Mokoena MP, Pillay B, Olaniran AO. Phenol hydroxylase from *Pseudomonas* sp. KZNSA: Purification, characterization and prediction of three-dimensional structure. Int J Biol Macromol. 2020; 146: 1000–1008. doi:10.1016/j.ijbiomac.2019.09.224.

45. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46: W296-W303. doi:10.1093/nar/gky427.

46. Bienert S, Waterhouse A, de Beer TAP, Tauriello G, Studer G, Bordoli L, Schwede T. The SWISS-MODEL Repository - new features and functionality. Nucleic Acids Res 45, D313–D319.

47. Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM. PDBsum: Structural summaries of PDB entries. Prot Sci. 2018; 27: 129-134.

48. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987; 4: 406–425.

49. Felsenstein, J. Confidence limits on phylogenies: An approach using the bootstrap. Evolution (N Y). 1985; 39: 783–791.

50. Zuckerkandl E. Pauling L. Evolutionary divergence and convergence in proteins. Edited in Evolving Genes and Proteins by V. Bryson and H.J. Vogel, pp. 97-166. Academic Press, New York. 1965.

51. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016; 33: 1870–1874. doi:10.1093/molbev/msw054.

52. Ciufo, S., Kannan, S., Sharma, S., Badretdin, A., Clark, K., Turner, S., et al. Using

average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. Int J Syst Evol Microbiol. 2018; 68: 2386–2392. doi:10.1099/ijsem.0.002809.

53. Federhen S, Rossello-Mora R, Klenk H-P, Tindall BJ, Konstantinidis KT, Whitman WB, et al. Meeting report: GenBank microbial genomic taxonomy workshop (12-13 May, 2015). BioMed Central; 2016.

54. Siltberg-Liberles J, Steen IH, Svebak RM, Martinez A. The phylogeny of the aromatic amino acid hydroxylases revisited by characterizing phenylalanine hydroxylase from *Dictyostelium discoideum*. Gene. 2008;427: 86–92.

55. Erlandsen H, Kim JY, Patch MG, Han A, Volner A, Abu-Omar MM, et al. Structural comparison of bacterial and human iron-dependent phenylalanine hydroxylases: similar fold, different stability and reaction rates. J Mol Biol. 2002;320: 645–661. doi:10.1016/s0022-2836(02)00496-5.

56. Yunyou S, Lifeng C, Brian B, Jian Y. The Catalytic product of Pentachlorophenol 4-monooxygenase is tetra-chlorohydroquinone rather than tetrachlorobenzoquinone. Open Microbiol J. 2008; 2: 100-106. doi:10.2174/1874285800802010100.

57. Simonet P, Gaget K, Parisot N, Duport G, Rey M, Febvay G, et al. Disruption of phenylalanine hydroxylase reduces adult lifespan and fecundity, and impairs embryonic development in parthenogenetic pea aphids. Sci Rep. 2016; 6. doi:10.1038/srep34321.

58. Han CS, Xie G, Challacombe JF, Altherr MR, Bhotika SS, Bruce D, et al. Pathogenomic sequence analysis of *Bacillus cereus* and *Bacillus thuringiensis* isolates closely related to *Bacillus anthracis*. J Bacteriol. 2006; 188: 3382–3390. doi:10.1128/JB.188.9.3382-3390.2006.

59. Leung KT, Campbell S, Gan Y, White DC, Lee HIY, Trevors JT. The role of the *Sphingomonas* species UG30 pentachlorophenol- 4-monooxygenase in *p*-nitrophenol degradation. FEMS Microbiol. Lett. 1999; 173: 247–253. doi:10.1016/S0378-1097(99)00083-X.

60. Wieser M, Wagner B, Ju¨ J, Eberspa¨cher J, Eberspa¨cher E, Lingens F. Purification and Characterization of 2,4,6-Trichlorophenol-4-monooxygenase, a dehalogenating enzyme from *Azotobacter* sp. Strain GP1 J Bacteriol. 1991; 173: 4447–4453. doi:10.1128/jb.173.14.4447-4453.1991.

61. TM Louie CM Webster LX. Genetic and biochemical characterization of a 2,4,6-

trichlorophenol degradation pathway in *Ralstonia eutropha* JMP134. J Bacteriol. 2002; 184: 3492–3500.

62. Xun L. Purification and characterization of Chlorophenol 4-Monooxygenase from *Burkholderia cepacia* AC1100. J. Bacteriol. 1996; 178(9): 2645-9. Doi:10.1128/jb.178.9.2645-2649.1996.
63. Cafaro V, Izzo V, Scognamiglio R, Notomista E, Capasso P, Casbarra A, et al. Phenol hydroxylase and toluene/o-xylene monooxygenase from *Pseudomonas stutzeri* OX1: Interplay between two enzymes. Appl Environ Microbiol. Am Soc Microbiol; 2004; 70: 2211–2219.
64. Kirchner U, Westphal AH, Müller R, van Berkel WJH. Phenol hydroxylase from *Bacillus thermoglucosidasius* A7, a two-protein component monooxygenase with a dual role for FAD. J Biol Chem. ASBMB; 2003; 278: 47545–47553.
65. Long Y, Yang S, Xie Z, Cheng L. Identification and characterization of phenol hydroxylase from phenol-degrading *Candida tropicalis* strain JH8. Can J Microbiol. NRC Research Press; 2014; 60: 585–591.
66. Zhang J-J, Liu H, Xiao Y, Zhang X-E, Zhou N-Y, Kamerbeek M, et al. Identification and characterization of catabolic para-Nitrophenol 4-monooxygenase and para-Benzoquinone reductase from *Pseudomonas* sp. Strain WBC-3. J Bacteriol. 2009; 191: 2703–2710. doi:10.1128/JB.01566-08.
67. Nakamura T, Motoyama T, Hirono S, Yamaguchi I. Identification, characterization, and site-directed mutagenesis of recombinant pentachlorophenol 4-monooxygenase. Biochim Biophys Acta - Proteins Proteomics. 2004; 1700: 151–159. doi:10.1016/j.bbapap.2004.04.008.
68. Wang H, Chen H, Hao G, Yang, Bo., Feng, Yun., Wang, Yu., et al. Role of the phenylalanine-hydroxylating system in aromatic substance degradation and lipid metabolism in the oleaginous fungus *Mortierella alpina*. Appl Environ Microbiol. 2013;79: 3225–3233. doi:10.1128/AEM.00238-13.
69. Bellin, CA., O'Connor, GA., Jin YT. Sorption and degradation of pentachlorophenol in sludge-amended soils. J Env Qual. 1990;19: 603–608.
70. Van Aken P, Lambert N, Van Den Broeck R, Degrève J, Dewil R. Advances in ozonation and biodegradation processes to enhance chlorophenol abatement in multisubstrate wastewaters: a review, Environ. Sci. Water Res. Technol. 2019; 5:444.

doi:10.1039/c8ew00562a.

71. MR Gisi LX. Characterization of chlorophenol 4-monooxygenase (TftD) and NADH: flavin adenine dinucleotide oxidoreductase (TftC) of Burkholderia cepacia AC1100. J Bacteriol. 2009;185: 2786–2792.

72. Das, B. and Patra S. Multisubstrate specific flavin containing monooxygenase from *Chlorella pyrenoidosa* with potential application for phenolic wastewater remediation and biosensor application. Environ Technol FEHLT. 2018; 39: 2073–2089. doi:10.1080/09593330.2017.1349838.

73. Flydal MI, Chatfield CH, Zheng H, Gunderson FF, Aubi O, Cianciotto NP, et al. Phenylalanine hydroxylase from *Legionella pneumophila* is a thermostable enzyme with a major functional role in Pyomelanin synthesis. PLoS One. 2012;7: e46209. doi:10.1371/journal.pone.0046209.

74. Luis AP, Martinez A. Iron binding effects on the kinetic stability and unfolding energetics of a thermophilic phenylalanine hydroxylase from *Chloroflexus aurantiacus*. J Biol Inorg Chem. 2009; 14: 521–531. doi:10.1007/s00775-009-0467-y.

75. Ekstrom F. X-ray characterization of PaPheOH, a bacterial phenylalanine hydroxylase Umeå University, 2003.

76. Schenk T, Muller R, Morsberger F, Otto MK, Lingens F. Enzymatic dehalogenation of pentachlorophenol by extracts from *Arthrobacter* sp. Strain ATCC 33790. J Bacteriol. 1989; 171(10): 5487–5491. doi: 10.1128/jb.171.10.5487-5491.1989.

77. Halcrow MA. Comprehensive coordination chemistry II. In: Jon A. McCleverty and Thomas J. Meyer, editor. Bio-coordination Chemistry. 2nd Editio. ScienceDirect; 2003. p. Volume 3, 2003, pp. 1-92. Available: https://doi.org/10.1016/B0-08-043748-6/02024-7

78. Panay AJ, Lee M, Krebs C, Bollinger JM, Fitzpatrick PF. Evidence for a high-Spin Fe(IV) species in the catalytic cycle of a bacterial phenylalanine hydroxylase. Biochemistry. 2011; 50: 1928–1933. doi:10.1021/bi1019868.

79. Lindhorst AC, Schütz J, Netscher T, Bonrath W, Kühn FE. Catalytic oxidation of aromatic hydrocarbons by a molecular iron-NHC complex. Catal Sci Technol. 2017; 7: 1902–1911. doi:10.1039/c7cy00557a.

80. Dai MH, Rogers JB, Warner JR, Copley SD. A previously unrecognized step in pentachlorophenol degradation in *Sphingobium chlorophenolicum* is catalyzed by

tetrachlorobenzoquinone reductase (PcpD). J Bacteriol. 2003; 185: 302–310. doi:10.1128/JB.185.1.302-310.2003.

81. Chenprakhon P, Wongnate T, Chaiyen P. Monooxygenation of aromatic compounds by flavin-dependent monooxygenases. Protein Science. 2019; 8–29. doi:10.1002/pro.3525.
82. Torres Pazmiño DE, Winkler M, Glieder A, Fraaije MW. Monooxygenases as biocatalysts: Classification, mechanistic aspects and biotechnological applications. J Biotechnol. 2010; 9–24. doi:10.1016/j.jbiotec.2010.01.021.
83. Xun L, Orser CS. Purification and properties of pentachlorophenol hydroxylase, a flavoprotein from *Flavobacterium* sp. strain ATCC 39723. J Bacteriol. 1991; 173: 4447–4453. doi:10.1128/jb.173.14.4447-4453.1991.
84. Wang H, Tiirola MA, Puhakka JA, Kulomaa MS. Production and characterization of the recombinant *Sphingomonas chlorophenolica* Pentachlorophenol 4-monooxygenase. Biochem Biophys Res Commun. 2001; 289: 161–166. doi:10.1006/bbrc.2001.5915.
85. Madsen NB, Shechosky S. Allosteric properties of Phosphorylase b II. comparision with a kinetic model. J Biol Chem. 1967; 242: 3301–3307.
86. Yandell JK, Fay DP, Sutin N. Mechanisms of the reactions of cytochrome c. II. Rate of reduction of horse-heart ferricytochrome c by chromium (II). J Am Chem Soc. 1973; 95: 1131–1137.
87. Lambeth DO, Palmer G. The kinetics and mechanism of reduction of electron transfer proteins and other compounds of biological interest by dithionite. J Biol Chem. 1973; 248: 6095–6103.
88. Strickland S, Massey V. The purification and properties of the flavoprotein melilotate hydroxylase. J Biol Chem. 1973; 248: 2944–2952.
89. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 2017; 45: D200-D203. doi:10.1093/nar/gkw1129.
90. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. Nucleic Acids Res. 2015; 43: D222-D226. doi:10.1093/nar/gku1221.
91. Marchler-Bauer A, Bryant SH. CD-Search: Protein domain annotations on the fly. Nucleic Acids Res. 2004; 32: W327-W331. doi:10.1093/nar/gkh454.
92. Holm L. Benchmarking fold detection by DaliLite v.5. Bioinformatics. 2019; 35(24):

5326-52-27. doi:10.1093/bioinformatics/btz536.

93. Gupta A, Mahalakshmi R. Helix-strand interaction regulates stability and aggregation of the human mitochondrial membrane protein channel VDAC3. J Gen Physiol. 2019; 151: 489–504. doi:10.1085/jgp.201812272.

94. Deber CM, Ng DP. Helix-helix interactions: Is the medium the message? Structure. 2015; 437–438. doi:10.1016/j.str.2015.02.004.

95. Sun L, Hu X, Li S, Jiang Z, Li K. Prediction of complex super-secondary structure βαβ motifs based on combined features. Saudi J Biol Sci. 2016; 23: 66–71. doi:10.1016/j.sjbs.2015.10.005.

96. Wu Q, Sanford RA, Löffler FE. Uranium(VI) reduction by *Anaeromyxobacter dehalogenans* strain 2CP-C. Appl Environ Microbiol. 2006; 72: 3608–3614. doi:10.1128/AEM.72.5.3608-3614.2006

97. He Q, Sanford RA. Characterization of Fe(III) reduction by chlororespiring *Anaeromxyobacter dehalogenans*. Appl Environ Microbiol. 2003; 69: 2712–2718. doi:10.1128/AEM.69.5.2712-2718.2003.

98. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 2019; 47: W636--W641. doi:10.1093/nar/gkz268.

99. Zhao G, Xia T, Song J, Jensent RA. *Pseudomonas aeruginosa* possesses homologues of mammalian phenylalanine hydroxylase and 4α-carbinolamine dehydratase/ DCoH as part of a three-component gene cluster. PNAS USA. 1994.

100. van Spronsen FJ, van Wegberg AMJ, Ahring K, Bélanger-Quintana A, Blau N, Bosch AM, Burlina A, Campistol J, Feillet F, Giżewska M, Huijbregts SC, Kearney S, Leuzzi V, Maillot F, Muntau AC, Trefz FK, van Rijn M, Walter JH, MacDonald A. Key European guidelines for the diagnosis and management of patients with phenylketonuria. The Lancet Diabetes and Endocrinology. 2017; 5; 743-756. doi: 10.1016/S2213-8587(16)30320-5.

101. Aregbesola OA, Mokoena MP, Olaniran AO. Biotransformation of pentachlorophenol by an indigenous *Bacillus cereus* AOA-CPS1 isolated from wastewater effluent in Durban, South Africa. Biodegradation. 2020 Dec;31(4-6):369-383. doi: 10.1007/s10532-020-09915-w. Epub 2020 Oct 3. PMID: 33011889.

5.11 SUPPLEMENTARY MATERIAL

Supplementary material 1a

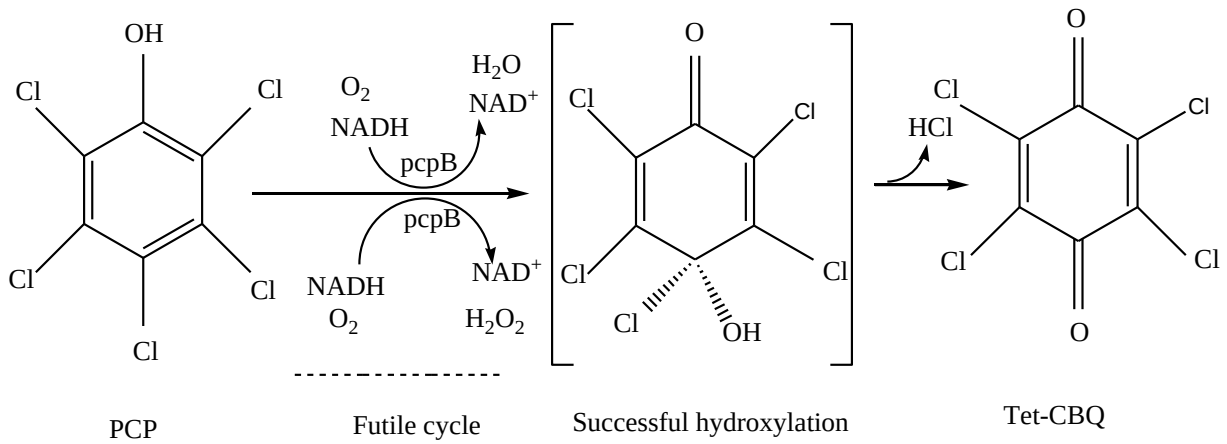


Fig. S1(a): Reaction catalysed by *Sphingobium chlorophenolicum* pentachlorophenol 4-monooxygenase (PcpB) (Chanama and Chanama, 2011, Asian J. Public Health. 2 (2011) 78–83.

Supplementary material 1b

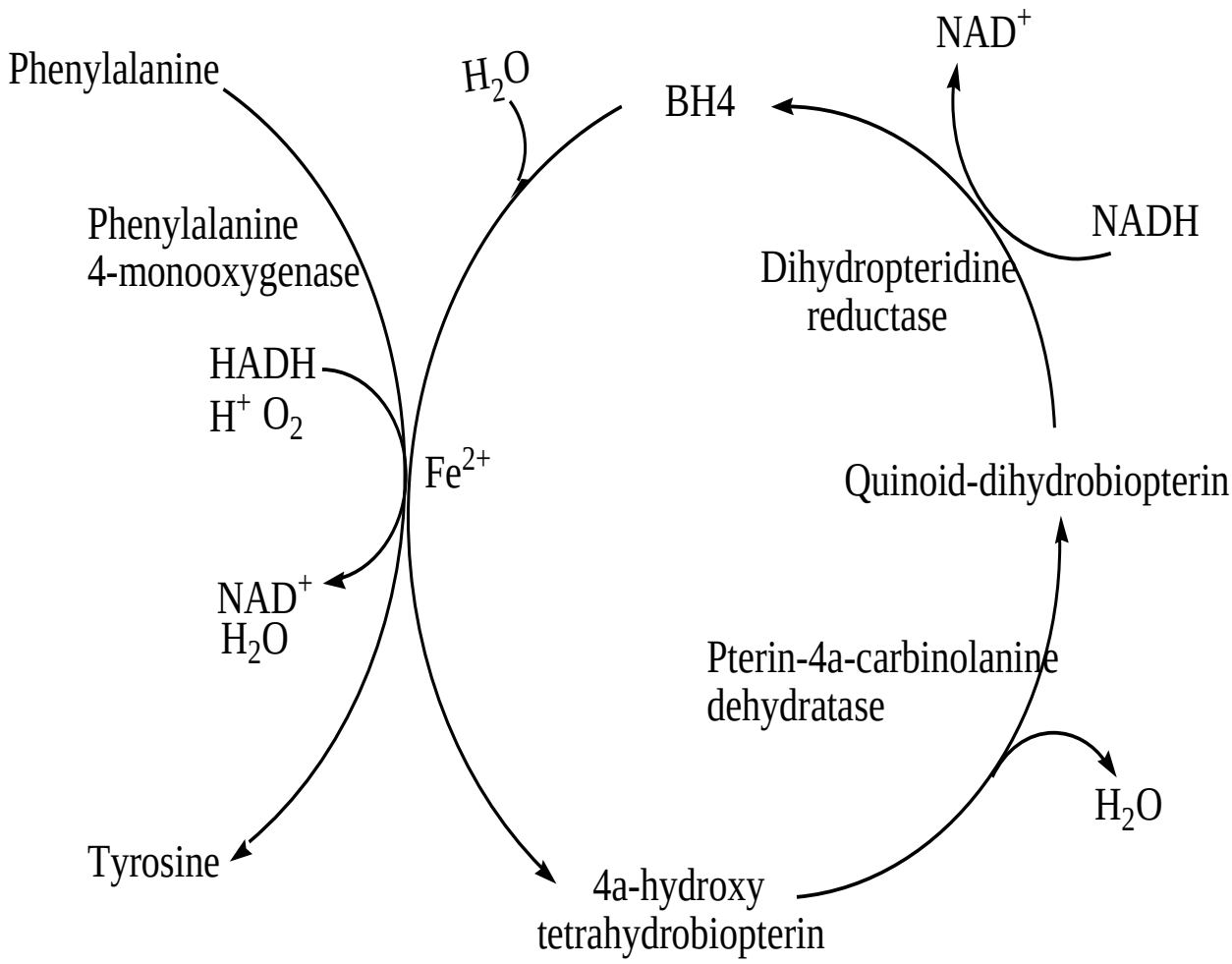


Fig. S1(b): A typical mammalian phenylalanine hydroxylating system showing the enzymes and the substrates used by the system and the reactions catalysed by each of the enzyme. The mammalian phenylalanine hydroxylating system is made up of phenylalanine 4-monooxygenase, pteri-4α-carbinolamine dehydratase and dihydropteridine reductase (van Spronsen et al., 2017, The Lancet Diabetes and Endocrinology, 5(9), 743–756.).

Supplementary material 2

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Manage Columns Show 100

select all *100 sequences selected* GenBank Graphics Distance tree of results

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Bacillus tropicus strain LM1212-W3 chromosome, complete genome	3203	3203	100%	0.0	99.60%	CP041071.1
Bacillus thuringiensis LM1212 chromosome, complete genome	3203	3203	100%	0.0	99.60%	CP024771.1
Bacillus cereus strain FM1, complete genome	3142	3142	100%	0.0	98.97%	CP009369.1
Bacillus cereus G9241, complete genome	3136	3136	100%	0.0	98.92%	CP009590.1
Bacillus cereus strain 03BB87, complete genome	3136	3136	100%	0.0	98.92%	CP009941.1
Bacillus cereus strain BHU2 chromosome	3131	3131	100%	0.0	98.86%	CP023726.1
Bacillus cereus ATCC 4342, complete genome	3109	3109	100%	0.0	98.63%	CP009628.1
Bacillus sp. SYJ chromosome, complete genome	3097	3097	100%	0.0	98.52%	CP036356.1
Bacillus cereus Q1, complete genome	3097	3097	100%	0.0	98.52%	CP000227.1
Bacillus cereus strain AR156, complete genome	3092	3092	100%	0.0	98.46%	CP015589.1
Bacillus paranthracis strain CFSAN068816 chromosome, complete genome	3086	3086	100%	0.0	98.40%	CP045777.1
Bacillaceae bacterium C05 chromosome, complete genome	3086	3086	100%	0.0	98.40%	CP045537.1
Bacillaceae bacterium C02 chromosome, complete genome	3086	3086	100%	0.0	98.40%	CP045533.1

Fig. S2: NCBI blast seach of cpsB gene fragment against the Genebank for similar gene

Supplementary material 3

Fig. S3

Liquid Chromatography-Mass Spectrometry (LC-MS) procedure

In-gel trypsin digestion

To perform the in-gel trypsin digestion, 50 µg of the pure protein stained with 5 µl of Coomassie blue (R250) loading dye (Merck), was loaded onto 12% SDS-PAGE, and run at 100 V for 120 min. The protein bands were excised and distained in a 1.5 ml Eppendorf tube with 200 mM of ammonium bicarbonate (NH_4HCO_3) and acetonitrile (Sigma, UK) in a ratio 1:1 until they become clear. The samples were dehydrated, desiccated, and reduced with 2 mM tris(2-carboxyethyl) phosphine (TCEP) solution (Sigma, UK) in 25 mM NH_4HCO_3 for 15 min with agitation at room temperature. Excess TCEP was removed and the gel pieces were re-dehydrated. The cysteine residues were thiomethylated with 20 mM *S*-Methyl methanethiosulfonate (Sigma, UK) in 25 mM NH_4HCO_3 at room temperature for 30 min. The gel pieces were dehydrated and washed with 25 mM NH_4HCO_3, followed by another dehydration step after thiomethylation.

The proteins were digested by rehydrating the gel pieces in 20 ng/µl Pierce trypsin protease solution (ThermoFisher Scientific, UK) and incubating at 37 °C overnight. Peptides were extracted from the gel pieces once with 50 µl water and once with 50% acetonitrile. The samples were then dried and resuspended in 30 µl of solvent (2% acetonitrile: water); 0.1% Pierce formic acid (FA) (ThermoFisher Scientific, UK). Residual digest reagents were removed using an in-

house manufactured C_{18} stage tip 'Empore Octadecyl C_{18} extraction discs; Supelco' (Sigma-Aldrich, UK). The samples were loaded onto the stage tip after activating the C_{18} membrane with 30 µl methanol (Sigma-Aldrich, UK) and equilibrated with 30 µl of 2% acetonitrile: water; 0.05% trifluoroacetic acid (TFA). The bound sample was washed with 30 µl of 2% acetonitrile: water; 0.1% TFA before elution with 30 µl of 50% acetonitrile: water; 0.05% TFA. The eluate was evaporated to dryness and the dried peptides were dissolved in 2%acetonitrile: water; 0.1% FA for LC-MS analysis.

Liquid chromatography (Dionex nano-RSLC)

Liquid chromatography was performed on a Dionex Ultimate 3000 RSLC system (Thermo Scientific, UK) equipped with a 5mm x 300 µm C_{18} trap column (Thermo Scientific, UK) and ACQUITY UPLC CSH 25 cm x 75 µm 1.7 µm particle size C_{18} column analytical column (Waters, UK). The solvent system employed was loading: 2% acetonitrile:water; 0.1% FA; Solvent A (2% acetonitrile:water; 0.1% FA) and Solvent B (100% acetonitrile:water). The samples were loaded onto the trap column using the loading solvent at a flow rate of 10 µl min^{-1} from a temperature-controlled autosampler set at 7 ^{o}C. The sample was loading for 5 min before eluted onto the analytical column. The flow rate was set to 300 nl min^{-1} and the gradient generated as follows: 2.0% -10%B for 4 min; 10-35%B from 4-60 min and 35%-50%B from 60-70 min. Chromatography was performed at 40 °C and the outflow was delivered to the mass spectrometer through a stainless-steel nano-bore emitter.

Mass spectrometry

Mass spectrometry (MS) was performed using a Fusion mass spectrometer equipped with a Nano-spray Flex ionization source (Thermo Scientific, UK). The sample was introduced through a stainless-steel emitter. Data were collected in positive mode with spray voltage set to 1.8 kV

and ion transfer capillary was set to 280 °C. Spectra were internally calibrated using polysiloxane ions at m/z = 445.12003 and 371.10024. MS1 scans were performed using the orbitrap detector set at 60 000 resolution over the scan range 350-1650 with automatic gain control (AGC) target at 5 E4, maximum injection time of 40 min and data were acquired in profile mode. MS2 acquisitions were performed using monoisotopic precursor selection for an ion with charges +2-+7 with error tolerance set to +/- 10 ppm. Precursor ions were excluded from fragmentation once for a period of 60 s. Precursor ions were selected for fragmentation in higher-energy collisional dissociation (HCD) mode using the quadrupole mass analyser with HCD energy set to 30%. Fragment ions were detected in the orbitrap mass analyser set to 15 000 resolution. The AGC target was set to 5E4, maximum injection time was set to 30 min and the data were acquired in centroid mode.

Data Analysis

The raw files generated by the mass spectrometer were imported into Proteome Discoverer v1.4 (Thermo Scientific) and processed using both Sequest and Amanda algorithm. Database interrogation was performed against a concatenated database created using the supplied database with the cRAP contaminant protein database. Semi-tryptic cleavage with two missed cleavages was allowed. Precursor mass tolerance was set to 10 ppm and fragment mass tolerance set to 0.02 Da. Deamidation (NQ), oxidation (M), and acetylation protein N-terminal were allowed as dynamic modifications and thiomethyl of C as a static modification. Peptide validation was performed using the Target-Decoy PSM validator node, the results files were imported into Scaffold 1.4.4. The LC-MS was done at the central analytical facility, Stellenbosch University, Stellenbosch, South Africa).

The *.mgf file generated from the Sequest and Amanda algorithm was imported as spectrum files into SearchGui v. 3.3.15 GitHub command line to generate *.cui files. The FASTA format protein databases for *Bacillus cereus* group closely related to the isolate used in this study were downloaded from UniProt (http://www.uniprot.org/) and were used as FASTA database in spectrum matching window of identification settings under the search settings option. All modifications, specific trypsin enzyme digestion under protease and fragmentation options were selected with other default settings. The search engines MS Amanda, X! Tandem, MyriMatch, MS-GF+, OMSSA, Comet, Tide and Andromeda were selected along with Novor and DirectTag for DeNovo algorithm. The SearchGui_out zip file generated from SearchGui were viewed with PeptideShaker (v. 1.16.36) for interpretation of the proteomics identification results from multiple search engines while re-calculating PTM localization scores and redoing the protein inference.

Supplementary material 4

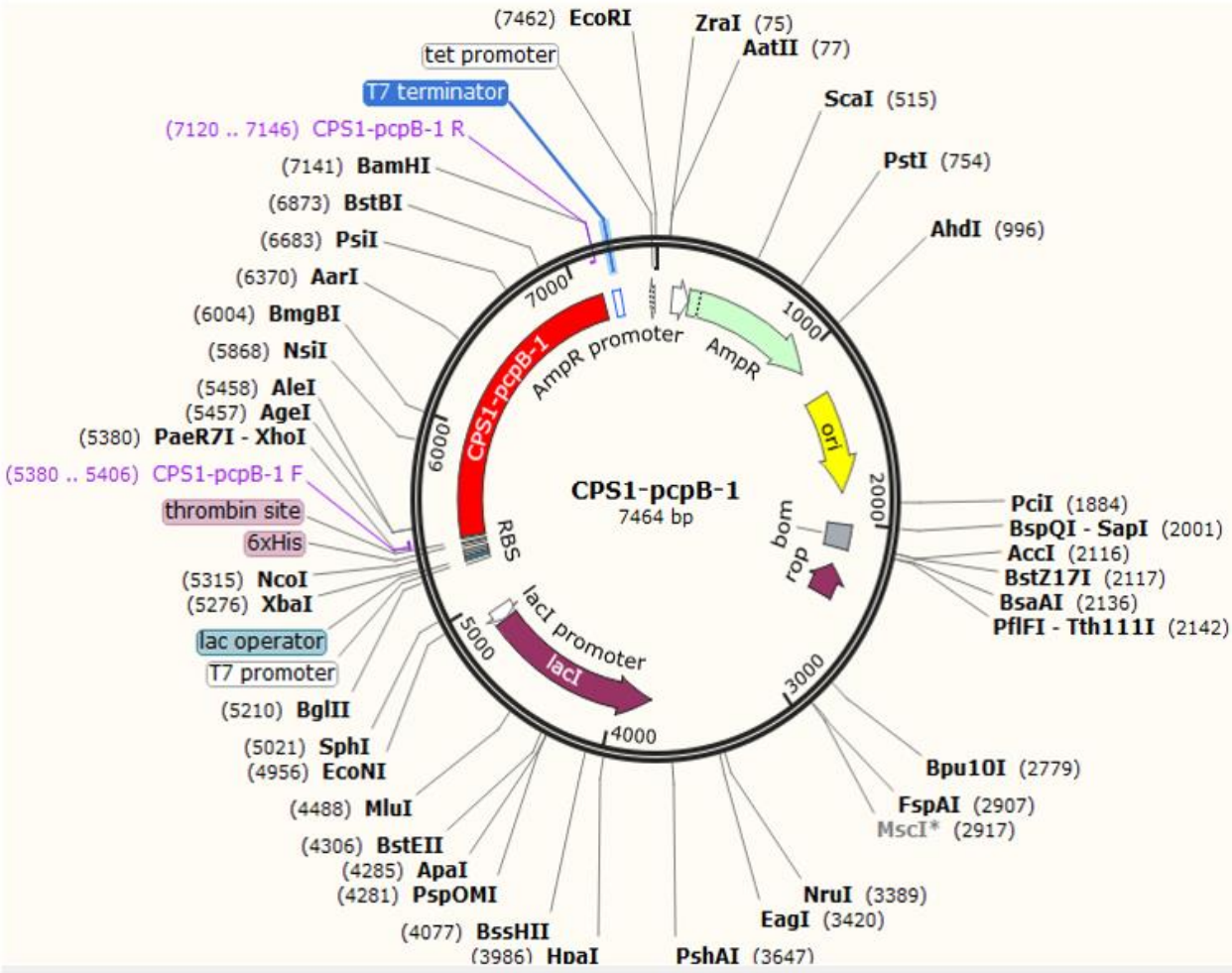


Fig. S4: Map of recombinant pET15b-CpsB vector showing Phe4MO (CpsB) gene fragment.

Supplementary material 5a

Fig. S5(a): Spectrum selection overview

Supplementary material 5b

Peptide Spectrum Matches

	ID	Sequence	Protein(s)	Confidence
1		pyro-EEHDAFEAVR-COOH	Q818B4	100

Spectrum Identification Results

Validated

	SE	Rnk	Sequence	Charge	Confidence
1		1	pyro-EEHDAFEAVR-COOH	2	100
2		10	NH2-DDWKMAYQK-COOH	2	0
3		6	NH2-DTNTYQVSEK-COOH	2	0
4		3	NH2-EEYNPYIEK-COOH	2	0
5		8	NH2-EIYPNYEEK-COOH	2	0
6		7	NH2-ESVQYTNTDK-COOH	2	0
7		5	NH2-GYTQFLHSTM-COOH	2	0
8		2	NH2-IGCVPCYTSNK-COOH	2	0
9		1	NH2-MLHSSEMHGR-COOH	2	0

OMSSA X!Tandem Andromeda MS Amanda MS-GF+ DirecTag Comet MyriMatch Tide Novor PeptideShaker

Fig. S5(b): Peptide spectrum matches and spectrum identification results from different data bases. .

Supplementary material 5c

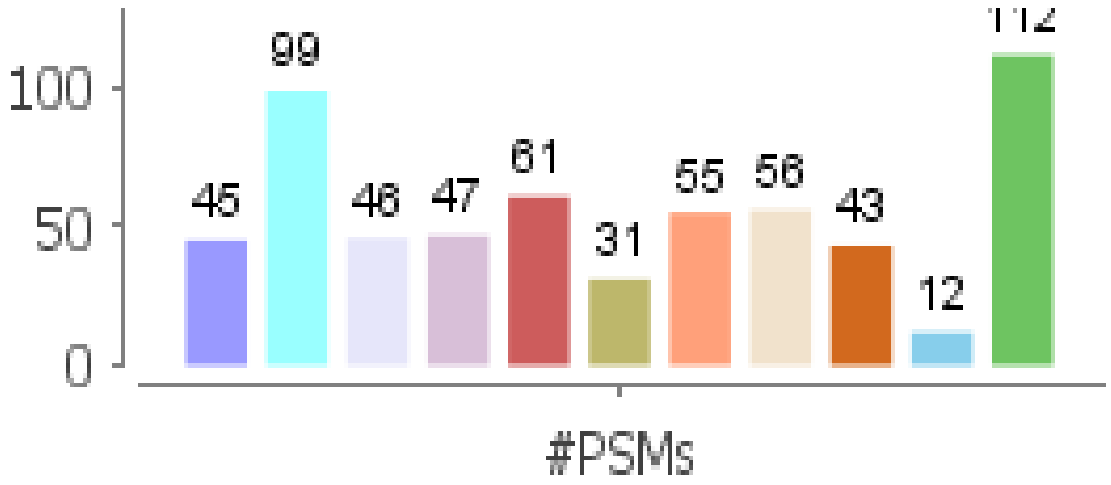


Fig. S5(c): Spectrum identification overview for PSMs

Supplementary material 5d

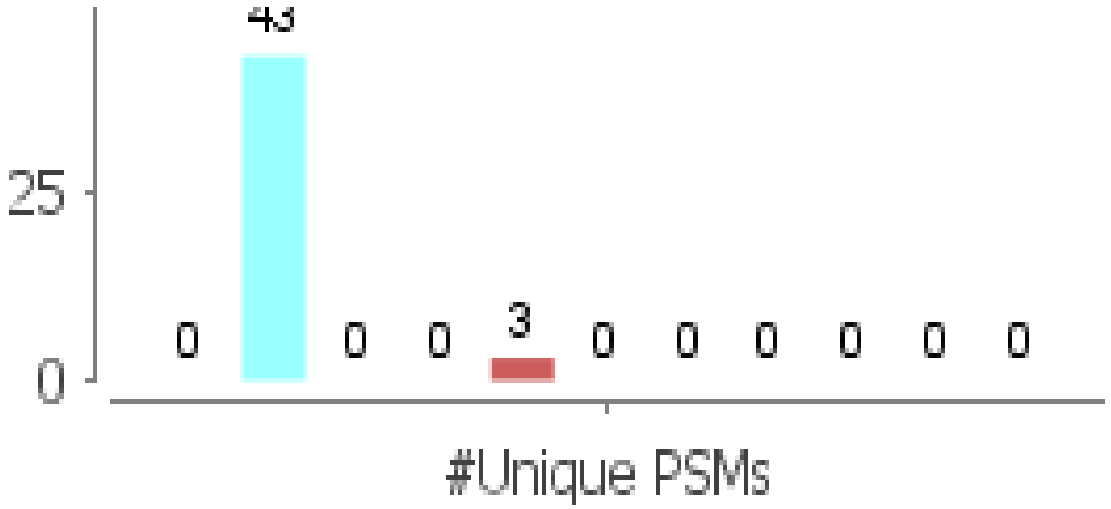


Fig. S5(d): Spectrum identification overview for the unique PSMs

Supplementary material 5e

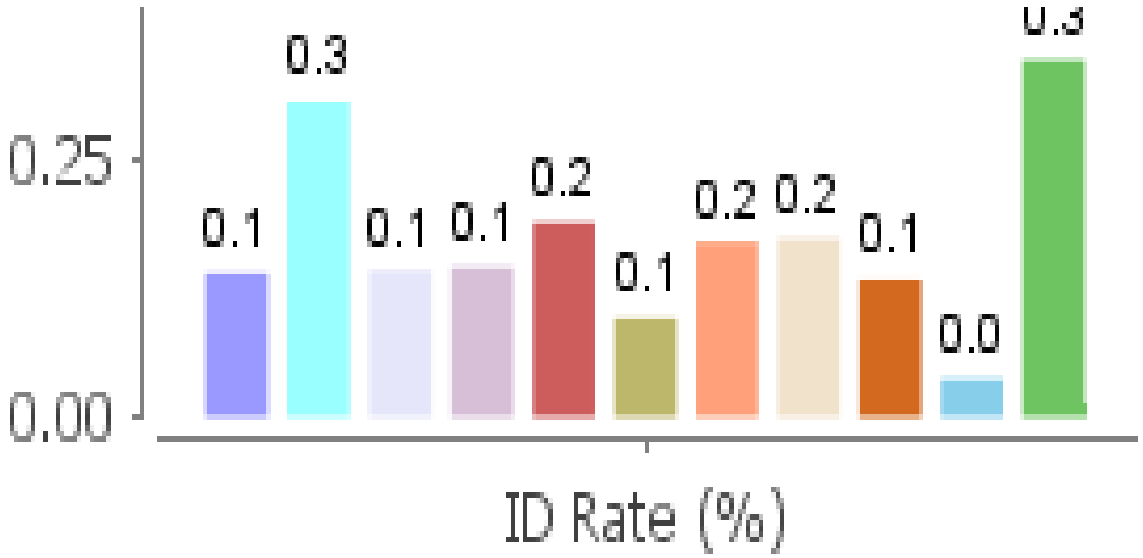


Fig. S5(e): Spectrum identification rate (%)

Supplementary material 5f

Fig. S5(f): Protein validation plots

Supplementary material 5g

Fig. S5(g): Peptide validation plots

Supplementary material 5h

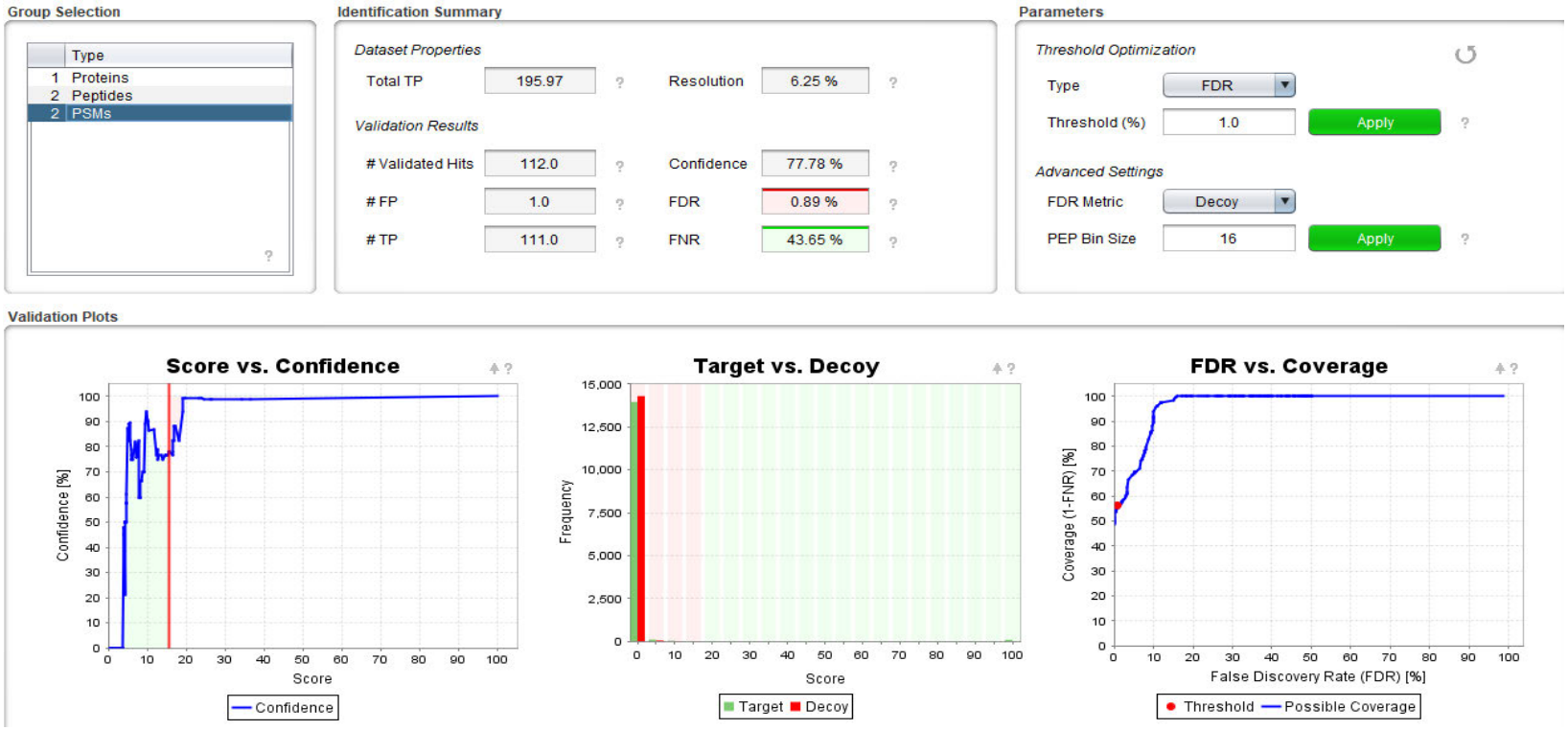


Fig. S5(h): Peptide validation plots

Supplementary material 5i

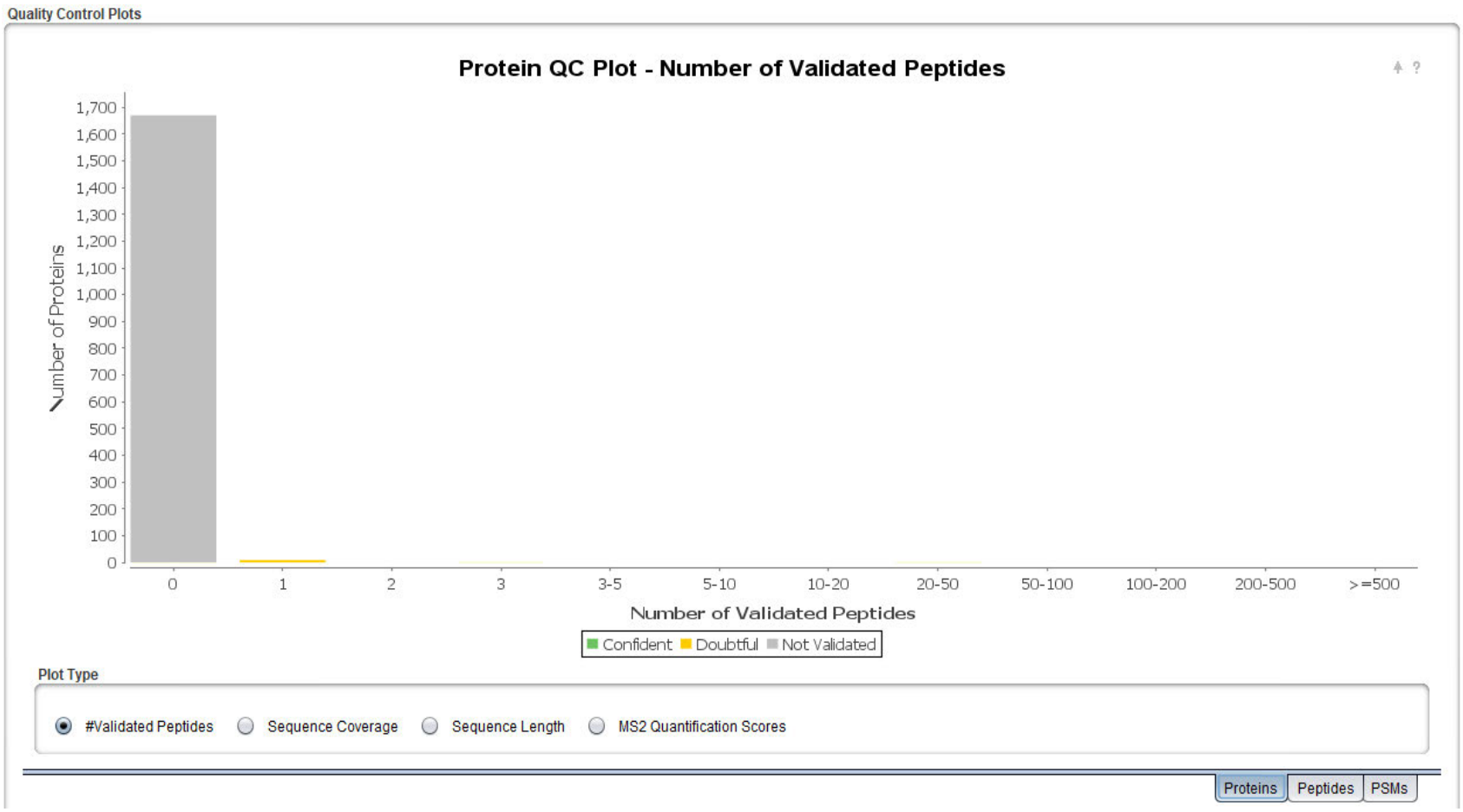


Fig. S5(i): Protein quality control plot statistic

Supplementary material 5j

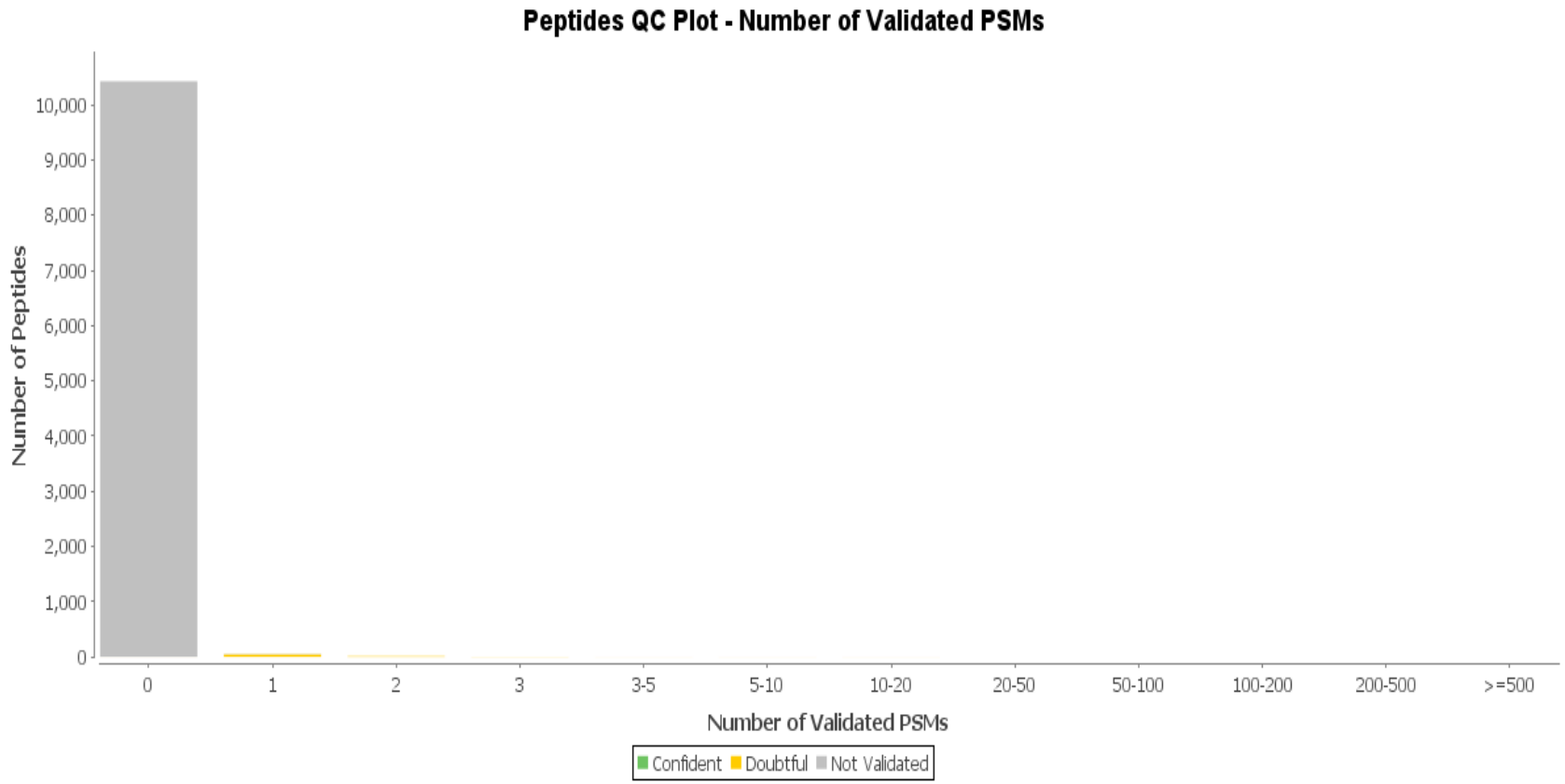


Fig. S5(j): Peptide quality control plot

Supplementary material 5k

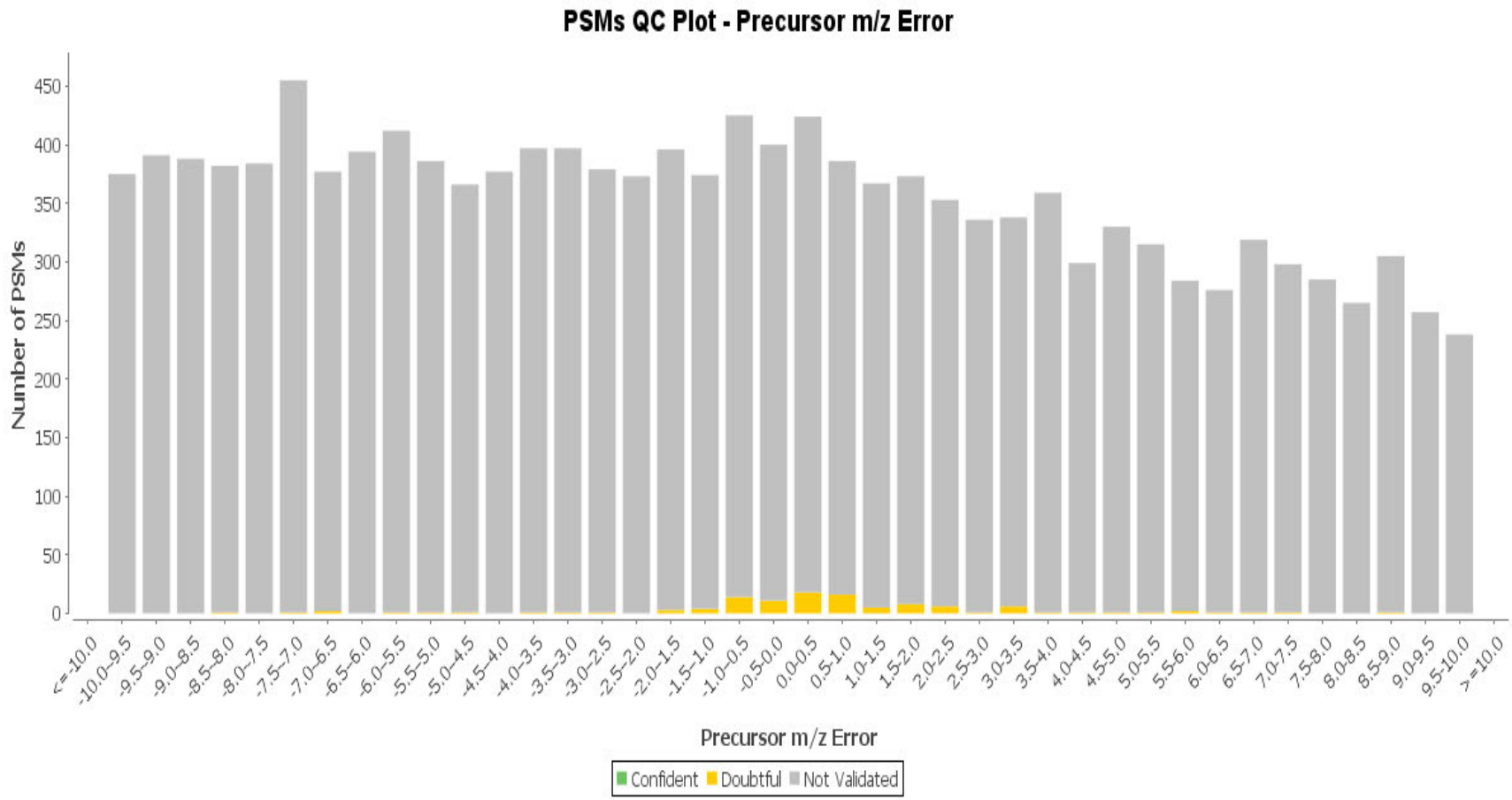


Fig. S5(k): PSMs QC plot

Supplementary material 5l

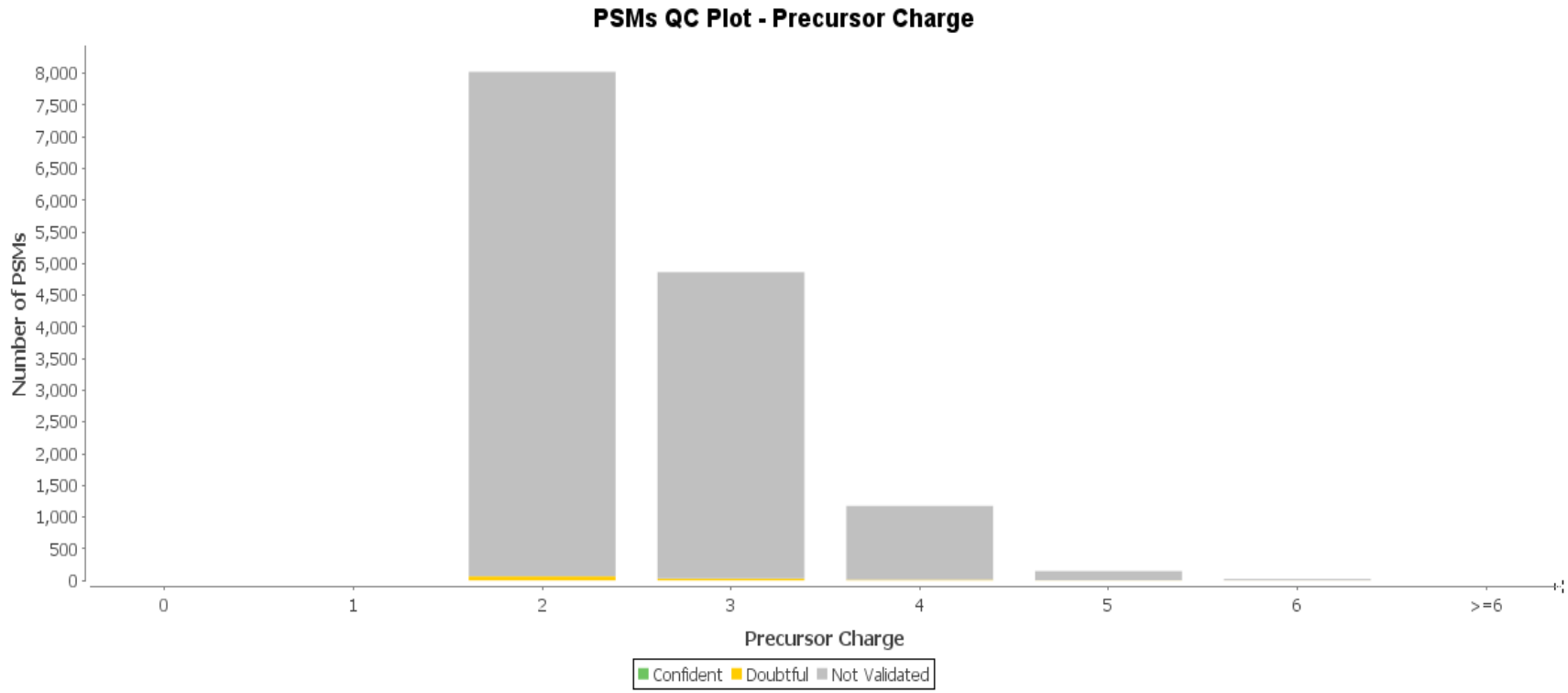


Fig. S5(l): PSMs quality control plot with precursor charge

Supplementary material 6a

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Manage Columns Show 100

select all *100 sequences selected* GenPept Graphics Distance tree of results Multiple alignment

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
aromatic amino acid hydroxylase [Bacillus sp. JAS102]	1206	1206	100%	0.0	100.00%	WP_149238481.1
aromatic amino acid hydroxylase [Bacillus cereus]	1204	1204	100%	0.0	99.83%	WP_073543352.1
aromatic amino acid hydroxylase [Bacillus anthracis]	1202	1202	100%	0.0	99.66%	WP_098214662.1
MULTISPECIES: aromatic amino acid hydroxylase [Bacillus cereus group]	1199	1199	100%	0.0	99.32%	WP_029439974.1
aromatic amino acid hydroxylase [Bacillus anthracis]	1197	1197	100%	0.0	99.14%	WP_098321511.1
aromatic amino acid hydroxylase [Bacillus cereus]	1197	1197	100%	0.0	99.14%	WP_061687105.1
MULTISPECIES: aromatic amino acid hydroxylase [Bacillus cereus group]	1196	1196	100%	0.0	98.97%	WP_042514673.1
aromatic amino acid hydroxylase [Bacillus sp. BD59S]	1196	1196	100%	0.0	98.97%	WP_143881553.1
aromatic amino acid hydroxylase [Bacillus thuringiensis]	1196	1196	100%	0.0	98.97%	WP_087954247.1
aromatic amino acid hydroxylase [Bacillus sp. AFS051223]	1195	1195	100%	0.0	98.80%	WP_098367927.1
aromatic amino acid hydroxylase [Bacillus cereus]	1195	1195	100%	0.0	98.80%	WP_061659337.1
phenylalanine 4-monooxygenase [Bacillus cereus]	1195	1195	100%	0.0	98.80%	ATI49527.1

Fig. S6(a): NCBI protein-protein blast search to determine similar proteins.

Supplementary material 6b

Conserved domains on [lcl|Query_1670]

View Standard Results

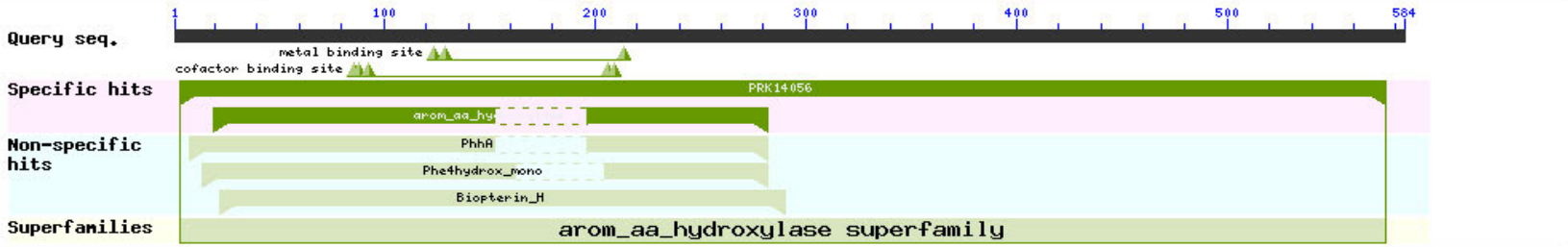
Local query sequence

Protein Classification

aromatic amino acid hydroxylase (domain architecture ID 10014562)
aromatic amino acid hydroxylase such as phenylalanine 4-monooxygenase (PhhA), which catalyzes the irreversible conversion of phenylalanine to tyrosine, using tetrahydrobiopterin (BH4) as a reducing agent

Graphical summary

Zoom to residue level | show extra options »



Search for similar domain architectures | Refine search

List of domain hits

Name	Accession	Description	Interval	E-value
[+] PRK14056	PRK14056	phenylalanine 4-monooxygenase; Provisional	3-575	0e+00
[+] arom_aa_hydroxylase	cd00361	Biopterin-dependent aromatic amino acid hydroxylase; a family of non-heme, iron(II)-dependent ...	19-282	1.37e-97
[+] PhhA	COG3186	Phenylalanine-4-hydroxylase [Amino acid transport and metabolism];	8-282	2.47e-46
[+] Phe4hydrox_mono	TIGR01267	phenylalanine-4-hydroxylase, monomeric form; This model describes the smaller, monomeric form ...	14-282	7.62e-44
[+] Biopterin_H	pfam00351	Biopterin-dependent aromatic amino acid hydroxylase; This family includes ...	22-290	5.70e-31

Blast search parameters

Data Source: Live blast search RID = Z9MMPF6S016
User Options: Database: CDSEARCH/cdd Low complexity filter: no Composition Based Adjustment: yes E-value threshold: 0.01 Maximum number of hits: 500

Fig. S6(b): Putative conserved domains have been detected, click on the image below for detailed results.

Supplementary file 6c

Fig. S6(c): Summary of the protein structural comparison search using the Dali search

S/N	Chain	*Z*-score	rmsd	lali	No. of residues	% ID	PDB	Description
1	1ltu-A	36.1	0.9	246	284	31	PDB	Molecule: Phenylalanine-4-hydroxylase
2	4bpt-A	31.2	1.6	233	252	29	PDB	Molecule: Phenylalanine-4-hydroxylase
3	2v27-B	30.2	1.6	233	272	29	PDB	Molecule: Phenylalanine-4-hydroxylase
4	5jk5-A	27.9	1.6	230	400	29	PDB	Molecule: Phenylalanine-4-hydroxylase
5	5tpg-A	27.6	2.1	233	271	24	PDB	Molecule: Tryptophan-5-hydroxylase
6	4v06	27.6	1.8	231	349	27	PDB	Molecule: Tryptophan-5-hydroxylase
7	2xsn-D	27.4	1.7	230	342	29	PDB	Molecule: Tyrosine-3-hydroxylase
8	5fgj-A	27.2	1.9	233	428	29	PDB	Molecule: Phenylalanine-4-hydroxylase
9	3e2t-A	24.9	2.3	231	308	26	PDB	Molecule: Tryptophan-5-hydroxylase
10	5j6d-B	22.4	2.0	201	248	25	PDB	Molecule: Tryptophan-5-hydroxylase

Supplementary file 7a

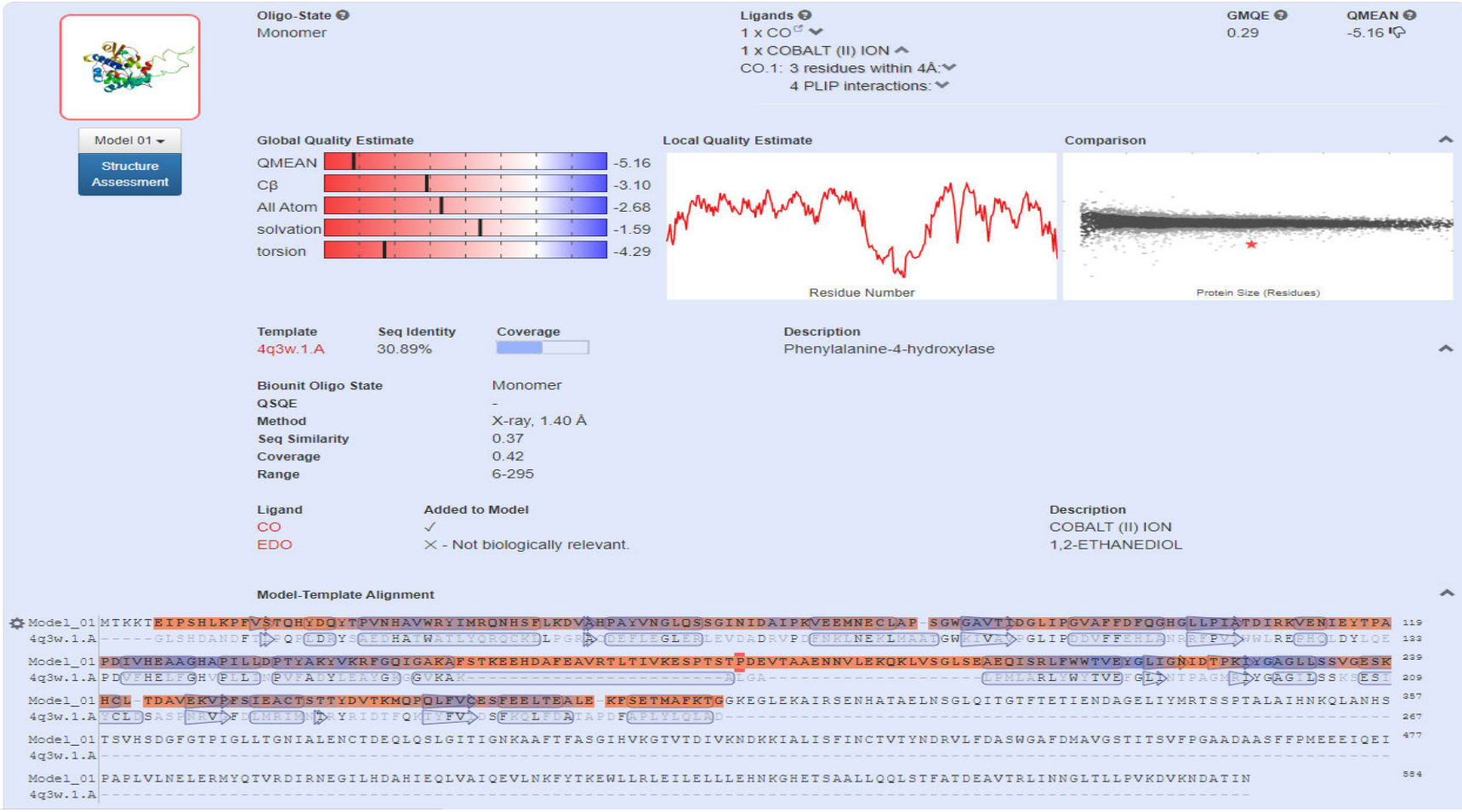


Fig. S7(a): CpsB model built using PDB entry 4q3w.1.A as a template

Supplementary file 7b

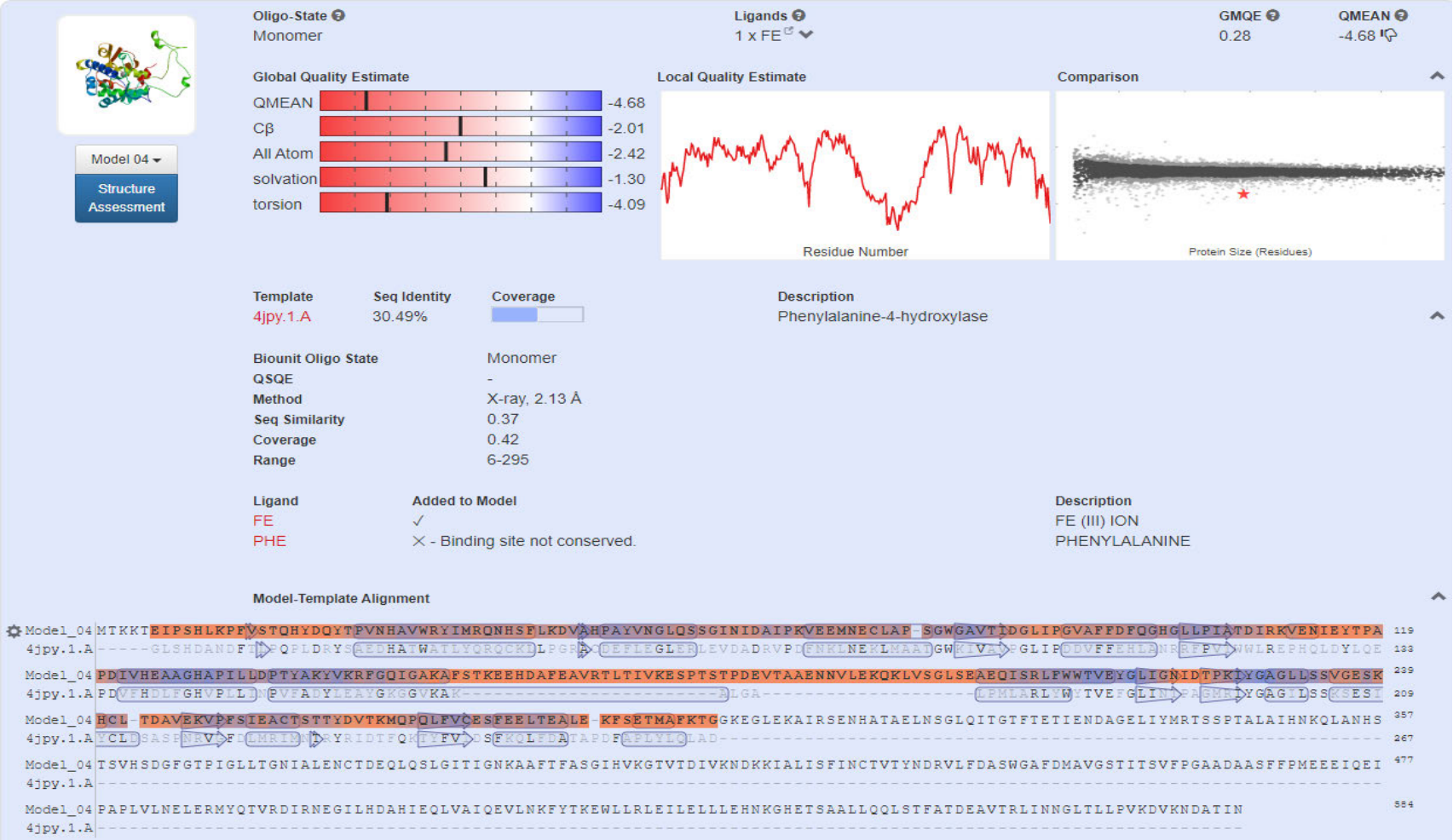


Fig. S7(b): CpsB model built using PDB entry 4q3w.1.A as a template.

Supplementary material 8a

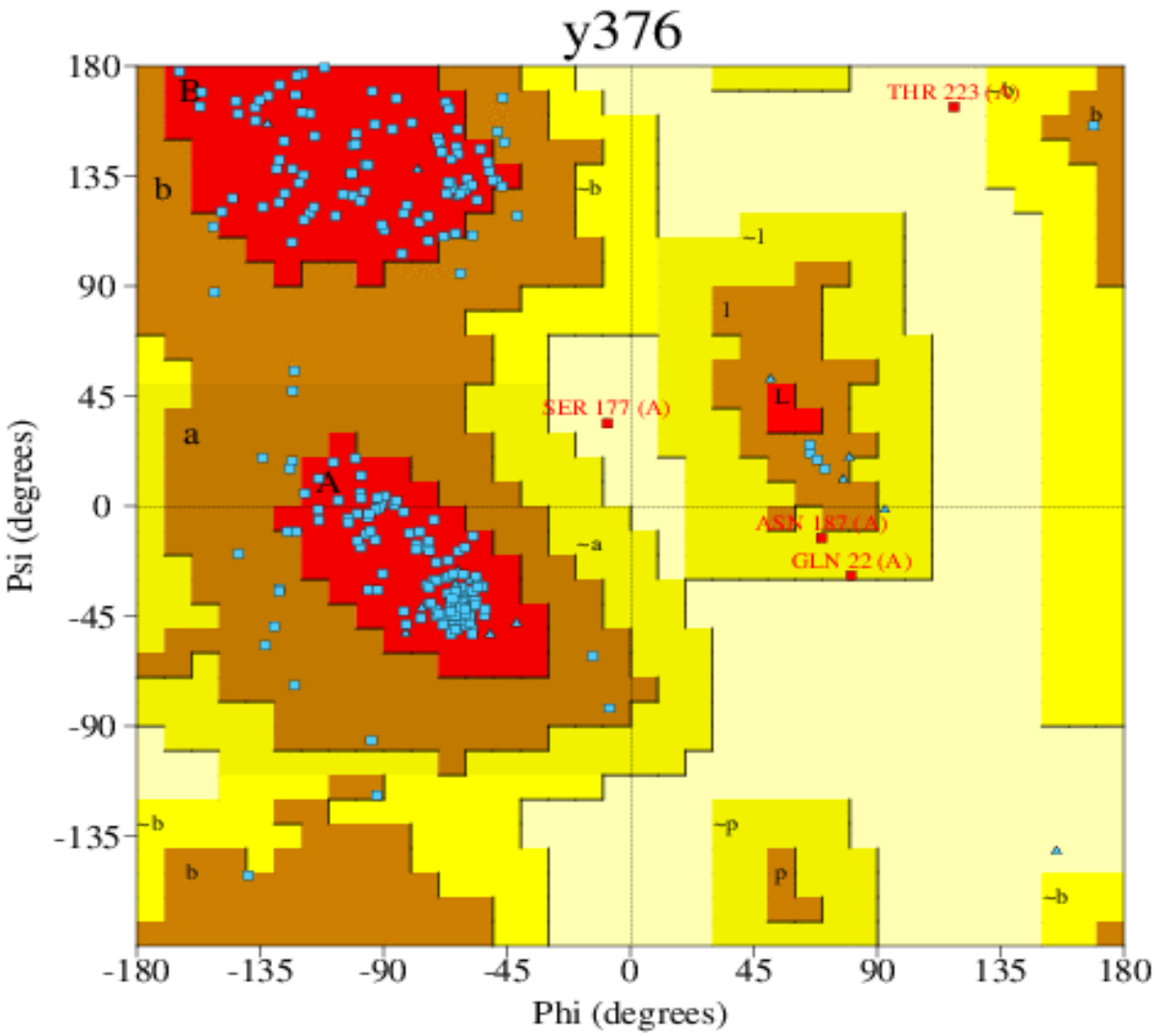


Fig. S8(a): Ramachandran Plot statistics of CpsB residues.

Supplementary material 8b

Fig. S8(b): Summary of CpsB secondary structure

Strand	Alpha helix	3-10 helix	Other	Total residues
27 (9.3%)	103 (35.5%)	9 (3.1%)	151 (52.1%)	290

Supplementary material 8c

Fig. S8(c): Lists of beta sheets in CpsB secondary structure

Sheet	No. strands	Type	Barrel	Topology
A	2	Parallel	N	1X
B	4	Mixed	N	-1 3 -1X

Supplementary material 8d

Fig. S8(d): Lists of beta-alpha-beta motifs in CpsB secondary structure

Strand 1			Strand 2			No. of	No. of residues	
Start	End	Length	Start	End	Length	helices	Loop	Helix
Gly 80	Ile 84	5	Leu 101	Ala 105	5	1	16	11
Glu 247	Pro 250	4	Gln 269	Cys 273	5	1	18	7

Supplementary material 8e

Fig. S8(e): Lists of beta hairpin in CpsB secondary structure

Strand 1			Strand 2			Hairpin
Start	**End**	**Length**	**Start**	**End**	**Length**	
Leu217	Asn220	4	Thr223	Ile226	4	2:2 I

Supplementary material 8f

Fig. S8(f): Lists of strands in CpsB secondary structure

S/N	Start	End	Sheet	No. of Residue
1	Gly80	Ile84	A	5
2	Leu101	Ala105	A	5
3	Leu217	Asn220	B	4
4	Thr223	Ile226	B	4
5	Glu247	Pro250	B	4
6	Gln269	Cys273	B	5

Supplementary material 8g

Fig. S8(g): Lists of helices in CpsB secondary structure

S/N	Start	End	Sheet	No. of Residue
1	Pro25	Phe41	H	17
2	Pro48	Ser56	H	9
3	Val67	Leu74	H	8
4	Ala75	Ser77	G	3
5	Gly90	Gly98	H	9
6	Val111	Asn113	G	3
7	Ile122	His129	H	8
8	Ala130	Leu134	G	5
9	Pro136	Ala152	H	17
10	Ala202	Trp210	H	9
11	Ala229	Leu232	H	4
12	Val235	His240	H	6
13	Ile253	Thr257	H	5
14	Phe276	Ala282	H	7
15	Glu288	Ala291	H	4

Supplementary material 8h

Fig. S8(h): Lists of helix-helix interactions in CpsB structure

							Number of interacting residues	
S/N	**Helices**		**Helix types**		**Interaction type**		**Helix 1**	**Helix 2**
1	A1	A2	H	H	C	N	1	1
2	A1	A7	H	H	I	N	2	4
3	A1	A11	H	H	I	C	1	1
4	A1	A12	H	H	N	N	3	1
5	A2	A3	H	H	C	I	1	1
6	A2	A7	H	H	I	C	2	2
7	A2	A8	H	G	I	I	5	3
8	A2	A11	H	H	I	I	1	1
9	A2	A14	H	H	n	n	1	1
10	A3	A8	H	G	C	C	2	3
11	A5	A9	H	H	I	I	4	6
12	A5	A10	H	H	N	I	3	2
13	A7	A8	H	G	c	n	5	4
14	A7	A10	H	H	c	c	1	1
15	A7	A11	H	H	I	I	3	3
16	A8	A9	G	H	c	n	3	4
17	A8	A14	G	H	c	n	1	1
18	A9	A10	H	H	c	n	5	4
19	A9	A14	H	H	I	I	4	4
20	A10	A13	H	H	I	C	3	2
21	A10	A15	H	H	N	C	1	2
22	A11	A12	H	H	C	N	2	2
23	A13	A14	H	H	n	c	1	1
24	A13	A15	H	H	I	C	2	1

Supplementary material 8i

Fig. S8(i): Lists of beta turns in CpsB secondary structure

S/N	Turn	Sequence	Turn type	H-bond
1	His10-Pro13	HLKP	IV	
2	Leu11-Phe14	LKPF	I	
3	His19-Gln22	HYDQ	IV	
4	Tyr20-Tyr23	YDQY	IV	
5	Leu42-Val45	LKDV	II	Yes
6	Lys43-Ala46	KDVA	IV	
7	Asn60-Ala63	NIDA	IV	
8	Asn113-Tyr116	NIEY	I	Yes
9	Ile114-Thr117	IEYT	I	
10	Asp160-Glu163	DAFE	I	
11	Ala161-Ala164	AFEA	IV	
12	Phe162-Val165	FEAV	I	Yes
13	Ala164-Thr167	AVRT	IV	
14	Leu168-Val171	LTIV	IV	Yes
15	Ile170-Glu173	IVKE	IV	
16	Ser174-Ser177	SPTS	I	Yes
17	Thr178-Glu181	TPDE	I	Yes
18	Pro179-Val182	PDEV	I	
19	Ala185-Asn188	AENN	IV	
20	Glu186-Val189	ENNV	IV	
21	Leu195-Gly198	LVSG	IV	
22	Leu199-Ala202	LSEA	IV	
23	Asn220-Thr223	NIDT	I	Yes
24	Asp262-Lys265	DVTK	IV	
25	Gln267-Leu270	QPQL	VIII	

Supplementary material 9a

Clefts

	Volume	R1 ratio	Accessible vertices		Buried vertices		Average depth		Residue..type							Ligands
1	9500.62	7.64	65.03	2	12.93	1	18.31	1	17	19	26	31	13	13	0	
2	1242.84	0.00	62.53	5	9.41	2	10.79	2	3	5	6	7	1	2	0	
3	1001.95	0.00	56.98	9	6.05	9	9.94	3	4	2	9	3	3	1	0	
4	969.05	0.00	56.95	10	5.45	10	8.02	6	3	5	3	5	2	4	1	
5	757.69	0.00	61.54	7	9.24	4	8.46	5	2	2	5	5	2	2	0	
6	713.39	0.00	59.38	8	7.08	6	8.62	4	3	2	4	3	0	2	0	
7	456.89	0.00	63.72	4	8.93	5	6.82	7	2	3	2	3	2	0	0	
8	367.88	0.00	61.66	6	6.47	8	6.03	9	3	0	1	1	1	5	0	
9	319.78	0.00	64.50	3	9.25	3	6.52	8	1	2	0	5	2	0	0	
10	224.86	0.00	65.67	1	6.69	7	5.75	10	2	1	1	1	1	2	0	

Protein structure

Residue-type colouring

Positive	Negative	Neutral	Aliphatic	Aromatic	Pro & Gly	Cysteine
H,K,R	D,E	S,T,N,Q	A,V,L,I,M	F,Y,W	P,G	C

Fig. S9(a): Properties of the clefts found in CpsD structure

Supplementary material 9b

Pores

	Radius	Free R	Length	HPathy	HPhob	Polar	Rel Mut	Residue..type							Ligands
1	1.18	2.24	28.3	0.25	0.19	11.6	86	2	3	1	4	1	0	0	
2	2.85	3.02	56.0	-0.44	0.00	12.4	76	1	3	1	6	2	1	0	
3	1.45	1.53	100.3	-0.12	-0.05	14.5	80	4	4	3	10	4	1	0	
4	1.26	2.04	141.1	-0.22	-0.01	12.1	82	4	6	4	9	5	2	0	

Residue-type_colouring

Positive	Negative	Neutral	Aliphatic	Aromatic	Pro & Gly	Cysteine
H,K,R	D,E	S,T,N,Q	A,V,L,I,M	F,Y,W	P,G	C

Fig. S9(b): Properties of the pores found in CpsB

Supplementary material 9c

Tunnels

	Radius	Free R	Length	HPathy	HPhob	Polar	Rel Mut	Residue..type							Ligands
1	1.21	2.52	40.6	0.65	0.64	9.8	73	2	2	3	10	8	1	0	

Residue-type_colouring

Positive	Negative	Neutral	Aliphatic	Aromatic	Pro & Gly	Cysteine
H,K,R	D,E	S,T,N,Q	A,V,L,I,M	F,Y,W	P,G	C

Fig. S9(c): Properties of the tunnel found in CpsB

Supplementary Material 10a

```
CpsB            MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN   60
YP002447937     MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN   60
YP003666549     ------------------------------------------------------------   0
NP834064        MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN   60
YP002369167     MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN   60
YP002751718     MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN   60
YP896672        MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN   60
YP003794078     ------------------------------------------------------------   0
YP030501        MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN   60
YP002817145     MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN   60
YP021230        MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN   60
YP002868648     MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN   60
NP846806        MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN   60
YP002453387     MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN   60
YP038409        MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN   60
NP980732        MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN   60
YP085681        MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN   60
YP002340416     MTKKREIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN   60
YP002531858     MTKKREIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN   60

CpsB            IDAIPKVEEMNECLAPSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP   120
YP002447937     IDAIPKVEEMNECLAPSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP   120
YP003666549     ----------NECLAPSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP   50
NP834064        IDAIPKVEEMNECLAPSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP   120
YP002369167     IDAIPKVEEMNECLAPSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP   120
YP002751718     IEAIPKVEEMNDCLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP   120
YP896672        IEAIPKVEEMNECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP   120
YP003794078     ----------NECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP   50
YP030501        IEAIPKVEEMNECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP   120
YP002817145     IEAIPKVEEMNECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP   120
YP021230        IEAIPKVEEMNECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP   120
YP002868648     IEAIPKVEEMNECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP   120
NP846806        IEAIPKVEEMNECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP   120
YP002453387     IEAIPKVEEMNECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP   120
YP038409        IEAIPKVEEMNECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP   120
NP980732        IDAIPKVEEMNECLAPSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP   120
YP085681        IEAIPKVEEMNECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP   120
YP002340416     IDAIPKVEEMNECLAPSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP   120
YP002531858     IDAIPKVEEMNECLAPSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP   120
                          +:*** ***********************************************

CpsB            DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD   180
YP002447937     DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD   180
YP003666549     DIVHEAAGHAPILLDPTYAKYVKHFGKIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD   110
NP834064        DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD   180
YP002369167     DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD   180
YP002751718     DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD   180
YP896672        DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD   180
YP003794078     DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVLTLTIVKESPTSTPD   110
YP030501        DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD   180
YP002817145     DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD   180
YP021230        DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD   180
YP002868648     DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD   180
NP846806        DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD   180
YP002453387     DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD   180
YP038409        DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD   180
NP980732        DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD   180
YP085681        DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD   180
YP002340416     DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD   180
YP002531858     DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD   180
                ************************:**:******************* **************
```

```
CpsB          EVTAAENNVLEKQKLVSGLSEAEQISRLFWWTVEYGLIGNIDTPKIYGAGLLSSVGESKH  240
YP002447937   EVKAAENAVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGNIDDPKIYGAGLLSSVGESKH  240
YP003666549   EVKAAENAVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGNIDDPKIYGAGLLSSVGESKH  170
NP834064      EVKAAENAVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGNIDDPKIYGAGLLSSVGESKH  240
YP002369167   EVKAAENAVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGNIDDPKIYGAGLLSSVGESKH  240
YP002751718   EVTAAENNVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGDIDNPKIYGAGLLSSVGESKH  240
YP896672      EVTAAENNVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGDIDNPKIYGAGLLSSVGESKH  240
YP003794078   EVTAAENNVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGDIDNPKIYGAGLLSSVGESKH  170
YP030501      EVTAAENNVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGDIDNPKIYGAGLLSSVGESKH  240
YP002817145   EVTAAENNVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGDIDNPKIYGAGLLSSVGESKH  240
YP021230      EVTAAENNVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGDIDNPKIYGAGLLSSVGESKH  240
YP002868648   EVTAAENNVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGDIDNPKIYGAGLLSSVGESKH  240
NP846806      EVTAAENNVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGDIDNPKIYGAGLLSSVGESKH  240
YP002453387   EVTAAENNVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGDIDNPKIYGAGLLSSVGESKH  240
YP038409      EVTAAENNVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGDIDNPKIYGAGLLSSVGESKH  240
NP980732      EVTAAENNVLEKQKLVSGLSEAEQISRLFWWTVEYGLIGNIDAPKIYGAGLLSSVGESKH  240
YP085681      EVTAAENNVLEKQKLVSGLSEAEQISRLFWWTVEYGLIGNIDAPKIYGAGLLSSVGESKH  240
YP002340416   EVTAAENNVLEKQKLVSGLSEAEQISRLFWWTVEYGLIGNIDAPKIYGAGLLSSVGESKH  240
YP002531858   EVTAAENNVLEKQKLVSGLSEAEQISRLFWWTVEYGLIGNIDAPKIYGAGLLSSVGESKH  240
              **.**** *:***:***************************:** *****************

CpsB          CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCESFEELTEALEKFSETMAFKTGGKEGL  300
YP002447937   CLTDAVEKVPFSIEACTGTTYDVTKMQPQLFVCESFEELTDALETFSKTMAFKTGGKEGL  300
YP003666549   CLTDAVEKVPFSIEACTGTTYDVTKMQPQLFVCKSFEELTDALETFSKTMAFKTGGKEGL  230
NP834064      CLTDAVEKVPFSIEACIGTTYDVTKMQPQLFVCESFEELTDALETFSKTMAFKTGGKEGL  300
YP002369167   CLTDAVEKVPFSIEACIGTTYDVTKMQPQLFVCESFEELTDALETFSKTMAFKTGGKEGL  300
YP002751718   CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCKSFEELTEALEKFAETMAFKTGGKEGL  300
YP896672      CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCKSFEELTEALEKFAETMAFKTGGKEGL  300
YP003794078   CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCKSFEELTEALEKFAETMAFKTGGKEGL  230
YP030501      CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCKSFEELTEALEKFAETMAFKTGGKEGL  300
YP002817145   CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCKSFEELTEALEKFAETMAFKTGGKEGL  300
YP021230      CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCKSFEELTEALEKFAETMAFKTGGKEGL  300
YP002868648   CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCKSFEELTEALEKFAETMAFKTGGKEGL  300
NP846806      CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCKSFEELTEALEKFAETMAFKTGGKEGL  300
YP002453387   CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCKSFEELTEALEKFAETMAFKTGGKEGL  300
YP038409      CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCKSFEELTEALEKFAETMAFKTGGKEGL  300
NP980732      CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCESFEELTEALEKFSETMAFKTGGKEGL  300
YP085681      CLTDAVEKVPFSIETCTSTTYDVTKMQPQLFVCESFEELTEALEKFSETMAFKTGGKEGL  300
YP002340416   CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCESFEELTEALEKFSETMAFKTGGKEGL  300
YP002531858   CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCESFEELTEALEKFSETMAFKTGGKEGL  300
              ***************:* .***************:******:***.*:.************

CpsB          EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKQLANHSTSV  360
YP002447937   EKAIRSENYATAELNSGLQITGTFSETIENDAGELIYMRTNSPTALALHNKQLANHSTSV  360
YP003666549   EKAIRSENYATAELNSGLQITGTFSETIENDAGELIYMRTNSPTALALHNKQLANHSTSV  290
NP834064      EKAIRSENYATAELNSGLQITGTFSETIENDAGELIYMRTNSPTALALHNKQLANHSTSV  360
YP002369167   EKAIRSENYATAELNSGLQITGTFSETIENDAGELIYMRTNSPTALALHNKQLANHSTSV  360
YP002751718   EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKQLANHSTAV  360
YP896672      EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKQLANHSTAV  360
YP003794078   EKAIRSENNATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKQLANHSTAV  290
YP030501      EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKELANHSTAV  360
YP002817145   EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKELANHSTAV  360
YP021230      EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKELANHSTAV  360
YP002868648   EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKELANHSTAV  360
NP846806      EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKELANHSTAV  360
YP002453387   EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKELANHSTAV  360
YP038409      EKAIRSENNATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKELANHSTAV  360
NP980732      EKAIRSENHATAELNSGLQITGTFTETIENDTGELIYMRTSSPTALAIHNKQLANHSTSV  360
YP085681      EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKQLANHSTSV  360
YP002340416   EKAIRSENHATAELNSGLQITGTFTETIENDADELIYMRTNSPTALAIHNKQLANHSTSV  360
YP002531858   EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTNSPTALAIHNKQLANHSTSV  360
              ******** ***************:******:.*******.******:***:******:*
```

```
CpsB            HSDGFGTPIGLLTGNIALENCTDEQLQSLGITIGNKAAFTFASGIHVKGTVTDIVKNDKK   420
YP002447937     HSDGFGTPIGLLTENIALENCTDEQLQSLGITIGTIAEFTFASGIHVKGTVTDIVKNDKK   420
YP003666549     HSDGFGTPIGLLTENIALENCTDEQLQSLGITIGTIAEFTFASGIHVKGTVTDIVKNDKK   350
NP834064        HSDGFGTPIGLLTENIALENCTDEQLQSLGITIGTIAEFTFASGIHVKGTVTDIVKNDKK   420
YP002369167     HSDGFGTPIGLLTENIALENCTDEQLQSLGITIGTIAEFTFASGIHVKGTVTDIVKNDKK   420
YP002751718     HSDGFGTPIGLLTENIALENCTDEQLQALGITIGNIAEFTFESGIHVKGTVIDIVKNDNK   420
YP896672        HSDGFGTPIGLLTENIALENCTDEQLQALGITIGNIAEFTFESGIHVKGTVIDIVKNHNK   420
YP003794078     HSDGFGTPIGLLTENIALENCTDEQLQALGITIGNIAEFTFESGIHVKGTVTDIVKNDNK   350
YP030501        HSDGFGTPIGLLTENIALENCTDEQLQALGITIGNIAEFTFESDIHVKGTVTDIVKNDNK   420
YP002817145     HSDGFGTPIGLLTENIALENCTDEQLQALGITIGNIAEFTFESDIHVKGTVTDIVKNDNK   420
YP021230        HSDGFGTPIGLLTENIALENCTDEQLQALGITIGNIAEFTFESDIHVKGTVTDIVKNDNK   420
YP002868648     HSDGFGTPIGLLTENIALENCTDEQLQALGITIGNIAEFTFESDIHVKGTVTDIVKNDNK   420
NP846806        HSDGFGTPIGLLTENIALENCTDEQLQALGITIGNIAEFTFESDIHVKGTVTDIVKNDNK   420
YP002453387     HSDGFGTPIGLLTENIALENCTDEQLQALGITIGNIAEFTFESDIHVKGTVTDIVKNDNK   420
YP038409        HSDGFGTPIGLLTENIALENCTDEQLQALGITIGNIAEFTFESGIHVKGTVTDIVKNDNK   420
NP980732        HSDGFGTPIGLLTENIALENCTDEQLQSLGITIGNVAEFTFASGIHVKGTVTDIVKNDKK   420
YP085681        HSDGFGTPIGLLTENIALENCTDEQLQSLGITIGNKAAFTFASGIHVKGTVTDIVKNDKK   420
YP002340416     HSDGFGTPIGLLTENIALENCTDEQLQSLGITIGNKAAFTFASGIHVKGTVTDIVKNDKK   420
YP002531858     HSDGFGTPIGLLTENIALENCTDEQLQSLGITIGNKAAFTFASGIHVKGTVTDIVKNDKK   420
                ************* *************:******. * *** *.******* *****.:*

CpsB            IALISFINCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAASFFPMEEEIQEIPAP   480
YP002447937     IALISFIDCTVTYNARVLFDASWGAFDMAVGSQITSVFPGAADAAAFFPMDEEVQGLPAP   480
YP003666549     IALISFIDCTVTYNARVLFDASWGAFDMAVGSQITSVFPGAADAAAFFPMDEEVQEIPAP   410
NP834064        IALISFIDCTVTYNARVLFDASWGAFDMAVGSQITSVFPGAADAAAFFPMDEEVHEIPAP   480
YP002369167     IALISFIDCTVTYNARVLFDASWGAFDMAVGSQITSVFPGAADAAAFFPMDEEVHEIPAP   480
YP002751718     IALISFINCTVTYKDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP   480
YP896672        IALISFINCTVTYKDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP   480
YP003794078     IALISFINCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP   410
YP030501        IALISFINCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP   480
YP002817145     IALISFINCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP   480
YP021230        IALISFINCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP   480
YP002868648     IALISFINCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP   480
NP846806        IALISFINCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP   480
YP002453387     IALISFINCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP   480
YP038409        IALISFINCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP   480
NP980732        IALISFISCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPTDEEVQKIPAP   480
YP085681        IALISFINCTVTYNDRVLFDASWGSFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP   480
YP002340416     IALISFINCTVTYNDRVLFDASWGSFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP   480
YP002531858     IALISFINCTVTYNDRVLFDASWGSFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP   480
                *******.*****: *********:******* *************:*** :**:: :***

CpsB            LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYTKEWLLRLEILELLLEHNK   540
YP002447937     LVLNELERMYQTVRDIRSEGILHNAHIDQLVAIQEVLNKFYAKEWLLRLEILELLLEHNK   540
YP003666549     LVLNELERMYQTVRDIRSEGILHEAHIDQLVAIQEVLNKFYAKEWLLRLEVLELLLEHNK   470
NP834064        LVLNELERMYQTVRDIRSEGILHDAHIDQLIAIQEVLNKFYAKEWLLRLEVLELLLEHNK   540
YP002369167     LVLNELERMYQTVRDIRSEGILHDAHIDQLVAIQEVLNKFYAKEWLLRLEVLELLLEHNK   540
YP002751718     LVLNELERMYQTVRDIRNEGILHDAHIDQLVAIQEVLNKFYAKEWLLRLEILELLLEHNK   540
YP896672        LVLNELERMYQTVRDIRNEGILHDAHIDQLVAIQEVLNKFYAKEWLLRLEILELLLEHNK   540
YP003794078     LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYTKEWLLRLEILELLLEHNK   470
YP030501        LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYTKEWLLRLEILELLLEHNK   540
YP002817145     LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYTKEWLLRLEILELLLEHNK   540
YP021230        LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYTKEWLLRLEILELLLEHNK   540
YP002868648     LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYTKEWLLRLEILELLLEHNK   540
NP846806        LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYTKEWLLRLEILELLLEHNK   540
YP002453387     LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYTKEWLLRLEILELLLEHNK   540
YP038409        LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYAKEWLLRLEILELLLEHNK   540
NP980732        LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKLYPKEWLLRLEILELLLEHNK   540
YP085681        LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYTKEWLLRLEILELLLEHNK   540
YP002340416     LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKLYPKEWLLRLEILELLLEHNK   540
YP002531858     LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKLYPKEWLLRLEILELLLEHNK   540
                ******************.*****:***:**:********:* ********:*********
```

```
CpsB            GHETSAALLQQLSTFATDEAVTRLINNGLTLLPVKDVKNDATIN  584
YP002447937     GHEASATLLQQLSTFTTDEAVTRLINNGLALLPVKDVKNDAKIN  584
YP003666549     GHETSAALLHQLSTFTTDEAVTRLINNGLALLPVKDVKNDAKIN  514
NP834064        GHETSAALLHQLSTFTTDEAVTRLINNGLALLPVKDVKNDAKIN  584
YP002369167     GHETSAALLHQLSTFTTDEAVTRLINNGLALLPVKDVKNDAKIN  584
YP002751718     GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKDVKNDATIN  584
YP896672        GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKDVKNDATIN  584
YP003794078     GHETSAALLQQLSTFTSDEAVTRLINNGLTLLPVKDVKNDATIN  514
YP030501        GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKGVKNDATIN  584
YP002817145     GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKGVKNDATIN  584
YP021230        GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKGVKNDATIN  584
YP002868648     GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKGVKNDATIN  584
NP846806        GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKGVKNDATIN  584
YP002453387     GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKGVKNDATIN  584
YP038409        GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKDVKNDATIN  584
NP980732        GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKDVKNDATIN  584
YP085681        GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKDVKNDATIN  584
YP002340416     GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKDVKNDATIN  584
YP002531858     GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKDVKNDATIN  584
                ***:**:**:*****::**************:*****.*****.**
```

Fig. S10(a): Multiple sequences alignments of Phe4MO's from all members of *Bacillus cereus* group in the Phe4MO superfamily. Members of the *Bacillus cereus* group in the Phe4MO were *Bacillus anthracis* str. Ames (NP846806), *B. cereus* ATCC 14579 (NP834064), *B. cereus* ATCC 10987 putative (NP980732), *B. thuringiensis* serovar konkukian str. 97-27 (YP038409), *B. anthracis* str. Sterne (YP030501), *B. anthracis* str. Ames Ancestor (YP021230), *B. cereus* E33L (YP085681), *B. thuringiensis* str. Al Hakam (YP896672), *B. cereus* B4264 (YP002369167), *B. cereus* AH187 (YP002340416), *B. cereus* G9842 (YP002447937), *B. cereus* AH820 (YP002453387), *B. cereus* Q1 (YP002531858), *B. cereus* 03BB102 putative (YP002751718), *B. anthracis* str. CDC 684 (YP002817145), *B. anthracis* str. A0248 (YP002868648), *B. thuringiensis* BMB171 (YP003666549) and *B. cereus* biovar *anthracis* str. CI (YP003794078). The cluster analysis was done using CLUSTAL O(1.2.4) multiple sequence alignment (Madeira et al., 2019).

Supplementary Material 10b

```
Feature 1                                                                                  # #    #
*CpsB       19 HYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSg----INIDAIPKVEEMNECLAP-SGWGAVTIDGLIPGVAF 093 CpsB
1DMW_A      58 PRVEYMEEEKKTWGTVFKTLKSLYKTHACYEYNHIFPLLekycgFHEDNIPQLEDVSQFLQTcTGFRLRPVAGLLSSRDF 137 human
AAD30115   242 PRTEYTSSERKTWGIIYRKLRELHKKHACKQFLDNFELLerhcgYSENNIPQLEDICKFLKAkTGFRVRPVAGYLSARDF 321 nematode
1PHZ_A     175 PRVEYTEEEKQTWGTVFRTLKALYKTHACYEHNHIFPLLekycgFREDNIPQLEDVSQFLQTcTGFRLRPVAGLLSSRDF 254 Norway rat
XP_641959  163 PRIDYTEEEIKTWGVVYNRLKELFPTNACHQHAYIFPLLeqncgYSPDNIPQLQDISNFLQEcTGWRIRPVQGLLSARDF 242 Dictyostelium
AAP82284   173 PRVEYTAEEKATWGTVFRELKTLYPTHACREHNRVFPLLekycgYREDNIPQLEDISHYLQScTGFRLRPVAGLLSSRDF 252 zebrafish
12644139   174 PHVDYTKEEIETWGIIFRNLTKLYKTHACREYNHVFPLLvdncgFREDNIPQLEDVSNFLRDcTGFTLRPVAGLLSSRDF 253 fruit fly
XP_786152  148 PRVTYTEIEIKTWNTIFTNLSDLFKTHACQEFNYVFPLLmencgYQPDNIPQLEDVSNFLKDcTGFTLRPVAGLLSSRDF 227 purple urchin
1MMT_A      73 PRVEYMEEEKKTWGTVFKTLKSLYKTHACYEYNHIFPLLekycgFHEDNIPQLEDVSQFLQTcTGFRLRPVAGLLSSRDF 152 human
6PAH        59 PRVEYMEEEKKTWGTVFKTLKSLYKTHACYEYNHIFPLLekycgFHEDNIPQLEDVSQFLQTcTGFRLRPVAGLLSSRDF 138 human

Feature 2
*CpsB       94 FDFQGHGLLPIATDIRKVENIEYTPAPDIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAfstkeehdafeavrtltivke 173 CpsB
1DMW_A     138 LGGLAFRVFHCTQYIRHGSKPMYTPEPDICHELLGHVPLFSDRSFAQFSQEIGLASLGA---------------------- 196 human
AAD30115   322 LAGLAYRVFFCTQYVRHHADPFYTPEPDTVHELMGHMALFADPDFAQFSQEIGLASLGA---------------------- 380 nematode
1PHZ_A     255 LGGLAFRVFHCTQYIRHGSKPMYTPEPDICHELLGHVPLFSDRSFAQFSQEIGLASLGA---------------------- 313 Norway rat
XP_641959  243 LNGLAFRVFHATQYIRHPSVPLYTPEPDCCHELLGHVPLLADPDFADFSQEIGLASIGA---------------------- 301 Dictyostelium
AAP82284   253 LAGLAFRVFHSTQYIRHSSKPMYTPEPDICHELLGHVPLFADPNFAQFSQEIGLASLGA---------------------- 311 zebrafish
12644139   254 LAGLAFRVFHSTQYIRHPSKPMYTPEPDVCHELMGHVPLFADPAFAQFSQEIGLASLGA---------------------- 312 fruit fly
XP_786152  228 LAGLAFRVFHSTQYIRHPSKPMYTPEPDVCHELLGHVPLFADPKFAQFSQEIGLLSLGA---------------------- 286 purple urchin
1MMT_A     153 LGGLAFRVFHCTQYIRHGSKPMYTPEPDICHELLGHVPLFSDRSFAQFSQEIGLASLGA---------------------- 211 human
6PAH       139 LGGLAFRVFHCTQYIRHGSKPMYTPEPDICHELLGHVPLFSDRSFAQFSQEIGLASLGA---------------------- 197 human

Feature 3                                   #  #
*CpsB      174 sptstpdevtaaennvlekqklvSglsEAEQISRLFWWTVEYGLIGNIDT-------------PKIYGAGLLSSVGESKH 240 CpsB
1DMW_A     197 ------------------------Pd-eYIEKLATIYWFTVEFGLCKQGDS-------------IKAYGAGLLSSFGELQY 239 human
AAD30115   381 ------------------------Se-eDLKKLATLYFFSIEFGLSSDDAAdspvkengsnherFKVYGAGLLSSAGELQH 436 nematode
1PHZ_A     314 ------------------------Pd-eYIEKLATIYWFTVEFGLCKEGDS-------------IKAYGAGLLSSFGELQY 356 Norway rat
XP_641959  302 ------------------------Sd-eDIQLLSTCYWFTVEFGLCKEGDT-------------IRAYGAGILSSTGEMEH 344 Dictyostelium
AAP82284   312 ------------------------Pd-eFIEKLATVYWFTVEFGLCKQGNE-------------VKAYGAGLLSSFGELQY 354 zebrafish
12644139   313 ------------------------Pd-dYIEKLSTIFWFTVEYGVCRQEGE-------------LKAYGAGLLSSYGELEY 355 fruit fly
XP_786152  287 ------------------------Pd-eYIQKLATCYWFTVEFGLCRQNGQ-------------VKAYGAGLLSSFGELKY 329 purple urchin
1MMT_A     212 ------------------------Pd-eYIEKLATIYWFTVEFGLCKQGDS-------------IKAYGAGLLSSFGELQY 254 human
6PAH       198 ------------------------Pd-eYIEKLATIYWFTVEFGLCKQGDS-------------IKAYGAGLLSSFGELQY 240 human
```

```
Feature 4
*CpsB      241 CLTDaVEKVPFSIeACTSTTYDVTKMQPQLFVCESFEELTEA 282 CpsB
1DMW_A     240 CLSEkPKLLPLELeKTAIQNYTVTEFQPLYYVAESFNDAKEK 281 human
AAD30115   437 AVEGsATIIRFDPdRVVEQECLITTFQSAYFYTRNFEEAQQK 478 nematode
1PHZ_A     357 CLSDkPKLLPLELeKTACQEYSVTEFQPLYYVAESFSDAKEK 398 Norway rat
XP_641959  345 FLTDkAKKLPFNPfDACNTEYPITTFQPLYYVAESFQKAKEQ 386 Dictyostelium discoideum AX4
AAP82284   355 CLTNePKLQPFEPeKTCQQKYPITEFQPVYFVAESFEDAKEK 396 zebrafish
12644139   356 CLTDkPQLKDFEPeVTGVTKYPITQFQPLYYVADSFETAKEK 397 fruit fly
XP_786152  330 CLSDePEIRPFDPeKTAVTDYPVTKFQPIYYLAESFEDAKEK 371 purple urchin
1MMT_A     255 CLSEkPKLLPLELeKTAIQNYTVTEFQPLYYVAESFNDAKEK 296 human
6PAH       241 CLSEkPKLLPLELeKTAIQNYTVTEFQPLYYVAESFNDAKEK 282 human
```

Fig. S10(b): Multiple sequences alignments between Phe4MO's from human, nematodes, rat, zebrafish fruit fly, purple urchin and CpsB.

Supplementary material 11a

List of protein-metal interactions

Fig. S11(a): Lists of hydrogen bonds in CpsB structure

	ATOM 1						ATOM 2					
S/N	**Atom No.**	**Atom name**	**Res name**	**Res No.**	**Chain**		**Atom No.**	**Atom name**	**Res name**	**Res No.**	**Chain**	**Distance**
1	939	NE2	HIS	124	A	<->	2266	FE	FE	1	-	2.02
2	948	OE2	GLU	125	A	<->	2266	FE	FE	1	-	3.33
3	972	NE2	HIS	129	A	<->	2266	FE	FE	1	-	2.00
4	1645	OE2	GLU	214	A	<->	2266	FE	FE	1	-	1.83

Supplementary material 11b

Fig. S11(b): Lists of non-bonded contacts in CpsB structure

	ATOM 1						ATOM 2					
S/N	**Atom No.**	**Atom name**	**Res name**	**Res No.**	**Chain**		**Atom No.**	**Atom name**	**Res name**	**Res No.**	**Chain**	**Distance**
1	937	CD2	HIS	124	A	<->	2266	FE	FE	1	-	3.07
2	938	CE1	HIS	124	A	<->	2266	FE	FE	1	-	2.98
3	939	NE2	HIS	124	A	<->	2266	FE	FE	1	-	2.02
4	948	OE2	GLU	125	A	<->	2266	FE	FE	1	-	3.33
5	970	CD2	HIS	129	A	<->	2266	FE	FE	1	-	2.97
6	971	CE1	HIS	129	A	<->	2266	FE	FE	1	-	3.04
7	972	NE2	HIS	129	A	<->	2266	FE	FE	1	-	2.00
8	1643	CD	GLU	214	A	<->	2266	FE	FE	1	-	2.47
9	1644	OE1	GLU	214	A	<->	2266	FE	FE	1	-	2.53
10	1645	OE2	GLU	214	A	<->	2266	FE	FE	1	-	1.83

CHAPTER SIX

This chapter has been ***submitted to the Journal: International Journal of Biological Macromolecules (IJBM)***

Biochemical and Structural Characterization of a Pentachlorophenol Hydroxylating Cytochrome P450 Monooxygenase from *Bacillus tropicus* strain AOA-CPS1

Oladipupo A. Aregbesola, Ajit Kumar, Mduduzi P. Mokoena and Ademola O. Olaniran*

Abstract

In this study, a 1236 bp gene encoding 47.1 kDa putative cytochrome P450 monooxygenase (P450CPS1) from the genome of a newly isolated pentachlorophenol (PCP) degrading *Bacillus tropicus* strain AOA-CPS1 (*Bt*AOA) was amplified, cloned, overexpressed, purified to homogeneity and characterized. The enzyme was purified to 1.4-fold with a total yield of 64.8% and specific activity of 607 $U \cdot mg^{-1}$ of protein. The enzyme showed optimum activity at pH 7.5 and 40°C, exhibiting >90% residual activity between pH 7.0-9.0 for 3 h while 100% residual activity was observed after 4 h between 25°C-30°C. P450CPS1 enzymatic activity was enhanced by the presence of Mn^{2+}, Fe^{2+} and Fe^{3+} in the reaction mixture but inhibited by piperonyl butoxide (1.5 mmol) and SDS (2 mmol). The reaction kinetics studies showed allosteric nature of P450CPS1 showing apparent v_{max}, K_m, k_{cat} and k_{cat}/K_m values of 0.069 $\mu mol \cdot s^{-1}$, 200 µmol, 0.011 s^{-1} and 5.42×10^{-5} $\mu mol^{-1} \cdot s^{-1}$, respectively, for the substrate PCP and 0.385 $\mu mol \cdot s^{-1}$, 56.46 µmol, 0.06 s^{-1} and 1.77×10^{-3} $\mu mol^{-1} \cdot s^{-1}$, respectively, for co-substrate NADH. The in-gel trypsin digestion experiments and bioinformatics tools confirmed the enzyme as cytochrome P450. This study provides experimental evidence of the involvement of a putative cytochrome P450 in *Bt*AOA in PCP biodegradation.

Keywords: Cytochrome P450 monooxygenase; *Bacillus tropicus* AOA-CPS1; Pentachlorophenol.

6.1. Introduction

Cytochrome P450 monooxygenases (CYPs) are found to be involved in detoxification of numerous recalcitrant compounds [1,2] and hydrolyse the inactivated/unreactive C–H bonds [3]. CYPs are undoubtedly one of the most promiscuous but adaptable biological catalysts with unique capacity to catalyse stereo-, chemo- and regiospecific oxidation/hydroxylation of a wide range of chemicals at a mild environmental and/or enzymatic condition [4]. These enzymes are specifically involved in hydroxylation, dehalogenation, epoxidation, de-alkylation, sulfoxidation, deamination, desulfuration and N-oxide reduction of many xenobiotics [5]. The initial step in the transformation of recalcitrant compounds, oxidation by oxygenase's, is common in both eukaryotes (fungi and animals) and prokaryotes (bacteria) [6,7]. The catalytic mechanism of oxidation of various compounds involve the reductive activation of molecular oxygen (O_2) to form a highly oxidized iron-oxo-complex [8].

This initial and rate-limiting oxidation step in the biotransformation of chemicals is catalysed by CYPs (mainly in eukaryotes) or CYPs and non-heme iron dioxygenases (mainly in prokaryotes) [5]. For example, hydroxylation of pentachlorophenol (PCP) at para-position is the first step involved in PCP degradation by many organisms [9,10]. This rate limiting step in PCP biodegradation is catalysed by PCP 4-monooxygenase [9] or fungi cytochrome P450 monooxygenase [10].

Bacterial CYPs perform prominent roles in biosynthesis and biodegradation pathways of wide ranges of secondary metabolites and diverse chemicals [3]. In the hydroxylation repertoire, bacteria CYPs are involved in the hydroxylation of fatty acid [11,12], terpenes [13,14], steroids [15,16], marine natural products [17,18], peptide [19] and heteroatom oxidation [20,21]. However, the major challenges in industrial applications of CYPs are instability, poor expression levels, low catalytic activity, expensive cofactor requirements, limited solvent tolerance, uncoupling between NADH oxidation, electron supply and product formation [4,22]. Fungal CYPs have been explored in the detoxification of organochlorine pesticides, such as 2,4-dichlorophenoxyacetic acid [23,24] and PCP transformation [10].

However, despite the wide range of reported xenobiotics degradation by bacterial CYPs, to the best of our knowledge, the exploitation of the catalytic promiscuity of CYPs in the biotransformation of chlorophenols such as PCP and other organochlorine pesticides has not

been explored. Investigation of PCP biotransformation by a recently isolated indigenous strain *Bacillus tropicus* strain AOA-CPS1 (*Bt*AOA), indicated that the strain can efficiently degrade PCP. The study further investigated the genes encoding the enzymes involved in the PCP degradation pathway in *Bt*AOA. Indirect experimental approaches and whole genome sequence of *Bt*AOA revealed the presence of genes involved in PCP degradation pathway (unpublished data).

The whole genome annotation data (GeneBank accession no: CP049019.1) did not indicate the presence of any PCP dehalogenase or PCP-4-monooxygense, therefore, the database of *Bt*AOA genome was searched for the presence of cytochrome P450 monooxygenase. The search results indicated the presence of a CYP in *Bt*AOA genome (GeneBank accession no: QIE37123.1). To explore the possible role of CYP in PCP degradation pathway in *Bt*AOA, the gene was cloned; the recombinant protein was overexpressed, purified, and characterized.

Further, the identity of the purified enzyme was confirmed using the bioinformatics tools and homology modelling. Hence, present manuscript reports the cloning, characterization, and structural homology modelling of PCP hydroxylating CYP (P450CPS1 in this manuscript) from an indigenous strain *Bt*AOA. Also, CYP's annotated in the whole genome of most of the *Bacillus tropicus* strains remains at the prediction level without proving their biological function. Therefore, this study reports on experimental evidence for the existence of a putative cytochrome P450 monooxygenase from *Bt*AOA and elucidated its role in PCP biodegradation.

6.2. Materials and methods

6.2.1. Materials

Pure pentachlorophenol (PCP, 98%), Na-azide (≥99%), NADH (≥97%), 2-Mercaptoethanol (≥99%), DTT (≥99.5%), EDTA (99.4%), SDS and PMSF (≥98.5%) were purchased from Merck (NJ, USA). Stock solutions (10 mmol) of PCP was prepared in 0.5 N NaOH and diluted with 50 mmol sodium phosphate (Na_2HPO_4-NaH_2PO_4) buffer, pH 7.0.

6.2.2. Isolation and identification of *Bacillus tropicus* strain AOA-CPS1 (*Bt*AOA)

The isolation and identification of *Bt*AOA is reported previously [25,26] Further, the whole genome data was generated (Inqaba Biotech, Pretoria, South Africa) using a combination of Sequel II System, PacBio Single-Molecule Real-Time (SMRT Link Version 7.0.1.66975) sequencing technology (Moine-Scientist and Applications Support, 2019), FALCON assembler and Hierarchical Genome Assembly Process 4 (HGAP4) de novo assembly analysis application. This Whole Genome Shotgun project of *Bt*AOA has been deposited in GenBank under accession number CP049019 (version CP049019.1) (unpublished data).

6.2.3. PCR amplification and cloning of cytochrome P450 monooxygenase (P450CPS1) gene

To amplify P450CPS1 from *Bt*AOA, PCR was performed using genomic DNA of the isolate as a template and the primer pair Forward 5'-GAGGGATCCCATGTCAATGAAAAACAA AGT-3' and reverse: 5'-TATGGATCCGAAAGTTAAAGGCAATTCC-3'). To design the primer pair, Whole Genome Shotgun project of *Bacillus tropicus* strain AOA-CPS1 deposited in GenBank under accession number CP049019 (version CP049019.1) was searched for the presence of cytochrome P450 monooxygenase gene. The whole-genome data showed the presence of putative cytochrome P450 monooxygenase (EC:1.14.14.1, gene locus_tag=GM610_09540) gene sequence with an accession no: QIE37123.1. The gene sequence was retrieved, and primers were synthesized to amplify the full-length gene with restriction sites *Not*I and *Bam*HI (underlined) targeting expression vector pET15b for cloning purpose. The in-frame cloning of P450CPS1 and conformation of recombinant pET15b-P450CPS1 plasmid was done as described previously [25,26], except the annealing temperature used for PCR was set at 58°C.

6.2.4. Overexpression and purification of P450CPS1

The transformation of *E. coli* BL21(DE3) with pET15b-P450CPS1, the overexpression of the enzyme and purification was done as described previously [25,26]. The recombinant 6 × His-tagged P450CPS1, 2.5 mL (≅42.5 mg total protein) of concentrated supernatant (cell lysate) was loaded in 5 mL HisPur Cobalt resin column (Thermo Fisher Scientific, USA) and purified as describes previously [25,26]. Thrombin Cleavage Capture Kit (Cat. No. #69022, Merck Millipore, NJ, USA) was used to remove 6 × His-tag. 12% SDS-PAGE [27] was used

to confirm the expression of P450CPS1, homogeneity of the purified protein and molecular mass. The enzymatic activity of P450CPS1 at each purification steps was determined as described below.

6.2.5. Enzyme activity assay

P450CPS1 activity was determined as described previously [9], with some modifications. One mL reaction mix contained 100 µM PCP, 160 µM NADH, 50 mM NaH_2PO_4-Na_2HPO_4 buffer (pH 7.0), 6.0 mmol 2-Mercaptoethanol and 300 µg P450CPS1 (6.36 µmol final concentration). 2-Mercaptoethanol was added to the reaction of PCP with P450CPS1 to trap Tet-CBQ formed as 2,3,5,6-tetrakis[(2-hydroxyethyl)thio]-1,4-hydroquinone (THTH). Transformation of PCP, oxidation of NADH and formation of THTH were monitored spectrophotometrically at wavelengths 320 nm, 340 nm, and 350 nm, respectively. The NADH was included in the blank reactions to prevent interference with the absorption peak of PCP. The absorbance was converted to molar concentrations of PCP, NADH and THTH by using the molar extinction coefficients of 935 $mol^{-1} \cdot cm^{-1}$ [28], 6200 $mol^{-1} \cdot cm^{-1}$ and 2175 $mol^{-1} \cdot cm^{-1}$ [9], respectively.

6.2.6. Determination of optimum pH, temperature and pH, temperature stability of purified P450CPS1

The optimum pH and pH stability as well as optimum temperature and thermal stability of purified P450CPS was determined as described previously [25,26] except 55 µL (300 µg, 6.36 µmol final concentration) of P450CPS1 was used all the assays. Small modifications are explained the figure legends.

6.2.7. Determination of kinetic parameters

To determine the kinetic parameters for the enzymatic reaction of P450CPS1 with substrate PCP, the purified P450CPS1 (300 µg, 6.37 µM final concentration) was incubated with initial concentrations of PCP (20, 40, 60, 80, 100, 120, 140, 160, 180 and 200 µmol) in the presence of 160 µmol NADH and the activity was determined at optimum conditions as described above. The purified P450CPS1 (300 µg, 6.37 µM final concentration) enzyme activity was also determined by incubating the enzyme with varying concentrations of NADH (20, 40, 60, 80 and 100 µmol) in the presence of 200 µmol PCP and the activity was determined at optimum conditions as described above. The values for K_m, v_{max}, k_{cat} and catalytic efficiency

(k_{cat}/K_m) were determined as described previously [25,26]. The Hill coefficient n_H, measuring homotropic cooperative effects, was estimated by applying Hill approximation equation [29–31]. The allosteric properties (M-W-C modal) and steady state reaction mechanism of the enzyme were analysed as described previously [32–34].

6.2.8. Effects of metal ions and inhibitors on P450CPS1 activity

Purified P450CPS1 (100 µg) was pre-incubated with varying concentrations metal ions inhibitors for 10 min. The enzyme assays were performed as described above [35]. Table footnote are referred for additional information.

6.2.9. In-gel trypsin digestion and identification of the purified P450CPS1 in ES-MS

The experimental procedures for In-gel trypsin digestion and identification of the purified P450CPS1 in ES-MS were followed as reported previously [25,26,36].

6.2.10. Template-based structure prediction and homology modelling for P450CPS1

Three-dimensional structure and homology modelling of P450CPS1 were predicted as described previously [25,26,37–39].

6.2.11. Evolutionary relationships of P450CPS1 with another cytochrome P450 monooxygenases

The evolutionary relatedness of P450CPS1 with other cytochrome P450 monooxygenases from P450 superfamily, the procedures were followed as described previously a phylogenetic based evolutionary analysis was performed using the Neighbour-Joining method [25,26,40–43].

6.2.12. The nucleotide sequence submission

This Whole Genome Shotgun project of *Bacillus tropicus* strain AOA-CPS1 has been deposited in GenBank under accession number CP049019 (version CP049019.1) (unpublished data). The putative cytochrome P450 monooxygenase gene sequence was submitted with an accession no: QIE37123.1 (EC:1.14.14.1, gene locus_tag=GM610_09540).

6.3. Results and discussion

6.3.1. Identification of *Bacillus tropicus* strain AOA-CPS1 (*Bt*AOA)

Initially, the isolate was identified as *Bacillus cereus* strain AOA-CPS1 (*Bc*AOA) based on the 16S rDNA sequence analysis (submitted to NCBI as accession number MH504118.1). However, a quality control test by NCBI for the submitted whole genome sequence of the strain, using an average nucleotide identity (ANI), which compares the submitted genome sequence against the whole genomes of the type strains that are already in GenBank [44,45], resulted in the renaming of *Bc*AOA as *Bacillus tropicus* strain AOA-CPS1 (*Bt*AOA). The ANI analysis indicated that the genome sequences of *Bc*AOA are 96.61% identical to the genome of the type strain of *Bacillus tropicus*, with 89.9% coverage of the genome. Consequently, *Bc*AOA was renamed as *Bt*AOA (based on the whole genome data submitted at NCBI under accession number CP049019).

6.3.2. Cloning, overexpression and purification of 6 × His-tagged P450CPS1

A 1236 bp gene fragment encoding P450CPS1 was amplified (Fig. 1A), cloned into an expression vector pET15b, and overexpressed in the expression host *E. coli* BL21(DE3) (Fig. 1B). The recombinant 6 × His-tagged P450CPS1 protein was purified to homogeneity and showed a single band of 47.1 kDa (Fig. 1B). The size of the P450CPS1 from *Bt*AOA is within the range (45-60 kDa) of monomeric P450 reported previously [46–48]. A single step purification strategy resulted in 1.4-fold purification of the enzyme with a total yield of 64.8% (Table 1). The total enzyme activity and specific activity of P450CPS1 at each purification step is shown in Table 1. The P450CPS1 gene sequence shared over 99% sequence homology with multispecies cytochrome P450 monooxygenases (CYPs) from another *Bacillus* spp. (Fig. S1) which are yet to be studied for their functional properties (but only predicted).

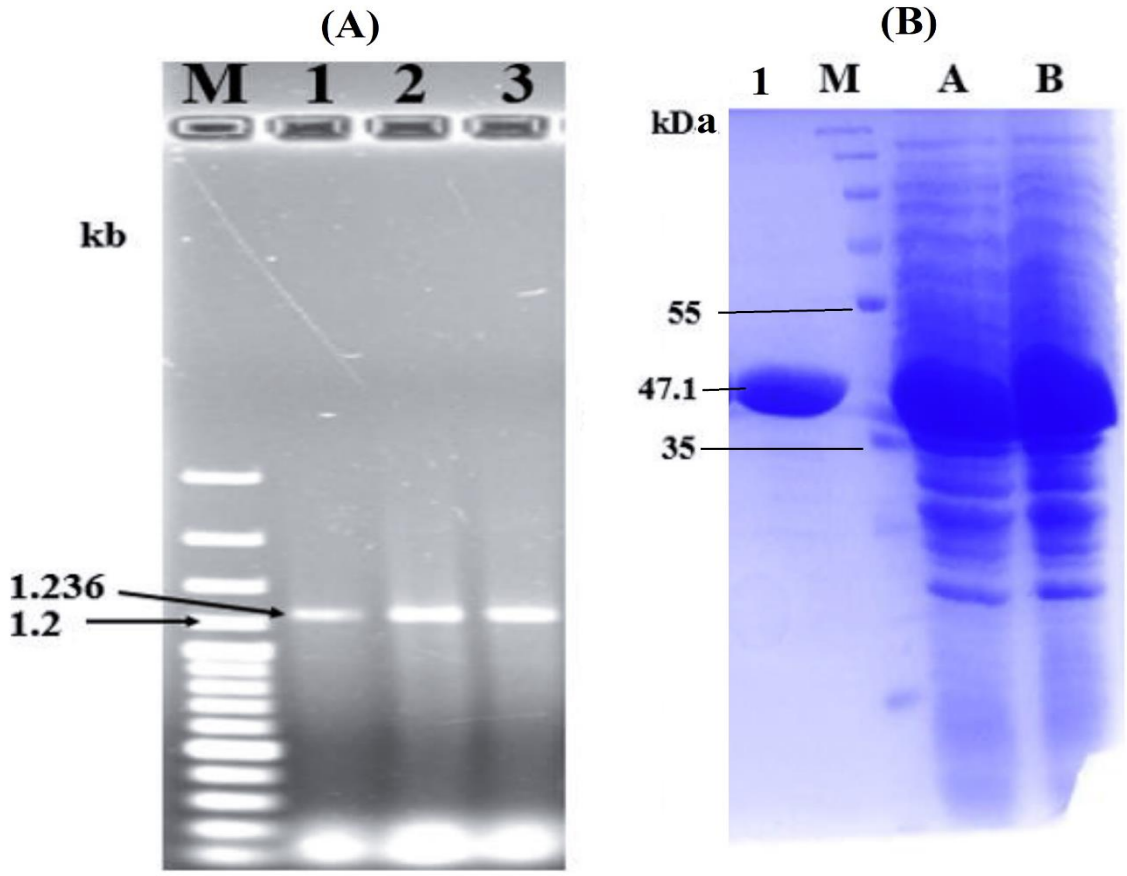


Fig. 1: Amplification, cloning and purification of P450CPS1. (A), Amplification of *P450CPS1* gene from *Bt*AOA, DNA marker (Lane M), amplified *P450CPS1* gene (Lanes 1-3); **(B),** 12% SDS-PAGE showing overexpressed P450CPS1, Protein marker (Lane M), cell lysate (Lane A, B), HisPur Cobalt Resin fractions of the purified recombinant 6 × His-tagged P450CPS1 (Lanes 1). ≅47.1 kDa molecular weight indicated is based on the biophysical properties calculated at ProtPram tool at ExPASy for P450CPS1 amino acid sequence.

Table 1: Purification scheme of P450CPS1 from *Bt*AOA.

Purification step	Total Protein (mg)	Protein ($mg \cdot mL^{-1}$)	Total activity (U*)	Activity ($U \cdot mL^{-1}$)	Specific activity ($U \cdot mg^{-1}$)	Purification (fold)	Yield[c] (%)
Crude (cell lysate)	179.22[a]	17.00[c]	66.51[d]	7390.00	434.71	1.0	100
HisPur Cobalt column	27.54[b]	2.75	91.747[e]	1668.12	606.59	1.4[f]	64.8[g]

[a]From 100 mL of crude *E. coli* cell lysate. [b]Total protein in 10 mL of eluted fractions pooled together after loading 42.5 mg of total protein in the column. [c] after concentrating the eluted fractions pooled together using ultrafiltration and reducing the volume to 10 mL. [d]in ≅9 µL (300 µg total protein) cell lysate. [e]in ≅55 µL fraction (300 µg total protein) showing single band on SDS-PAGE (Fig 1B, Lane 1). [f]from 606.59/434.71. [g]Yield (%) = (27.54 mg/42.5 mg) ×100. *One unit of P450CPS1 activity was defined as the 1 µmole of THTH produced per min under standard assay conditions.

6.3.3. Optimum pH and pH stability of P450CPS1

The enzyme showed optimum activity at pH 7.5 but was active at slightly acidic to wide ranges (6.5-11.0) of pH (Fig. 2A). The activity was very low at pH 5.0-6.0 but increased exponentially at pH 6.5 and reached the optimum at pH 7.5. The optimum pH of P450CPS1 is similar to previously reported Cytochrome P450 102A2 [49,50]. Enhance steroid hydroxylation by P450 enzyme in the presence of alkaline buffer has also been reported [51]. The enzyme was stable at a wide range of alkaline pH, retaining more than 90% of its residual activity between pH 7.0-11.0 for 90 min (Fig. 2B). The enzyme maintained 100% residual activity at pH 7.0-9.0 for 180 min. However, the residual activity of P450CPS1 decreased to about 76% and 54% at pH 10.0 and pH 11.0, respectively, after 180 min. However, the enzyme activity was completely lost at pH 12.0 after 150 min. The ability of other CYPs to hydroxylates recalcitrant compounds in wide ranges of pH had also been reported previously [52].

6.3.4. Optimum temperature and temperature stability of P450CPS1

The P450CPS1 exhibited good activity between 25°C-60°C (Fig. 2C) with optimum activity at 40°C which is similar to previous report [52]. A slight decrease of 4.10% and 2.21% in P450CPS1 activity was observed at 35°C and 45°C, relative to the activity at 40°C which is similar to that of cytochrome P450 from *B. subtilis* strain 102A2 which is involved in SDS biodegradation [49]. The enzyme was found to be stable between 25°C-30°C, retaining 100% of its residual activity after 240 min. The enzyme was 100% stable for 180 min at 35°C, but activity decreased to about 84% after 240 min (Fig. 2D). Also, P450CPS1 retained >50% of its activity between 45°C-50°C after 240 min, in agreement with previous report [53].

The enzyme however lost 100% activity at 60°C within 30 min (data not shown), also in accordance with previous report [49]. The complete denaturation of P450CPS1 at 60°C and the mesophilic nature of its catalytic activity further buttress the earlier reports that most mesophilic (regular) cytochrome P450s denature around 50°C to 60°C [54]. The optimum temperature and stability of CYPs mostly depends on the environmental factors and the habitat of their host organisms [55]. Most natural environments are comparatively mild but most P450s enzymes have not been characterized in nature for their ability to tolerate temperatures above 40°C [56].

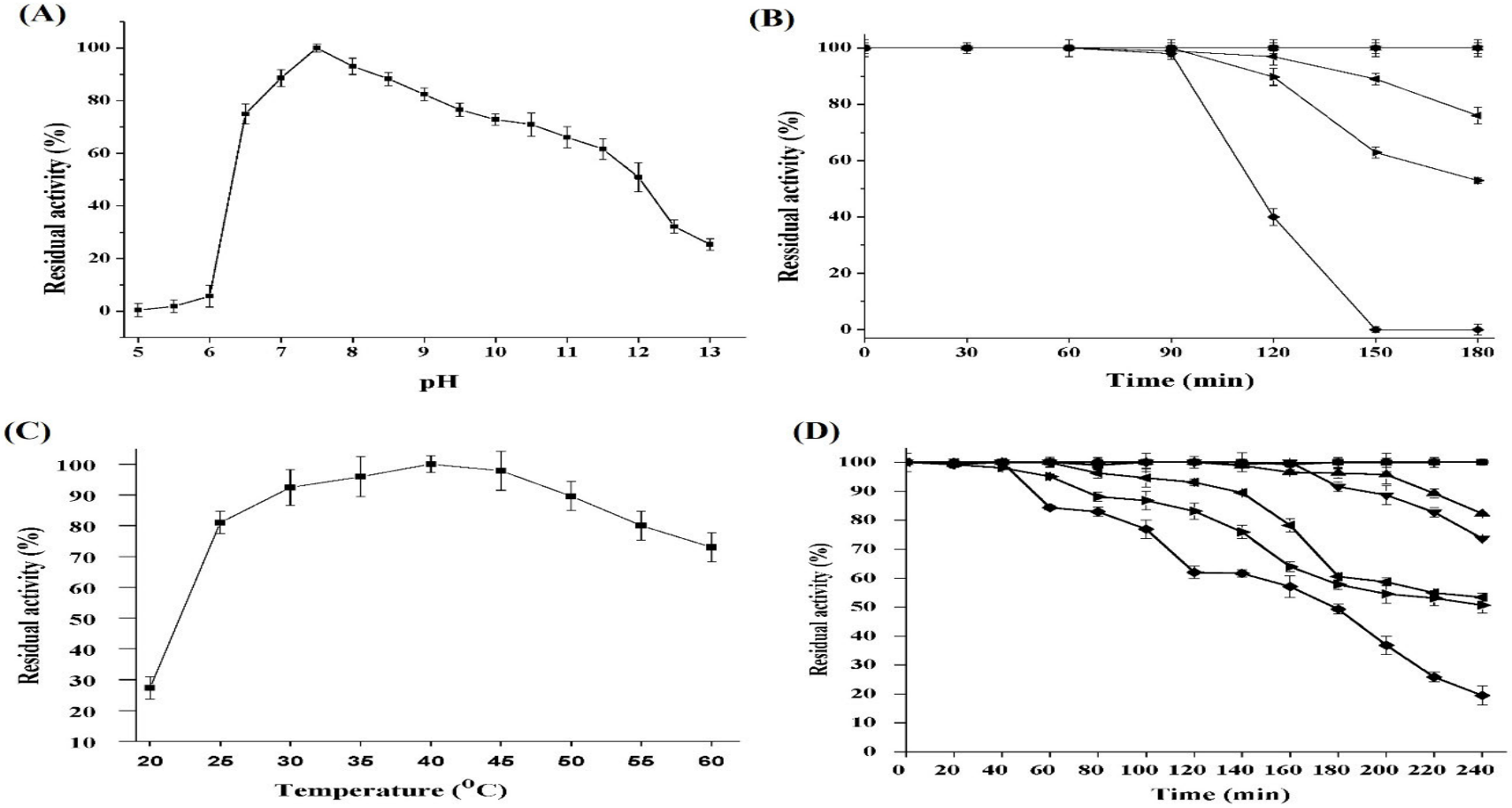


Fig. 2: Characterization of P450CPS1 from *Bt*AOA. (A), Optimum pH; **(B)**, Functional stability of P450CPS1 at pH 7.0 (●), 8.0 (▲), 9.0 (▼), 10.0 (◄), 11.0 (►) and 12.0 (♦), over time; **(C),** Optimum temperature; **(D),** Functional stability at 25 ºC (■), 30 ºC (●), 35 ºC (▲), 40 ºC (▼), 45 ºC (◄), 50 ºC (►) and 55 ºC (♦), over time.

6.3.5. Effect of metal ions and inhibitors on P450CPS1 activity

P450CPS1 exhibited higher activity in the presence of Fe^{2+}, Ca^{2+}, Mn^{2+} and Fe^{3+} as compared to the activity in buffer only (Table 2). Metal ions enhanced substrate binding and formation of the heme active centre of the CYPs compared to P450 devoid of metal ions (Manna and Mazumdar, 2008). Unlike, other CYPs, P450CPS1 activity was stimulated by both Fe^{2+} and Fe^{3+} ions confirming the earlier report that CYPs are present in both ferrous (Fe^{2+}) and Ferric (Fe^{3+}) forms in a resting state within cells and hepatocytes [57].

In addition, non-heme iron is mainly in the ferric state and substrate binding to the Fe^{3+} resting state facilitates the first electron reduction of the heme [57]. However, partial inactivation of P450 in the presence of Ca^{2+} and Mg^{2+} have also been reported [58], meaning metal ions may only stimulate P450 activity in a concentration dependent manner.

Also, previous studies have shown that both Fe^{2+} and Fe^{3+} forms of P450's were expressed heterologous in the absence of substrates, by bacterial (strains CYP176A1 and CYP101A1) and mammalian (CYP1A2, CYP2C19, CYP2A6, CYP3A4 and CYP2C9) CYPs [57]. Cytochrome P450 generally require a cofactor as a source of electrons, which are normally acquired from reduced NAD(P)H.

However, P450 enzymes cannot acquired electrons directly from the cofactors, they require a redox shutter(s) (partner proteins) to shuttle the electron(s) to the P450 heme iron [59,60]. The activity of P450CPS1 was estimated in the presence of different concentrations of inhibitors and results shown in Table 3. The activity of P450CPS1 was stimulated by 0.5 - 2.0 mM phenylhydrazine (PHZ), Na-azide, DTT, PMSF and EDTA. PHZ stimulated the activity of P450CPS1 by about 11 folds which is higher than the previous report [61].

Also 0.5-1.5 mmol piperonyl butoxide (PBO) stimulated the activity of P450CPS1 while 2.0 mmol of the same compound completely inhibited the activity of the enzyme. However, only 1.5 mmol of SDS stimulated the activity of P450CPS1. Previous report also showed that 2.0 mmol PHZ did not have inhibitory effect on P450CPS1 activity.

Another finding also demonstrated PHZ-mediated induction of heme oxygenase activity by CYP in rat liver and kidney [62]. Although PBO is a known inhibitor of CYP's activity [61],

0.5-1.5 mM PBO was found to stimulate the activity of P450CPS1 in this study, in agreement with the previous report [61].

The observed complete inhibition of P450CPS1 activity by 2.0 mmol of PBO in this study agrees with another report on PCP transforming CYP from *Phanerochaete chrysosporium* [10]. Also, increase in P450 enzyme in the presence of PMSF has also been report [63,64] in according with the finding of this report. The finding of this report also indicated that P450CPS1 is insensitive to the protease inhibitors used except the PBO that is specific for P450s and it is on inhibitory to P450CPS1 at high concentrations.

Table 2: Effects of metal ions on P450CPS1 from *Bt*AOA

Metal ions	Residual activity (%)
Control	100
Fe^{2+}	149.22 ± 4.85
Ca^{2+}	155.55 ± 2.13
Cu^{2+}	0.00
Mg^{2+}	31.579 ± 4.39
Pb^{2+}	0.00
Ni^{2+}	0.00
Co^{2+}	83.23 ± 2.94
Mn^{2+}	181.268 ± 1.58
Na^{2+}	22.68 ± 2.03
Hg^{2+}	0.00
Fe^{3+}	137.49 ± 5.36

Control is the enzyme reaction without metal ion

Table 3: Residual activity (%) of P450CPS1 in the presence of different concentrations of inhibitors.

Inhibitor (mM)	EDTA	DTT	Na-Azide	PMSF	SDS	PBO	PHZ
0.5	149.17±0.4	436.81±4.8	98.67±0.1	391.01±0.4	97.89±0.5	419.70±0.4	815.64±0.4
1.0	259.81±0.9	299.08±0.9	541.94±0.0	949.53±0.9	89.98±5.1	206.74±0.3	1088.36±0.9
1.5	203.81±0.9	256.81 ±0.4	712.36±0.4	952.79±0.4	148.18±0.9	119.94±0.9	1148.48±0.9
2.0	163.84±0.02	134.74±0.2	738.41±0.2	798.61±0.2	86.24±0.2	0.00±0.00	167.36±0.9

The residual activity of assay without inhibitor was recorded as 100%, Phenylhydrazine (PHZ), piperonyl butoxide (PBO); dithiothreitol (DTT), phenylmethylsulfonyl fluoride (PMSF),

6.3.6. Steady-state kinetic parameters of P450CPS1

The enzyme substrate reaction exhibited pre-steady state at low PCP concentrations (20-140 µmol) but showed a steady state at high PCP concentrations (140-200 µmol), exhibiting the S-shaped curve (Fig. 3A). The Lineweaver-Burk double reciprocal plot was not found to be linear but fits a parabola. It becomes linear, however, if $1/[PCP]^2$ is plotted against $1/v^2$ resulting in calculated v_{max}, K_m, k_{cat} and k_{cat}/K_m values of 0.069 µmol·s^{-1}, 200 µmol, 0.011 s^{-1} and 5.42 × 10^{-5} µmol^{-1}·s^{-1}, respectively (Fig. 3B). The Lineweaver-Burk double reciprocal plot for $1/v$ versus 1/[NADH] was fit linearly exhibiting the v_{max}, K_m, k_{cat} and k_{cat}/K_m values of 0.385 µmol·s^{-1}, 56.46 µmol, 0.06 s^{-1} and 1.77 × 10^{-3} µmol^{-1}·s^{-1}, respectively (Fig. 3C). The results indicate that optimum PCP and NADH concentrations for steady state kinetics of PcpB substrate reaction must be above 200 µmol and 100 µmol, respectively, for each. The catalytic efficiency of P450CPS1 for PCP and NADH when used individually as variable substrates were found to be higher than that of CYP from strain 102A2 [49]. Also, the K_m and v_{max} of the enzymes using both substrates were similar to those of fatty acids hydroxylating P450 monooxygenase CYP102A7 from *B. licheniformis* [12] and P450BM3 (CYP102A1) from *B. megaterium* [65], but slightly different from that of bisphenol degrading CYP from *Sphingomonas* sp. strain AO1 [66] and strain CYP505D6 involved in the hydroxylation of various xenobiotics [1].

Plotting the enzyme substrate and enzyme co-substrate data as Hill plots provided additional useful information about reaction kinetics. The Hill plot (Fig. 4A) was nonlinear over the range of substrate concentrations studied, showing two prominent straight lines, one from the substrate range of 20-120 µmol (Fig. 4B) and second from 120-200 µmol (Fig. 4C). The plots were used to calculate $K_{0.5}$, considered as K_m [67] and showed that the enzyme has two different K_m values i.e. 58 µmol (Fig. 4B) at low substrate concentrations (20-120 µmol) and 186 µmol (Fig. 4C) at high substrate concentrations (100-200 µmol). These values coincide with the values calculated from the Lineweaver-Burk plots. The results indicate that P450CPS1 has low binding affinity at low substrate range, but suddenly show burst in product formation due to the high binding affinity at high substrate concentrations, indicating primarily its allosteric nature, and has homotropic cooperative interaction effect. The Hill plot (Fig. 4D) for NADH interaction with P450CPS1 suggest equal affinity at different reaction conditions and may have heterotropic cooperative interaction effects.

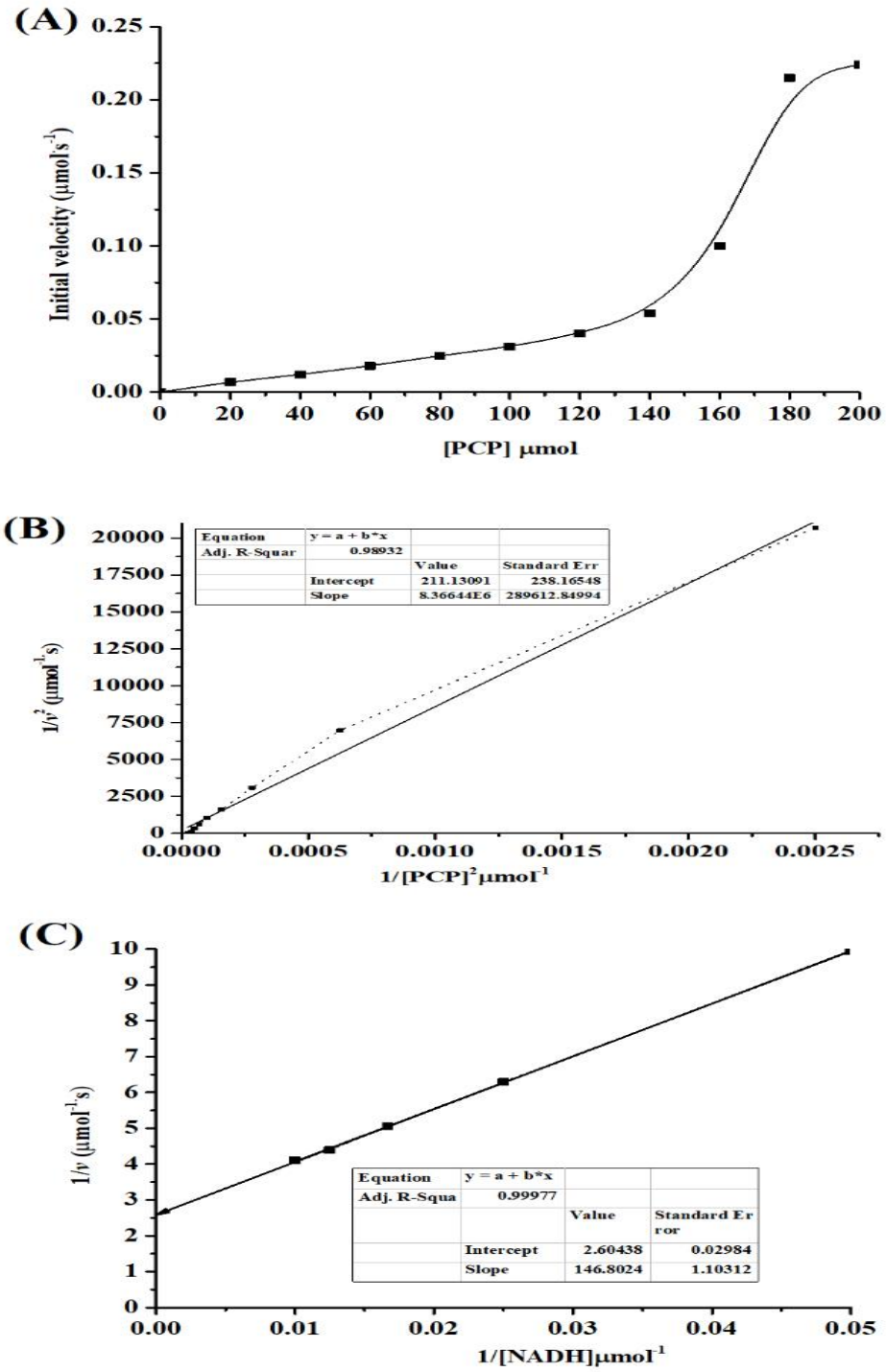


Fig: 3: **Steady-state kinetic parameters of P450CPS1.** Initial velocity of P450CPS1 (300 µg) reaction with PCP at different initial concentration (20-200 µmol) in the presence of 160 µmol NADH**. (B)**, The double reciprocal plot of data shown in 3A, $1/v^2$ *vs* $1/[PCP]^2$. **(C)** the double reciprocal plot of initial velocity of P450CPS1 (300 µg) reaction with PCP (200 µmol) in presence of different NADH initial concentration (20, 40, 60, 80 and 100 µmol).

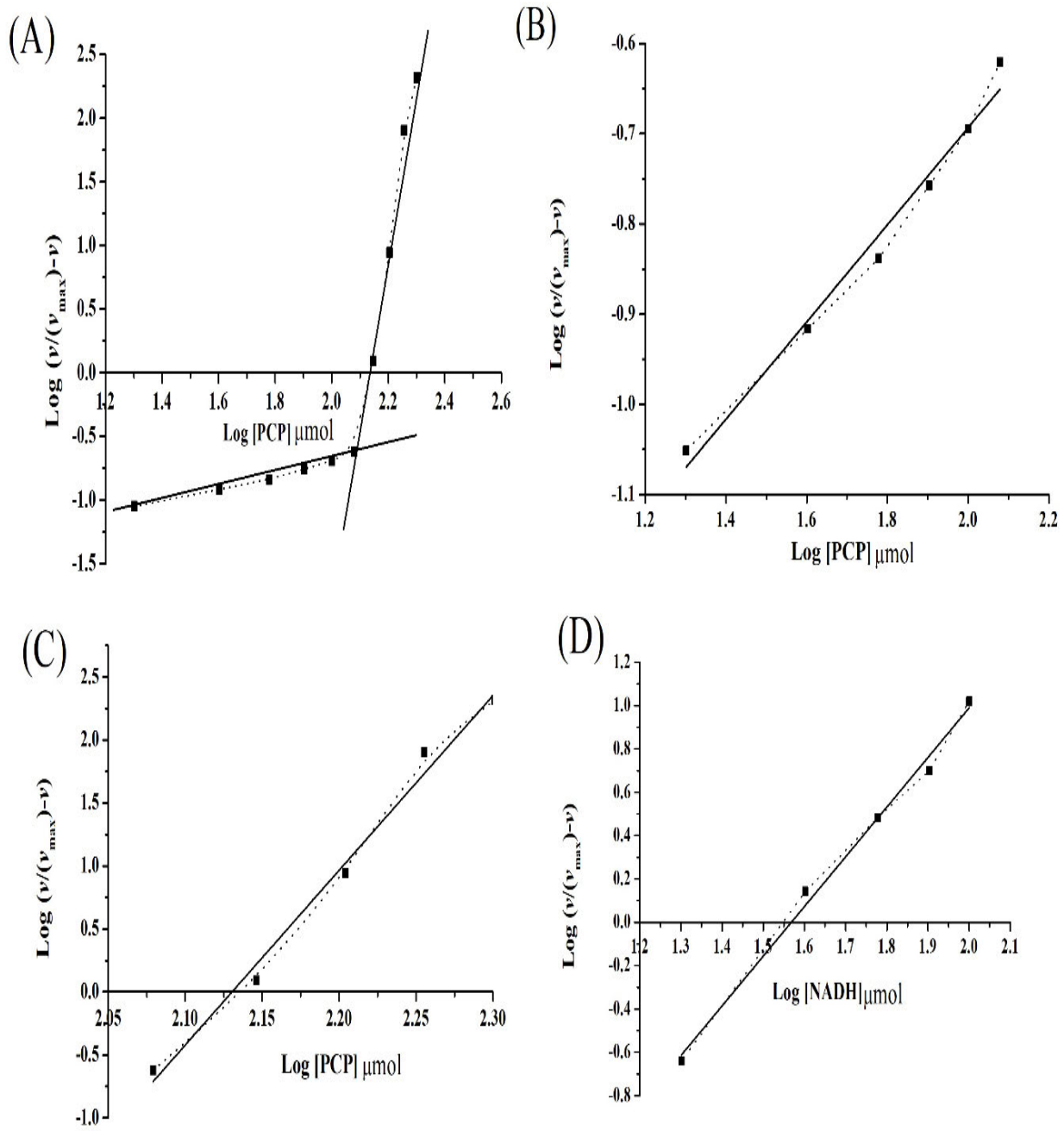


Fig. 4: Enzyme reaction kinetics using Hill plots. (A), Hill plot for the data shown in 3A. **(B)**, Hill plot for the data shown in 4A, but only for the PCP at low [PCP] (20-120 µmol). **(C)**, Hill plot for the data shown in 4A, but only at high [PCP] 120-200 µmol. **(D),** Hill plot for the data shown in 3C.

6.3.7. Allosteric properties of P450CPS1

The enzyme-substrate interaction plots (Fig. 3A and 4A) confirmed the allosteric nature of P450CPS1. The curve was significantly found to be sigmoidal in shape indicating slow formation of product at first followed by a sudden increase before reaching saturation. The Hill plot for the co-substrate NADH was found to be linear over a wide range of concentrations indicating that P450CPS1 has equal affinity all the time for NADH. As previously reported, the Hill plot [29-31] is used to analyse allosteric effects where slopes values (n_H) indicate the number of interacting sites and equals the number of binding sites for the ligand when the interaction is very strong [67].

The results shown in Fig. 4B shows n_H=0.53, indicate that all the binding sites are not occupied at low substrate concentrations, but n_H=10.83 indicates that most of the binding sites are occupied and has homotropic substrate-substrate interaction. This also shows that P450CPS1 may exists in multimeric form while reacting with the substrate. The value of n_H=2.33 from Fig. 4D indicate that P450CPS1 has two binding sites for NADH. Although the results shown in Fig 1C show its monomeric form, the multimeric forms of P450CPS1 need to be confirmed in further studies.

The plots drawn using the M-W-C model [32,67] at low and high substrate concentrations show that the curve was more sigmoidal at low substrate range and shifted towards hyperbola shape at high substrate range (Fig. 5A) shifting the L (allosteric constant) value from 0 to 1000. Bigger shift in L value shows higher allosteric effect. The n_H shift from 0.53 to 10.83 also support the homotropic allosteric effect of increased substrate concentrations.

The change in NADH concentrations from 50 to 100 µmol showed a shift in L value from 0 to 100, indicating the heterotropic allosteric effect, but there was no shift observed when the NADH concentration was above 100 µmol (Fig. 5A). The results clearly indicate, that NADH is not only necessary for the formation of enzyme substrate complex, but also acts as an allosteric activator, binding at a different site.

The results also clearly show that P450CPS1 belongs to the allosteric enzymes. It possesses the properties of a K-system with homotropic effects of the substrate (PCP), and heterotropic effects of the activator (NADH). The physiological significance of these results with respect to metabolic control functions remains to be elucidated.

To the best of our knowledge, this is the first study to provide experimental evidence of the involvement of a putative cytochrome P450 in PCP biodegradation, although the function has been predicted (annotated) in the genome of many organisms. Therefore, further insight into the reaction kinetics of P450CPS1 with PCP may provide a better understanding of the application of this enzyme in different areas.

Two reaction mechanisms (mechanism I and mechanism II) are proposed to further discuss the nature of steady state kinetics of P450CPS1 and PCP interaction. Analysis of both mechanisms by the steady state method shows that, in general, simple behaviour is not to be found. However, in both cases, the assumption that small k_4 value, but >0, leads to the observed hyperbolic dependence of k_{obs} versus [S] (Fig. 5B).

The results shown in Fig. 5B, indicate that P450CPS1 reaction with PCP follow two-step reversible equilibrium mechanism (Mechanism II) (Fig. 5B). In the case of a simple reversible equilibrium, k_{obs} is a linear function of [S] with slope k_1 and intercept at zero [S]= k_2 [68] (theoretical slop 1 in Fig. 5A).

In contrast, such plots are nonlinear for the two-step Mechanism II as can be seen from steady state equation [34]. At high [S], the limiting rate is the sum of the rate constants of the second equilibrium, k_3+k_4, and at zero [S], the limiting rate is k_4. This plot is very useful to differentiate between one- and two-step equilibrium processes, and when extrapolated, provides numerical values for k_2 or k_4, respectively.

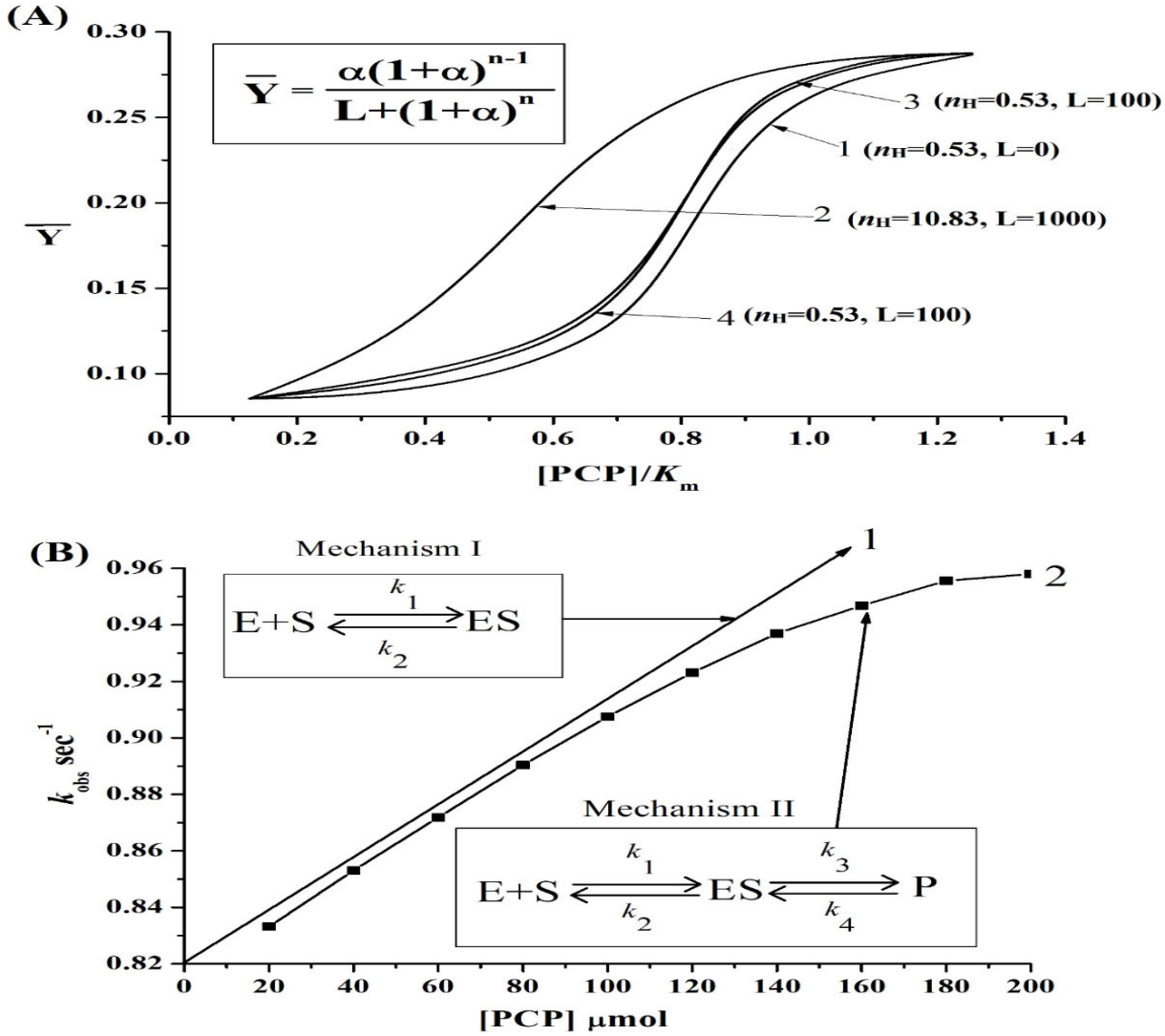


Fig. 5: Allosteric properties of P450CPS1. (A), M-W-C model [32,67] at low (1, from data shown in fig. 4B) and high (2, data shown in Fig. 4C) [PCP], in the presence of 50 µmol NADH, (3, from data shown in fig. 4E) in the presence of 100 µmol NADH, (4, from data shown in fig. 4E) in the presence of 150 µmol NADH; n_H= Hill coefficient calculated from the respective graphs, L=allosteric constant. **(B),** The plot of k_{obs} versus [PCP] (20-200 µmol). The value of k_{obs} was calculated from semilogarithmic plots of P450CPS1 (300 µg) reaction with PCP at different initial concentration (20-200 µmol) versus time obtained from continuous spectrophotometric methods over the time range of initial 10 min. The strait plot represents the theoretical Line (1) when the reaction is in a simple one step reversible equilibrium conditions (mechanism I). The curve (2) represent when the reaction is in a two-step reversible equilibrium condition (mechanism II).

6.3.8. Reaction stoichiometry and UV spectrum of P450CPS1

The UV-visible spectrum of the P450CPS1 shows that the enzyme absorbs maximum light at wavelength 420 nm as well showed some peaks at wavelengths 460 nm, 310 nm, 280 nm and 220 nm (Fig. 6A). The reduction in concentrations of PCP and NADH by P450CPS1 while enzyme assays is shown in Fig. 6B. The decrease in absorbance at wavelength 320 nm represents a reduction in PCP concentration while the reduction in absorption at wavelength 340 nm represents oxidation of NADH during PCP transformation (Fig. 6C). The stoichiometry of the reaction showed that about 30 µmol of NADH was oxidized for every 9.0 µmol of PCP transformed, i.e. for every 1.0 µmol of PCP transformed, about 3.0 µmol of NADH was oxidized.

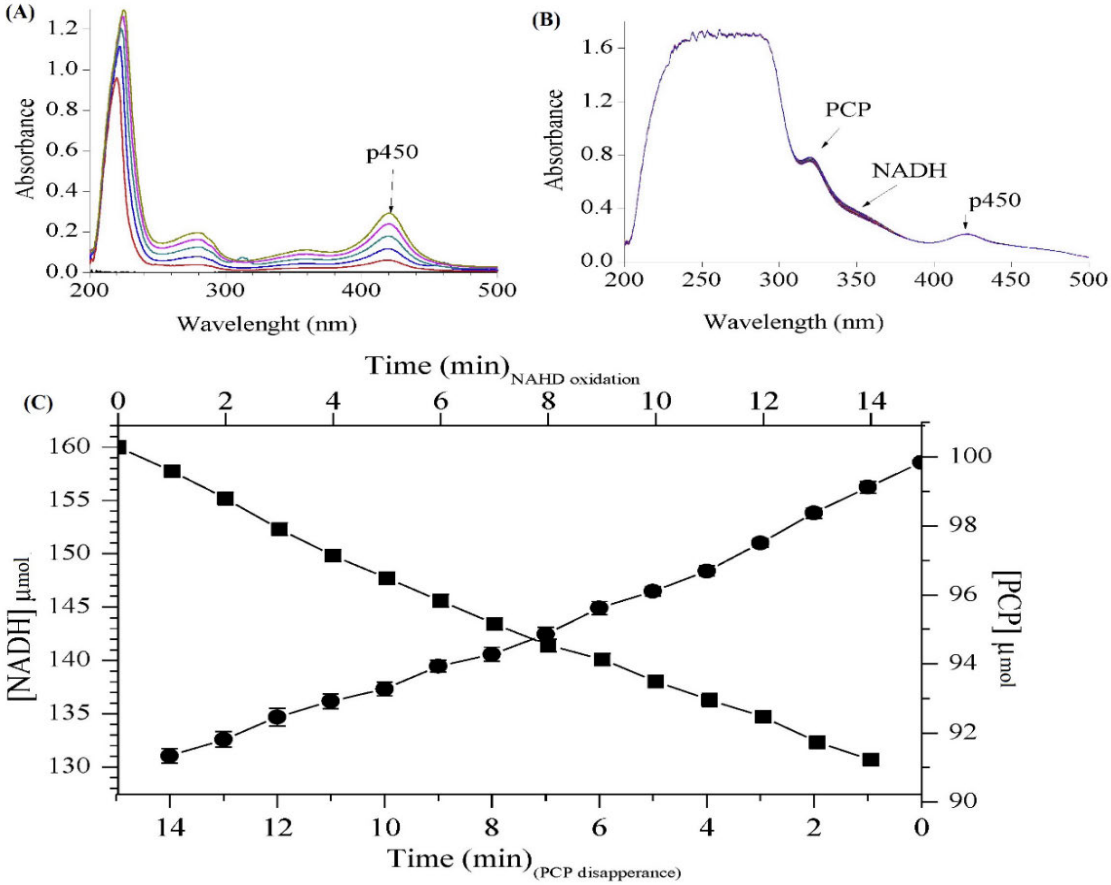


Fig. 6: UV-visible spectrum of P450CPS1 and reaction of PCP with P450CPS1 over time. **(A),** UV spectrum of 50 (–), 100 (–), 150 (–), 200 (–), and 250 (–) µM P450CPS1 in 50 mmol Na-PO_4 buffer, pH 7.0; **(B),** continous spectrophotometric reaction of 100 µmol PCP with 200 µg P450CPS1 in the presence of 160 µmol NADH; **(C),** graph showing PCP transformation (●) and NADH oxidation (■) in a continous spectrophotometric reaction.

6.3.9. Confirmation of the identity of purified protein as P450CPS1

The LC-MS analysis of tryptic digested pure protein band revealed that the peptides spectrum matched with that of the multi species CYPs (UniProt:Q81CYO) from *B. cereus* strains ATCC 14579 [69], with 72.13% coverage, and a 100% peptide matching confidence. Although the protein Q81H02 was annotated, it was only reported at prediction level with no existing experimental evidence. However, the present study report P450CPS1 at experimental level with its characterization and functional properties. The results from LC-MS analysis are shown in Fig. S2a-g.

Furthermore, NCBI protein BLAST (protein to protein) search indicated that P450CPS1 shared 99.69% protein sequence homology with multispecies cytochrome P450 (putative) from *Bacillus* species. (Fig. S3a). This confirmed that P450CPS1 expressed in this study, by the cloned *P450CPS1* gene fragment from *Bt*AOA is the P450CPS1 of interest. Putative conserved domains detected in P450CPS1 structure indicated that P450CPS1 belongs to a P450 heme-thiolate superfamily (accession number: cl12078), with a specific hits CypX and non-specific hits P450_cycloAA_1, P450 and PLN02302, respectively (Fig. S3b) [70–72].

The P450 superfamily are heme-thiolate proteins involved in the oxidative degradation of various recalcitrant organic compounds and are particularly known for their involvement in the transformation of recalcitrant compounds and mutagens [73]. The P450 superfamily are divided into four classes based on the method by which electrons from NAD(P)H are delivered to the catalytic active site [73,74]. However, sequences conservation among this protein family is relatively low within the family; only three residues are absolutely conserved, but their structural and topography and are generally conserved [70]. The conserved core is composed of a coil (meander), a four-helix bundle, helices J and K, and two sets of β-sheets [73]. This makes the heme-binding loop (with a conserved cysteine which serves as the fifth ligand for the heme-iron) of the proton-transfer groove and the conserved EXXR motif in helix K [73]. Their general enzymatic function is to catalyse region-specific and stereospecific oxidation of non-activated hydrocarbons at physiological temperatures [70]. The CypX (accession: COG2124) are specific for secondary metabolites biosynthesis, transport and catabolism, defence mechanisms [70].

The P450_cycloAA_1 subfamily is cytochrome P450 enzymes that occur next to tRNA-dependent cyclodipeptide synthases. The conserved protein domain family P450 (accession:

pfam00067) are heme-thiolate proteins involved in the oxidative degradation of various compounds, while PLN02302 (accession: cl12078) is ent-kaurenoic acid oxidase that catalyses three successive oxidations of ent-kaurene [73,75]. The P450CPS1 has been added to P450 (pfam00067) and PLN02302 (accession: cl12078) superfamily, respectively.

6.3.10. Sequence homology comparison and structural modelling of P450CPS1

The Dali [76] search signified that P450CPS1 was most similar to polyketide biosynthesis cytochrome P450 PKSS (4YZR) from *B. subtilis* 168 with a *Z*-score of 64.1 and 44% sequence identity (Fig. S4a), followed by an acyl carrier protein (3EJB) and cytochrome P450 CYP267B1 protein (6JK5) from *Sorangium cellulosum* SO CE56 with *Z*-scores of 49.7 and 49.0 and sequences identities of 39% and 38%, respectively.

The P450CPS1 structural model was built (by using polyketide biosynthesis cytochrome P450 PKSS (4YZR) as a template) by submitting the amino acid sequence of the protein at SWISS-MODEL server for an automatic modelling of the protein structure. The built model (Fig. S4) was once again submitted at PDBsum database [39] for secondary structure prediction and determination of the theoretical properties of the protein model. Ramachandran Plot statistics indicated: 87.4% of P450CPS1 residues in favoured regions, 2.0% generously allowed regions, 10.3% additional allowed regions, 0.3% disallowed regions (Fig. S5a-k). The predicted P450CPS1 secondary structure comprises of 393 residues of 4-sheets, 4-β-hairpins, 2-β-bulges, 10-strands, 24 α-helixes, 40-helix-helix interacts, 29-β-turns, and 2-ψ-turns (Fig. 7A-B). The 3D sequences and structural alignments between 4YZR, 3EJB, 6JK5 and P450ACPS1 showed that the sequences and structures are conserved (Fig. 8A-B). The 3D graphical structures and models of P450CPS1 generated at PDBsum database is shown in Fig. 8B. The clefts (10) and pores (7) found in the structure built are shown in Figs. 8D-F and 8G-I, respectively, while their properties are shown in Fig. S6a-b. The large pores may improve the efficacy of the protein catalytic function (Fig. 8D-F). The amino acids in the channels are mostly hydrophilic as indicated by the hydropathy index per each amino acid (Fig. S6b).

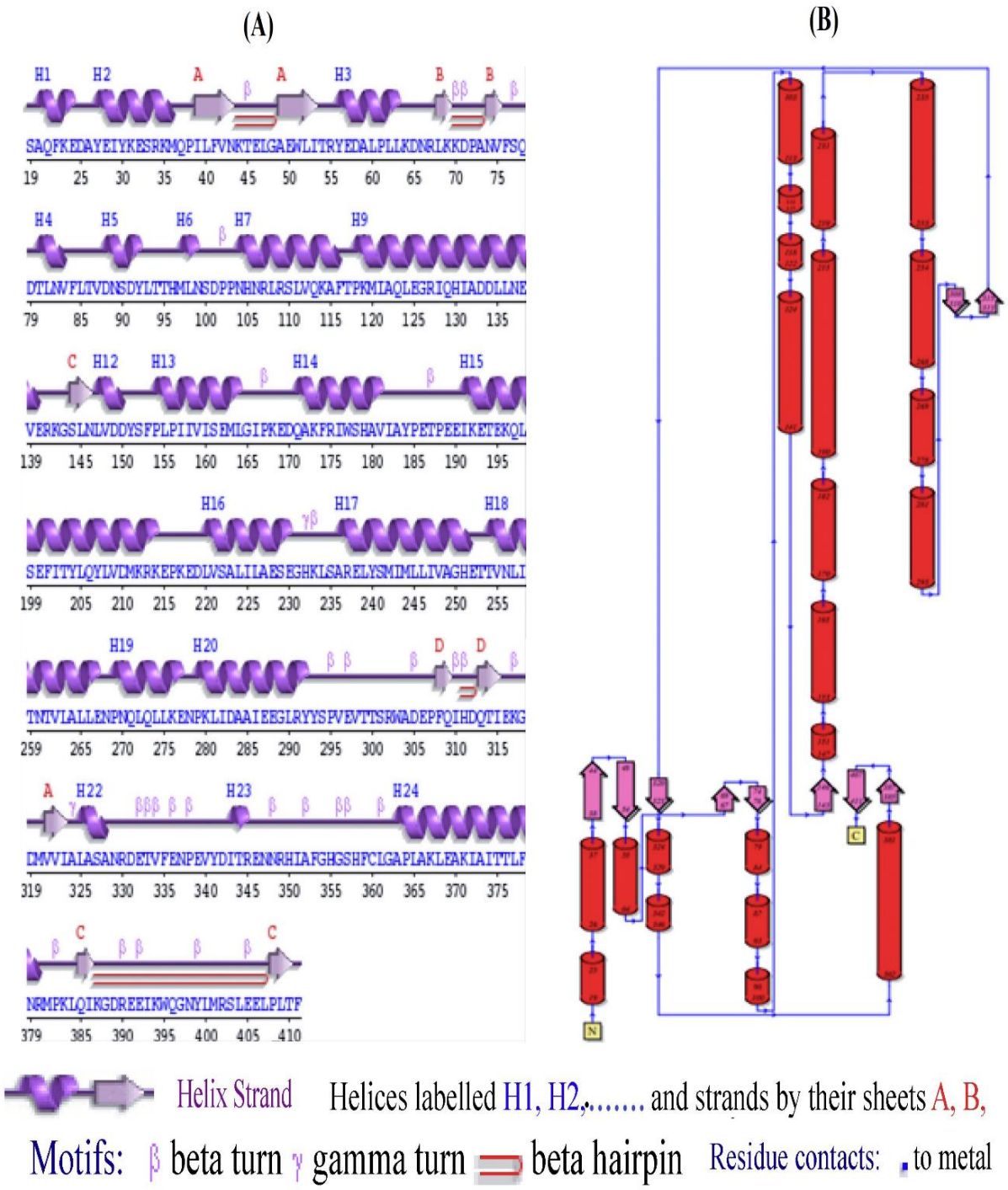


Fig. 7: The predicted 3D and secondary structural framework of P450CPS1. (A), secondary structural elements of P450CPS1 obtained using PDBsum indicating the secondary structure of P450CPS1; **(B),** topology of P450CPS1 obtained using PDBsum showing the motifs repeats.

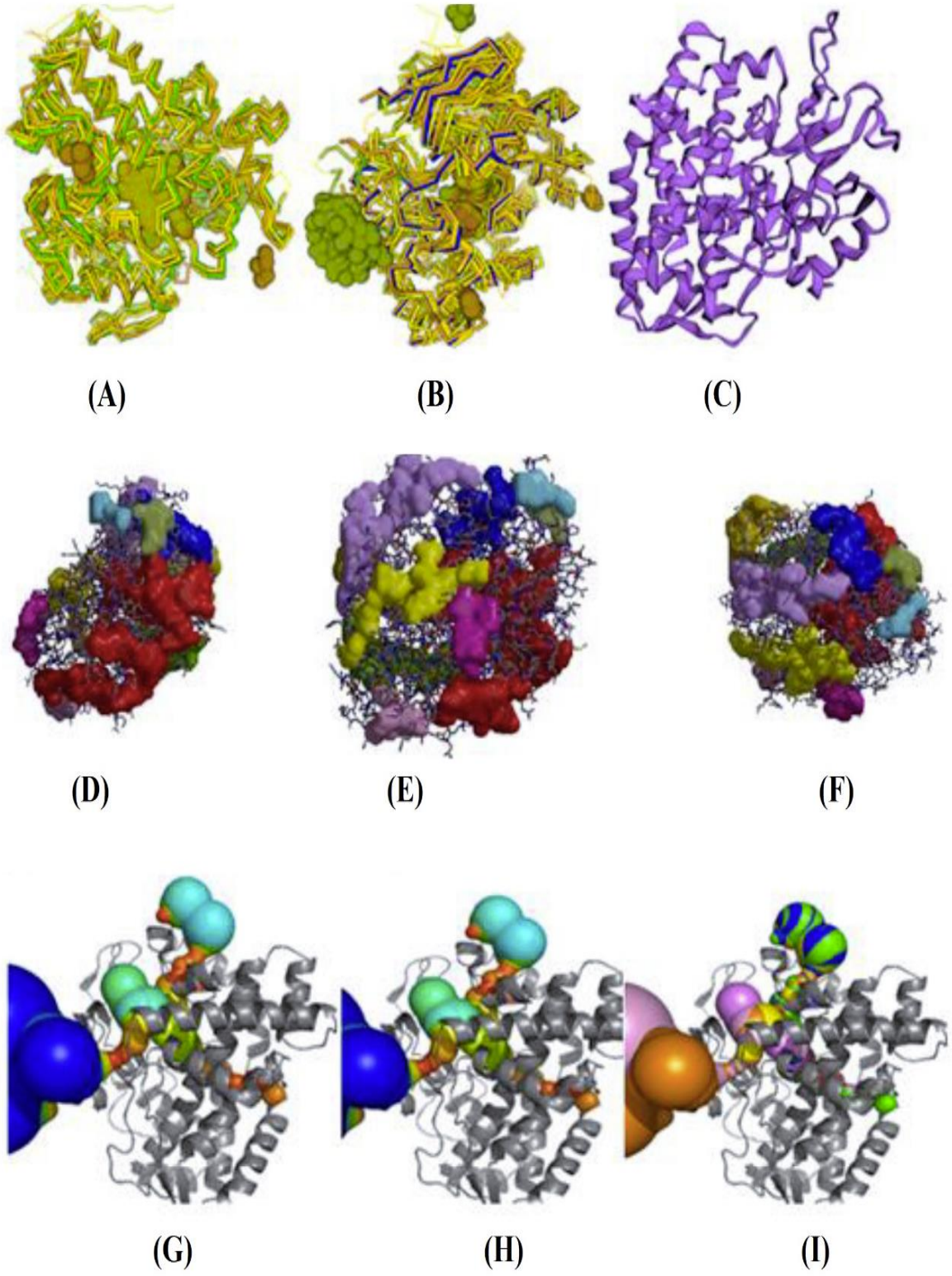


Fig. 8: Graphics of the 3D Superimposition of other CYP's and P450CPS1. (A), the conserved sequence; **(B)**, conserved structures; **(C)**, 3D structure of P450CPS1; **(D, E, F)**, the clefts; (**G, H, I)** pores found in P450CPS1 theoretical structure.

6.3.11. Multiple sequences alignments and evolutionary relationship of P450CPS1 with other CYPs.

Multiple sequences alignments of P450CPS1 from *Bt*AOA (shaded in yellow) and other members of the conserved protein domain family P450 i.e., heme-thiolate proteins in pfam00067 superfamily is shown in Fig. S7a. The multiple sequences alignments were constructed using Cluster Omega (1.2.4) multiple sequence alignment [77]. The alignments showed that only six residues (E288, R291, F353, G356, C360 and G362, i.e., Glu288, Arg291, Phe353, Gly356, Cyc360 and Gly362) that were conserved in P450CPS1 from *Bt*AOA were also conserved in all members of P450 superfamily, in agreement with the previous reports [73], that sequence conservation among cytochrome P450 superfamily is comparatively low, which may be due to the fact that members of this family are from diverse groups. Phylogenetic based evolutionary relatedness of P450CPS1 with other members of P450 superfamily indicated that P450CPS1 is 26% related to aromatases strains P19098 from *Gallus gallus* and P28649 from *Mus musculus*. The enzyme also shared the same branch with cytochrome P450 strains P-450C27/25 (Q02318), SCC (P05108), 11A1 (P00189), 4A10 (P08516), P450-LTB-omega (Q08477), 4F1 (P33274) and benzoate 4-monooxygenase (P17549) (Fig. 9).

Cluster analysis of the alignments between P450CPS1 and the aromatases strains P28649 and P19098 show that P450CPS1 is evolutionarily related with and shared the same leaf which showed that lots of residues (including the six residues that were conserved among all members of P450 superfamily) were conserved in these proteins (Fig. 10). Aromatase (P28649) is a cytochrome P450 enzyme that catalyses three consecutive hydroxylation reactions that converts androgens to aromatic estrogens [78–80]. The reaction utilised three molecular oxygen ($3O_3$) and three reduced NADPH [80]. This was consistent with our observation on the stoichiometry of the reactions between PCP and P450CPS1 in the presence of NADH as described earlier (Fig. 6C). This further strengthens our experimental evidence on the enzymatic degradation of PCP by P450CPS1 in this study (Fig. 6C) and results of the GC-MS analysis of the metabolites of PCP degradation by *Bt*AOA in which P450CPS1 played a prominent role in the hydroxylation of PCP (data not shown).

Furthermore, multiple sequences alignments of P450CPS1 from *Bt*AOA with all other members of the conserved protein domain PLN02302 superfamily i.e., ent-kaurenoic acid oxidase (Fig. S7b), showed that more residues (32 residues) were conserved in all members

of this family compared to that in P450 (pfam00067) superfamily (Fig. S7a). This shows that the more the members in a group, the diverse the conserved residues in that group.

The phylogenetic analysis of P450CPS1 with other members of PLN02302 superfamily (Fig. 11) show that it is only related to kaurenoic acid oxidase from *Selaginella moellendorffii* (XP002988455) and on the same branch with a hypothetical protein SELMODRAFT_1840, partial sequence from *S. moellendorffii* (XP002993510), Os06g0110000 from *Oryza sativa Japonica* Group (NP001056579), and cytochrome P450 88A1 from *Sorghum bicolor* (XP002436354). Cluster analysis of the alignment between P450CPS1 and kaurenoic acid oxidase (Fig. 12) also showed that there were more conserved residues between the two proteins. Similarly, *ent*-kaurenoic acid oxidase, a cytochrome P450 enzyme also catalysed a three-step reaction in the gibberellin biosynthetic pathway from *ent*-kaurenoic acid to GA12 [81,82] similar to that of aromatase which is in agreement with the stoichiometry of P450CPS1 catalysed hydroxylation of PCP in this study.

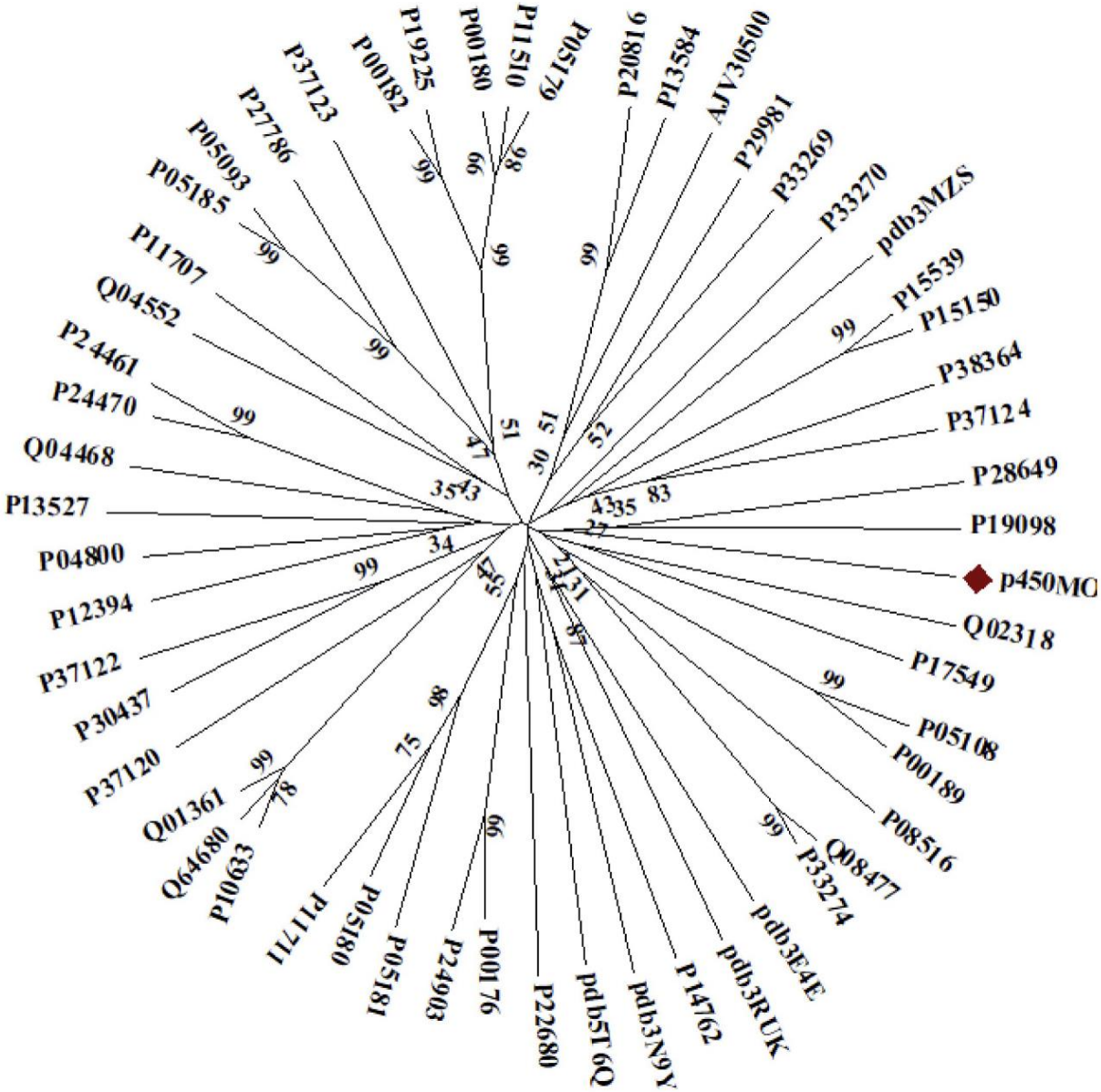


Fig. 9: The evolutionary relationships of P450CPS1 with other members of P450 (pfam00067) superfamily. The optimal tree with the sum of branch length (53,089) is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved 55 amino acid sequences. All positions containing gaps and missing data were eliminated. There was a total of 410 positions in the final dataset. Members of P450 (pfam00067) superfamily include Benzoate 4-monooxygenase (P17549), Erg11p from *Saccharomyces cerevisiae* YJM1307 (AJV30500),

cholesterol side-chain cleavage enzyme (3N9Y), Steroid 17-α-hydroxylase/17, 20 lyase (3RUK) and cytochrome P450 strains 2e1 (3E4E), P-450C27/25 (Q02318), 11B2 (P15539), 11B1 (P15150), SCC (P05108), 11A1 (P00189), 19A1 (P28649), P-450AROM (P19098), P450-LTB-omega (Q08477), 2D14 (Q01361), P450-CMF1A (P10633), CYPXVII (P12394), P450-C17 (P30437), P450c17 (P05185), P450c17 (P05093), 17A1 (P27786), BM-1 (P14762), 7A1 (P22680), 57A2 (P38364), 6A1 (P13527), P450 6a2 (P33270), 6B1 (Q04552), 3A1 (P04800), 3A6 (P11707), 4C1 (P29981), 4A2 (P20816), 4A10 (P08516), 4B1 (P13584), 4F1 (P33274), Cyp11a1 (3MZS), 4B1 (5T6Q), 4d1 (P33269), 77A2 (P37124), 77A1 (P37123), 73 (Q04468), 76A2 (P37122), 75A2 (P37120), 2C12 (P11510), 2C7 (P05179), 2C1 (P00180), 2C3 (P00182), 2C70 (P19225), 2C23 (P24470), 2E1 (P05181), 2H1 (P05180), 2A1 (P11711), 2G1 (P24461), 2B1 (P00176), 2F1 (P24903), 2D4 (Q64680), and P450CPS1.

```
AOA-CPS1    -------------------------------------------MSMK------NKVGLRI    11
P28649      MFLEMLNPMQYNVTIMVPETVTVSAMPLLLIMGLLLLIWNCESSSSIPGPGYCLGIGPLI    60
P19098      MIPETLNPLNYF-TSLVPDLMPVATVPIIILICFLFLIWNHEETSSIPGPGYCMGIGPLI    59
                                                       *          :*  *

AOA-CPS1    ED------GINLASAQFKEDAYEIYKESRKMQPILFVNKTELGAEWLITRYEDALPLLKD    65
P28649      SHGRFLWMGIGSACN-YYNKMYG-----------------EFMRVWISGE--ETLII---    97
P19098      SHGRFLWMGVGNACN-YYNKTYG-----------------EFVRVWISGE--ETFII---    96
            ..      *:. *.  : :. *                  *:   *:  .  ::: :

AOA-CPS1    NRLKKDPANVFSQDTLNVFLTVDNSDY----------------LTTHMLNSDPPNHNRLR    109
P28649      ------------SKSSSMFHVMKHSHYISRFGSKRGLQCIGMHENGIIFNNNPSLWRTIR    145
P19098      ------------SKSSSVFHVMKHWNYVSRFGSKLGLQCIGMYENGIIFNNNPAHWKEIR    144
                        ..: .:* .:.: .*                 .  ::*.:*   . :*

AOA-CPS1    SLVQKAFTPKMIAQ-----LEGRIQHIADDLLNEVERKGSLNLVDDYSFPLPIIVISEML    164
P28649      PFFMKALTGPGLVRMVEVCVESIKQHL-DRLGEVTDTSGYVDVLTLMRHIMLDTSNMLFL    204
P19098      PFFTKALSGPGLVRMIAICVESTIVHL-DKLEEVTTEVGNVNVLNLMRRIMLDTSNKLFL    203
             :. **::   :.:     :*.   *: * * : .   * ::::      :       :*

AOA-CPS1    GIPKE-------DQAKFRIWSHAVIA-------------YPETPEEIKETEKQLSEFITY    204
P28649      GIPLDESAIVKKIQGYFNAWQALLIKPNIFFKISWLYRKYERSVKDLKDEIAVLVEK--K    262
P19098      GVPLDESAIVLKIQNYFDAWQALLLKPDIFFKISWLCKKYEEAAKDLKGAMEILIEQ--K    261
            *:* :        *  *  *.  ::              * .: :::*     * *

AOA-CPS1    LQYLVDMKRKEPKEDLVSALILAESEGHKLSARELYSM----------------IMLLIV    248
P28649      RHKVSTAEKLEDCMDFATDLIFAERRGDLTKENVNQCILEMLIAAPDTMSVTLYFMLLLV    322
P19098      RQKLSTVEKLDEHMDFASQLIFAQNRGDLTAENVNQCVLEMMIAAPDTLSVTLFIMLILI    321
             : :   :: :   *:.: **:*: .*.    .   .:                :**:::|

AOA-CPS1    AGHETTVNLITNTVLALLENPNQLQLLKENPKLIDAAIEEGLRYYSPVEVTTSRWADEPF    308
P28649      AEYPEVEAAILKEIHTVVGDRDIKIEDIQNLKVVENFINESMRYQPVVDLVMRR-ALEDD    381
P19098      ADDPTVEEKMMREIETVMGDREVQSDDMPNLKIVENFIYESMRYQPVVDLIMRK-ALQDD    380
            *    .   : . : ::: : :       * *:::  * *.:**   *::   : * :

AOA-CPS1    QIHDQTIEKGDMVVIALASANRDETVFENPEVYDI-----TRENNRHIAFGHGSHFCLGA    363
P28649      VIDGYPVKKGTNIILNIGRMHRL-EYFPKPNEFTLENFEKNVPYRYFQPFGFGPRGCAGK    440
P19098      VIDGYPVKKGTNIILNIGRMHKL-EFFPKPNEFSLENFEKNVPSRYFQPFGFGPRGCVGK    439
             *..  ::**  ::: :.  ::    * :*: : :     .   . .  **.* : * *

AOA-CPS1    PLAKLEAKIAITTLFNRMPKLQIKGDRE-EIKWQGNYLMRSLEELPLTF-----------    411
P28649      YIAMVMMKVVLVTLLRRFQVKTLQKRCIENIPKKNDLSLHPNEDRHLVEIIFSPRNSDKY    500
P19098      FIAMVMMKAILVTLLRRCRVQTMKGRGLNNIQKNNDLSMHPIERQPLLEMVFTQEAQTRI    499
             :* :  *  :.**:.*     ::     :*  :.:  ::  *   *

AOA-CPS1    --------     411
P28649      LQQ-----     503
P19098      RVTKVDQH     507
```

Fig. 10: Cluster analysis of p450CPS1 with aromatases strain P28649 and P19098 that shared the same leaf with p450CPS1.

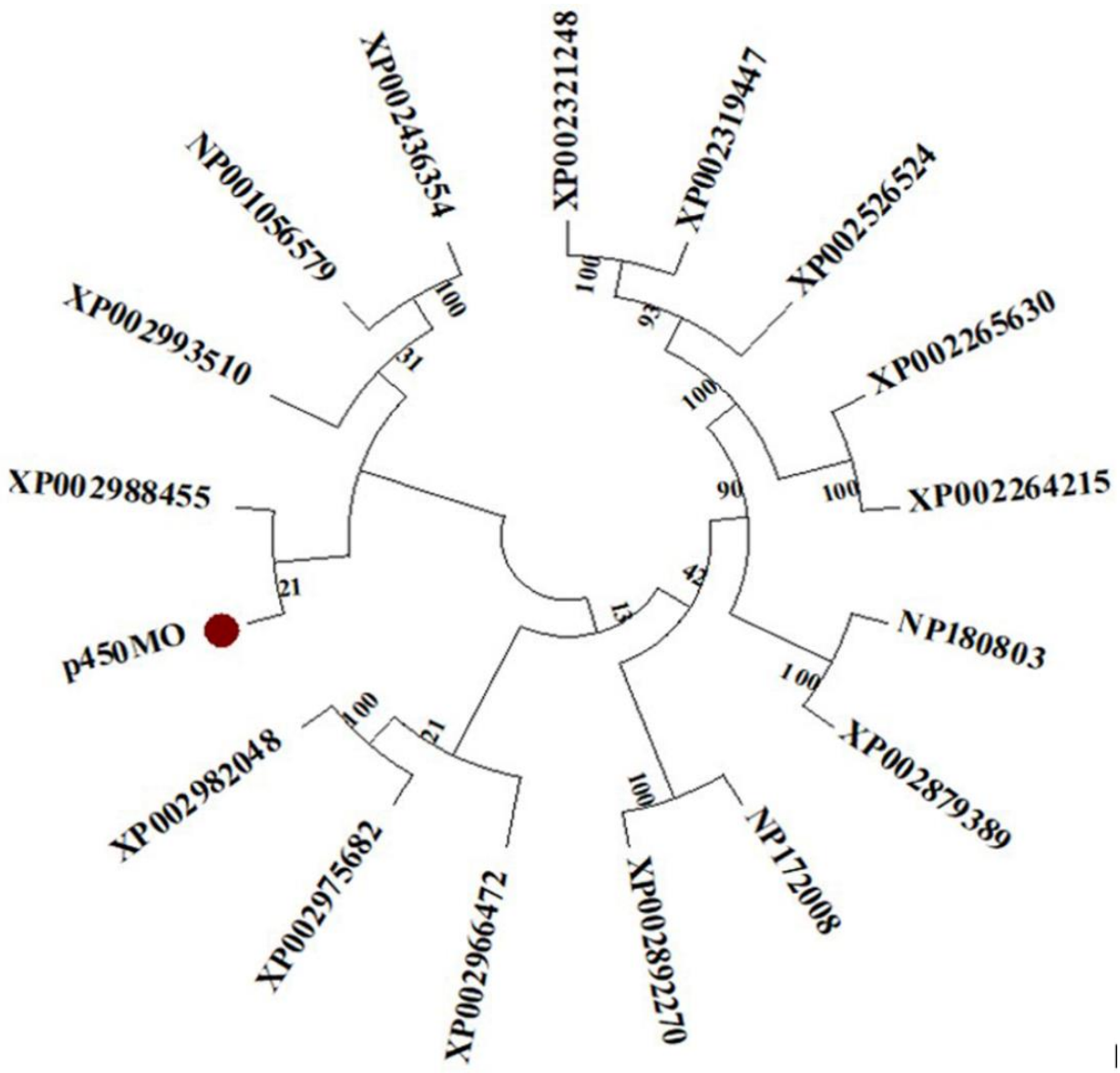


Fig 11: The evolutionary relationships of P450CPS1 with other members of PLN02302 superfamily. The optimal tree with the sum of branch length (14.79) was shown. The evolutionary distances were computed using the Poisson correction method [42] and are in the units of the number of amino acid substitutions per site. The analysis involved 17 amino acid sequences. All positions containing gaps and missing data were eliminated. There was a total of 411 positions in the final dataset. Members of PLN02302 superfamily (cl12078) include Os06g0110000 from *Oryza sativa Japonica* Group (NP001056579), cytochrome P450, family 88, subfamily A, polypeptide 3 from *Arabidopsis thaliana* (NP_172008), ent-kaurenoic acid hydroxylase 2 *A. thaliana* (NP180803), cytochrome P450, probable ent-kaurenoic acid oxidase from *Populus trichocarpa* (XP_002321248), ent-kaurenoic acid oxidase family protein from *P. trichocarpa* (XP_002319447), ent-kaurenoic acid oxidase 2 (predicted) from *Vitis vinifera*

(XP_002265630), ent-kaurenoic acid oxidase 1 (predicted) from *V. vinifera* (XP_002264215), cytochrome P450 88A1 from *Sorghum bicolor* (XP_002436354), ent-kaurenoic acid oxidase 2 from *Ricinus communis* (XP_002526524), ent-kaurenoic acid hydroxylase 1 from *Arabidopsis lyrata* subsp. *Lyrata* (XP_002892270), ent-kaurenoic acid hydroxylase from *A. lyrata* subsp. *Lyrata* (XP_002879389), kaurenoic acid oxidase from *Selaginella moellendorffii* (XP_002966472), ent-kaurenoic acid oxidase 2 isoform X1 from *S. moellendorffii* (XP_002975682), hypothetical protein SELMODRAFT_1840 from *S. moellendorffii* (XP_002993510), hypothetical protein SELMODRAFT_115926 from *S. moellendorffii* (XP_002982048), kaurenoic acid oxidase from *S. moellendorffii* (XP_002988455) and P450CPS1 from BcCPS1.

```
AOA-CPS1       ----------------------------------MSMKNKVGLRIEDGINLASAQFKE--  24
XP_002988455   MILGVAVVAVTALLTFSSFFNKWYYEPVLKPGQPPLPPGSLGWPVFGNMAAFLRAFKSGR  60
                                                  ..:*  : ..:     **.

AOA-CPS1       ------------DAYEIYKESRKMQPILFV----------NKTELGAEWLITRYEDALPL  62
XP_002988455   PDTFMAHYVAKYNRVGFYKAFLFWQPTVLAATPEACKFVLSKDSFETGWPES----AVEL  116
                          :   :**     ** ::.         .* .: : *  :    *: *

AOA-CPS1       LKDNRLKKDPANVFSQDTLNVFLTVDNSDYLTTHMLNSDPPNHNRLRSLVQK--------  114
XP_002988455   MGRN----------------SFAGLT-------------GESHFKLRKLTEPAVNSPKAL  147
               :  *                *  :               .* :**.*.:

AOA-CPS1       -AFTPKMIAQLEGRIQHI-AD---DLLNEVER-------------KGSLNLVDDYSFPLP  156
XP_002988455   EQYVPLIVNNIKACLARWSAQDKIVLLTEMRRFTFLTVLHILYGKDSSLDVDE--TFSLY  205
                 :.* :: :::. : :  *:    **.*:.*            ..**:: :  :* *

AOA-CPS1       IIVISEMLGIPKEDQAKFRIWSHAVIAYPETPEEIKETEKQLSEFITYLQYLVDMKRKE-  215
XP_002988455   YIVNQGIRALP--------------INFPGTAYN--KALKARRKLIKLIQDVINQRRASG  249
                ** . : .:*             * :* *  :  :: *   ::*. :* ::: :* .

AOA-CPS1       -PKEDLVSALIL--AESEGHKLSARELYSMIMLLIVAGHETTVNLITNTVLALLENPNQL  272
XP_002988455   KPQETNILSLLMDQLDDKGEALEDAQIIDVLNMYMNAGHDSTAHVIMWLMIFLKRNPDVL  309
                *:*  : :*::   :.:*. *.  :: .:: : : ***::*.::*   :: * .**: *

AOA-CPS1       QLLKENPKLIDAAIEEG----------LRYYSPVEVTTSRWADEP----------FQIHD  312
XP_002988455   EKVKTEQDGIAKCISEGEMLNLSDIKRMRYLSSVVDETLRLANISPMVFRRALVDVEFNG  369
               : :* : . *  .*.**         :** * *   * * *:           .:::.

AOA-CPS1       QTIEKGDMVVIALASANRDETVFENPEVYDITR------ENNRHIAFGHGSHFCLGAPLA  366
XP_002988455   FTIPKGWHAEAWLRQVHMDPHVHPDPEKFDPERWEKYGASPFTFMPFGMGNRTCPGNELA  429
                ** **  .   * ..: *  *. :** :*  *     .   .: ** *.: * *  **

AOA-CPS1       KLEAKIAITTLFNRMPKLQIKGDREEIKWQGNYLMRSLEEL----PLTF---------   411
XP_002988455   KLQIFIVVHYFVTGY------------RWTALNPNSKVSYLPHPRPRDFYSVRVSKLL   475
               **:  *.:  :..             :* .    .:. *    *  *
```

Fig. 12: Multiple sequences alignments between P450CPS1 and kaurenoic acid oxidase (XP_002988455).

6.4. Conclusion

The present study provided experimental evidence on the involvement of a putative CYP P450CPS1 in PCP degradation and insights into the reaction stoichiometry and the steady-states kinetics. To the best of our knowledge, this is the first report on the involvement of bacterial CYP in chlorophenol biotransformation and provided a comparative analysis with PCP hydroxylation by fungi CPS's counterpart. This study provides information on biological function of the multispecies family putative protein CYP from several species of *Bacillus* but whose function remain putative to date. Though, the in-gel trypsin digestion experiments and bioinformatics tools confirmed that the reported enzyme is cytochrome P450 monooxygenase, further experimental studies including measurement of heme content (with Fe^{2+}-ions) per 47.1 kDa protein and uv-vis spectra of the protein in the presence and absence of CO, are needed to confirm whether P450CPS1 meets minimum criteria for a cytochrome P450 monooxygenase. The enzyme (P450CPS1) portrayed a promising biocatalyst for the oxidation of the rate limiting step in PCP biodegradation. The catalytic efficiency of the enzyme is comparatively higher than those of other PCP hydroxylating monooxygenases and its heterogonous expression level is relatively high, indicating that the enzyme can be produced in bulk and used in enzymatic remediation and other biotechnological applications.

6.5. Acknowledgments:

Author thank CAF, Stellenbosch, South Africa for the Liquid-Chromatography Mass-Spectrometry proteomics analysis of the protein.

6.6. Author contributions:

O.A. and A.O conceived and designed the project; O.A. and A.K. designed the experiments; O.A. performed the experiments; M.P.M. contributed reagents and materials; O.A., A.K., M.P.M. and A.O. wrote the manuscript; all the authors have read and approved the manuscript.

6.7. Funding: National Research Foundation, South Africa (Grant No: 94036 and 92803).

6.8. Conflict of interest: All the authors declare no conflict of interest.

6.9. Ethical statement:

This article does not contain any studies with human participants or animals performed by any of the authors.

6.10. References:

[1] K. Sakai, F. Matsuzaki, L. Wise, Y. Sakai, S. Jindou, H. Ichinose, N. Takaya, M. Kato, H. Wariishi, M. Shimizu, C.K. Sakai, E. Ning-Yi Zhou, S. Jiao Tong, Biochemical characterization of CYP505D6, a self-sufficient cytochrome P450 from the white-rot fungus *Phanerochaete chrysosporium*, Appl. Environ. Microbiol. 84 (2018) e01091-18.

[2] B.R. Shah, W. Xu, J. Mraz, Cytochrome P450 1B1: Role in health and disease and effect of nutrition on its expression, RSC Adv. 9 (2019) 21050–21062. https://doi.org/10.1039/c9ra03674a.

[3] A. Greule, J.E. Stok, J.J. De Voss, M.J. Cryle, Unrivalled diversity: the many roles and reactions of bacterial cytochromes P450 in secondary metabolism, Nat. Prod. Rep. 35 (2018) 757–791. https://doi.org/10.1039/c7np00063d.

[4] P. Durairaj, J.-S. Hur, H. Yun, Versatile biocatalysis of fungal cytochrome P450 monooxygenases, Microb. Cell Fact. 15 (2016) 125. https://doi.org/10.1186/s12934-016-0523-6.

[5] K. Syed, A. Porollo, Y.W. Lam, P.E. Grimmett, J.S. Yadav, CYP63A2, a catalytically versatile fungal P450 Monooxygenase capable of oxidizing higher molecular weight polycyclic aromatic hydrocarbons, alkylphenols, and alkanes, Appl. Environ. Microbiol. 79 (2013) 2692–2702. https://doi.org/10.1128/AEM.03767-12.

[6] P.F.X. Corvini, A. Schäffer, D. Schlosser, Microbial degradation of nonylphenol and other alkylphenols - Our evolving view, Appl. Microbiol. Biotechnol. 72 (2006) 223–243. https://doi.org/10.1007/s00253-006-0476-5.

[7] R.-H. Peng, A.-S. Xiong, Y. Xue, X.-Y. Fu, F. Gao, W. Zhao, Y.-S. Tian, & Quan, H. Yao, Q.-H. Yao, Microbial biodegradation of polyaromatic hydrocarbons, FEMS Microbiol Rev. 32 (2008) 927–955. https://doi.org/10.1111/j.1574-6976.2008.00127.x.

[8] V.B. Urlacher, M. Girhard, Cytochrome P450 monooxygenases in biotechnology and synthetic biology, Trends Biotechnol. 37 (2019) 882–897. https://doi.org/10.1016/j.tibtech.2019.01.001.

[9] K. Hlouchova, J. Rudolph, J.M.H. Pietari, L.S. Behlen, S.D. Copley, Pentachlorophenol hydroxylase, a poorly functioning enzyme required for degradation of pentachlorophenol by *Sphingobium chlorophenolicum*, Biochemistry. 51 (2012)

3848–3860. https://doi.org/10.1021/bi300261p.

[10] D. Ning, H. Wang, Involvement of cytochrome P450 in pentachlorophenol transformation in a white rot fungus *Phanerochaete chrysosporium*, PLoS One. 7 (2012) e45887. https://doi.org/10.1371/journal.pone.0045887.

[11] M. Dietrich, S. Eiben, C. Asta, T.A. Do, J. Pleiss, V.B. Urlacher, Cloning, expression and characterisation of CYP102A7, a self-sufficient P450 monooxygenase from *Bacillus licheniformis*, Appl. Microbiol. Biotechnol. 79 (2008) 931–940. https://doi.org/10.1007/s00253-008-1500-8.

[12] S.D. Munday, N.K. Maddigan, R.J. Young, S.G. Bell, Characterisation of two self-sufficient CYP102 family monooxygenases from *Ktedonobacter racemifer* DSM44963 which have new fatty acid alcohol product profiles, Biochim. Biophys. Acta - Gen. Subj. 1860 (2016) 1149–1162. https://doi.org/10.1016/j.bbagen.2016.01.023.

[13] J.H.Z. Lee, S.H. Wong, J.E. Stok, S.A. Bagster, J. Beckett, J.K. Clegg, A.J. Brock, J.J. De Voss, S.G. Bell, Selective hydroxylation of 1,8- and 1,4-cineole using bacterial P450 variants, Arch. Biochem. Biophys. 663 (2019) 54–63. https://doi.org/10.1016/j.abb.2018.12.025.

[14] B. Unterweger, N. Drinkwater, P. Johanesen, D. Lyras, G.J. Dumsday, S. McGowan, X-ray crystal structure of cytochrome P450 monooxygenase CYP101J2 from *Sphingobium yanoikuyae* strain B2, Proteins Struct. Funct. Bioinforma. 85 (2017) 945–950. https://doi.org/10.1002/prot.25227.

[15] D.J. Frank, Y. Zhao, S.H. Wong, D. Basudhar, J.J. De Voss, P.R.O. De Montellano, Cholesterol analogs with degradation-resistant alkyl side chains are effective *Mycobacterium tuberculosis* growth inhibitors, J. Biol. Chem. 291 (2016) 7325–7333. https://doi.org/10.1074/jbc.M115.708172.

[16] X. Ke, G.J. Ding, B.X. Ma, Z.Q. Liu, J.F. Zhang, Y.G. Zheng, Characterization of a novel CYP51 from *Rhodococcus triatomae* and its NADH-ferredoxin reductase-coupled application in lanosterol 14α-demethylation., Process Biochem. 62 (2017) 59–68. https://doi.org/10.1016/j.procbio.2017.07.030.

[17] A.R. Carroll, B.R. Copp, R.A. Davis, R.A. Keyzers, M.R. Prinsep, Marine natural products, Nat. Prod. Rep. 36 (2019) 122–173. https://doi.org/10.1039/c8np00092a.

[18] J.D. Rudolf, C.Y. Chang, M. Ma, B. Shen, Cytochromes P450 for natural product biosynthesis in Streptomyces: sequence, structure, and function, Nat. Prod. Rep. 34 (2017) 1141–1172. https://pubs.rsc.org/en/content/articlehtml/2017/np/c7np00034k.

[19] H. Zhang, J. Chen, H. Wang, Y. Xie, J. Ju, Y. Yan, H. Zhang, Structural analysis of HmtT and HmtN involved in the tailoring steps of himastatin biosynthesis, FEBS Lett. 587 (2013) 1675–1680. https://doi.org/10.1016/j.febslet.2013.04.013.

[20] N.M. Gaudelli, D.H. Long, C.A. Townsend, β-Lactam formation by a non-ribosomal peptide synthetase during antibiotic biosynthesis., Nature.Com. 520 (2015) 383–401. https://www.nature.com/articles/nature14100.

[21] A. Miyanaga, Michael additions in polyketide biosynthesis, Nat. Prod. Rep. 36 (2019) 531–547. https://doi.org/10.1039/c8np00071a.

[22] R. Bernhardt, V.B. Urlacher, Cytochromes P450 as promising catalysts for biotechnological application: chances and limitations, Appl Microbiol Biotechnol. 98 (2014) 6185–6203. https://doi.org/10.1007/s00253-014-5767-7.

[23] J. Nykiel-Szymańska, P. Stolarek, P. Bernat, Elimination and detoxification of 2,4-D by *Umbelopsis isabellina* with the involvement of cytochrome P450, Environ. Sci. Pollut. Res. 25 (2018) 2738–2743. https://doi.org/10.1007/s11356-017-0571-4.

[24] V.I. Lushchak, T.M. Matviishyn, V. V Husak, J.M. Storey, K.B. Storey, Pesticide toxicity: a mechanistic approach., Excli J. 17 (2018) 1101–1136. https://doi.org/10.17179/excli2018-1710.

[25] O.A. Aregbesola, A. Kumar, M.P. Mokoena, A.O. Olaniran, Role of tetrachloro-1, 4-benzoquinone reductase in phenylalanine hydroxylation system and pentachlorophenol degradation in *Bacillus cereus* AOA-CPS1, Int. J. Biol. Macromol. 161 (2020) 875–890.

[26] O.A. Aregbesola, A. Kumar, M.P. Mokoena, A.O. Olaniran, Cloning, overexpression, purification, characterization and structural modelling of a metabolically active Fe2+ dependent 2, 6-dichloro-p-hydroquinone 1, 2-dioxygenase (CpsA) from *Bacillus cereus* strain AOA-CPS_1, Int. J. Biol. Macromol. 161 (2020) 247–257.

[27] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of Bacteriophage T4, Nature. 227 (1970) 680–685. http://dx.doi.org/10.1038/227680a0.

[28] S.J. Silva, M. I., Burrows, H. D., Formosinho, D. Ferreira, L., Alves, A., Godinho, M. J., Antunes, Photocatalytic degradation of chlorophenols using $Ru(bpy)_3^{2+}/S2O_8^{2-}$, Env. Chem Lett. 5 (2007) 143–149. https://doi.org/10.1007/s10311-007-0096-z.

[29] W.E.L. Brown, A.V. Hill, The oxygen-dissociation curve of blood, and its thermodynamical basis, Proc. R. Soc. London. Ser. B, Contain. Pap. a Biol. Character. 94 (1923) 297–334.

[30] L. Endrenyi, F.H.F. Kwong, C. Fajszi, Evaluation of Hill slopes and Hill coefficients when the saturation binding or velocity is not known, Eur. J. Biochem. 51 (1975) 317–328.

[31] A. Horn, H. Börnig, G. Thiele, Allosteric properties of the Mg^{++}-dependent inorganic Pyrophosphatase in mouse liver cytoplasm, Eur. J. Biochem. 2 (1967) 243–249.

[32] J. Monod, J. Wyman, J.-P. Changeux, On the nature of allosteric transitions: a plausible model, J Mol Biol. 12 (1965) 88–118.

[33] A. Horn, H. Börnig, Analysis of kinetic data of allosteric enzymes by a linear plot, FEBS Lett. 3 (1969) 325–329.

[34] S. Strickland, G. Palmer, V. Massey, Determination of dissociation constants and specific rate constants of enzyme-substrate (or protein-ligand) interactions from rapid reaction kinetic data, J. Biol. Chem. 250 (1975) 4048–4052.

[35] S. Liu, S. Tiantian, Z. Cong, Z. Wen-Mao, Z. Deyu, S. Jing, Tiandi Wei, W. Kang, H. Yan, G. Liming, S. Xu, N.-Y. Zhou, Crystal structure of PnpCD, a two-subunit Hydroquinone 1, 2-dioxygenase, reveals a novel structural class of Fe^{2+}-dependent Dioxygenases, J. Biol. Chem. 290 (2015) 24547–24560. http://www.jbc.org/content/290/40/24547.full.pdf.

[36] B. Setlhare, A. Kumar, M.P. Mokoena, B. Pillay, A.O. Olaniran, Phenol hydroxylase from *Pseudomonas* sp. KZNSA: Purification, characterization and prediction of three-dimensional structure, Int. J. Biol. Macromol. 146 (2020) 1000–1008. https://doi.org/10.1016/j.ijbiomac.2019.09.224.

[37] S. Bienert, A. Waterhouse, T.A.P. de Beer, G. Tauriello, G. Studer, L. Bordoli, T. Schwede, The SWISS-MODEL Repository—new features and functionality, Nucleic Acids Res. 45 (2017) D313--D319. https://doi.org/10.1093/nar/gkw1132.

[38] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F.T. Heer, T.A.P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: homology modelling of protein structures and complexes, Nucleic Acids Res. 46 (2018) W296--W303. https://doi.org/10.1093/nar/gky427.

[39] R.A. Laskowski, J. Jabłońska, L. Pravda, R.S. Vařeková, J.M. Thornton, PDBsum: Structural summaries of PDB entries, Protein Sci. 27 (2018) 129--134.

[40] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees., Mol. Biol. Evol. 4 (1987) 406–425.

[41] Felsenstein, J., Confidence limits on phylogenies: An approach using the bootstrap, Evolution (N.Y). 39 (1985) 783–791.

[42] E. Zuckerkandl, P. Linus, Evolutionary divergence and convergence in proteins. Edited in Evolving genes and proteins by V. Bryson and H.J. Vogel, pp. 97-166. Academic Press, New York., in: 1965.

[43] S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets, Mol. Biol. Evol. 33 (2016) 1870–1874. https://doi.org/10.1093/molbev/msw054.

[44] C. Stacy, K. Sivakumar, S. Shobha, B. Azat, C. Karen, T. Seán, B. Slava, L.S. Conrad, K. Avi, Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI, Int. J. Syst. Evol. Microbiol. 68 (2018) 2386–2392. https://doi.org/10.1099/ijsem.0.002809.

[45] S. Federhen, R. Rossello-Mora, H.-P. Klenk, B.J. Tindall, K.T. Konstantinidis, W.B. Whitman, D. Brown, D. Labeda, D. Ussery, G.M. Garrity, Others, Meeting report: GenBank microbial genomic taxonomy workshop (12-13 May, 2015) BioMed Central, (2016).

[46] D. Liu, X. Zhou, M. Li, S. Zhu, X. Qiu, Characterization of NADPH-cytochrome P450 reductase gene from the cotton bollworm, *Helicoverpa armigera*, Gene. 545 (2014) 262–270. https://doi.org/10.1016/j.gene.2014.04.054.

[47] X. Wen-Sheng, W. Xiang-Jing, R. Tian-Rui, C. Su-Qin, Purification of recombinant wheat cytochrome P450 monooxygenase expressed in yeast and its properties, Protein Expr. Purif. 45 (2006) 54–59. https://doi.org/10.1016/j.pep.2005.07.004.

[48] H. Tomita, Y. Katsuyama, H. Minami, Y. Ohnishi, Identification and characterization of a bacterial cytochrome P450 monooxygenase catalyzing the 3-nitration of tyrosine in rufomycin biosynthesis, Enzyme Res. 2010 (2017) 1–7. https://doi.org/10.1074/jbc.M117.791269.

[49] I. Axarli, A. Prigipaki, N.E. Labrou, Cytochrome P450 102A2 catalyzes efficient oxidation of sodium dodecyl sulphate: A molecular tool for remediation, Enzym. Res. 2010 (2010) 1–8. https://doi.org/10.4061/2010/125429.

[50] M. Kuzu, M. Çiftci, Purification and characterization of NADPH-cytochrome P450 reductase from lake van fish liver microsomes and investigation of some chemical and metals' effects on the enzyme activity, Turkish J. Chem. 39 (2015) 149–158. https://doi.org/10.3906/kim-1404-76.

[51] W. Lu, J. Feng, X. Chen, Y.-J. Bao, Y. Wang, Q. Wu, D. Zhu, Distinct regioselectivity of fungal p450 enzymes for steroidal hydroxylation, Appl. Environ. Microbiol. 85 (2019) e01182-19. https://doi.org/10.1128/AEM.01182-19.

[52] C. Amutha, P. Subramanian, Toxicological and environmental chemistry effect of temperature, salinity, pH and naphthalene on ethoxyresorufin-o-deethylase activity of Oreochromis mossambicus., Toxicol. Environ. Chem. 92 (2010) 127–135. https://doi.org/10.1080/02772240903109092.

[53] S. Hayakawa, H. Matsumura, N. Nakamura, M. Yohda, H. Ohno, Identification of the rate-limiting step of the peroxygenase reactions catalyzed by the thermophilic cytochrome P450 from *Sulfolobus tokodaii*strain 7, FEBS J. 281 (2014) 1409–1416. https://doi.org/10.1111/febs.12712.

[54] Z. Liu, S. Lemmonds, J. Huang, M. Tyagi, L. Hong, N. Jain, Entropic contribution to enhanced thermal stability in the thermostable P450 CYP119, Proc. Natl. Acad. Sci. U. S. A. 115 (2018) E10049--E10058. https://doi.org/10.1073/pnas.1807473115.

[55] S. Mondal, S. Baksi, A. Koris, G. Vatai, Journey of enzymes in entomopathogenic fungi, Pacific Sci. Rev. A Nat. Sci. Eng. 18 (2016) 85–99. https://doi.org/10.1016/j.psra.2016.10.001.

[56] L.H. Kurt, E.S.T. Raine, J.S. Silja, G. Yosephine, M.J.G. Elizabeth, Determinants of thermostability in the cytochrome P450 fold, Biochim. Biophys. Acta - Proteins

Proteomics. 1866 (2018) 97–115. https://doi.org/10.1016/j.bbapap.2017.08.003.

[57] W.A. Johnston, D.J.B. Hunter, C.J. Noble, G.R. Hanson, J.E. Stok, M.A. Hayes, J.J. De Voss, E.M.J. Gillam, Cytochrome P450 is present in both ferrous and ferric forms in the resting state within intact *Escherichia coli* and Hepatocytes, J. Biol. Chem. 286 (2011) 40750–40759. https://doi.org/10.1074/jbc.M111.300871.

[58] S.K. Manna, S. Mazumdar, Reversible inactivation of cytochrome P450 by alkaline earth metal ions: Auxiliary metal ion induced conformation change and formation of inactive P420 species in CYP101, J. Inorg. Biochem. 102 (2008) 1312–1321. https://doi.org/10.1016/j.jinorgbio.2008.01.013.

[59] D.C. Lamb, M.R. Waterman, Unusual properties of the cytochrome P450 superfamily, Philos. Trans. R. Soc. B Biol. Sci. 368 (2013). https://doi.org/10.1098/rstb.2012.0434.

[60] M.J. Paine, N.S. Scrutton, A.W. Munro, A. Gutiérrez, Paine MJI, Scrutton NS, Munro AW, Gutierrez A, Roberts GCK, Wolf CR: Electron transfer partners of cytochrome P450, in: Cytochrome P450 Struct. Mech. Biochem. 3rd Ed. Ed. by Ortiz Montellano PR. Kluwer Acad. Plenum Publ., 2005: pp. 115–148.

[61] F. Vinchi, G. Ingoglia, D. Chiabrando, S. Mercurio, E. Turco, L. Silengo, F. Altruda, E. Tolosano, Heme exporter FLVCR1a regulates heme synthesis and degradation and controls activity of cytochromes P450, Gastroenterology. 146 (2014) 1325–1338. https://doi.org/10.1053/j.gastro.2014.01.053.

[62] M.D. Maines, J.C. Veltman, Phenylhydrazine-mediated induction of haem oxygenase activity in rat liver and kidney and development of hyperbilirubinaemia. Inhibition by zinc-protoporphyrin, Biochem. J. 217 (1984) 409–417. https://doi.org/10.1042/bj2170409.

[63] N.S.K. Münteha, L. Elif, Purification of alkaline serine protease from local *Bacillus subtilis* M33 by two steps: a novel organic solvent and detergent tolerant enzyme, J. Sci. 32 (2019) 116–129. http://dergipark.gov.tr/gujs.

[64] P.T. Sanatan, P.R. Lomate, A.P. Giri, V.K. Hivrale, Characterization of a chemostable serine alkaline protease from *Periplaneta americana*, BMC Biochem. 14 (2013) 32. https://doi.org/10.1186/1471-2091-14-32.

[65] T. Kitazume, N. Takaya, N. Nakayama, H. Shoun, *Fusarium oxysporum* fatty-acid

subterminal hydroxylase (CYP505) is a membrane-bound eukaryotic counterpart of *Bacillus megaterium* cytochrome P450BM3, J. Biol. Chem. 275 (2000) 39734–39740. https://doi.org/10.1074/jbc.M005617200.

[66] M. Sasaki, A. Akahira, K.-I. Oshiman, T. Tsuchido, Y. Matsumura, Purification of cytochrome P450 and ferredoxin, involved in bisphenol a degradation, from *Sphingomonas* sp. Strain AO1., Appl. Environ. Microbiol. 71 (2005) 8024–8030. https://doi.org/10.1128/AEM.71.12.8024-8030.2005.

[67] N.B. Madsen, S. Shechosky, Allosteric properties of Phosphorylase b II. comparision with a kinetic model, J. Biol. Chem. 242 (1967) 3301–3307.

[68] S. Strickland, V. Massey, The purification and properties of the flavoprotein melilotate hydroxylase, J. Biol. Chem. 248 (1973) 2944–2952.

[69] M. Il Kim, C. Lee, J. Park, B.Y. Jeon, M. Hong, Crystal structure of *Bacillus cereus* flagellin and structure-guided fusion-protein designs, Sci. Rep. 8 (2018) 1–10. https://doi.org/10.1038/s41598-018-24254-w.

[70] A. Marchler-Bauer, Y. Bo, L. Han, J. He, C.J. Lanczycki, S. Lu, F. Chitsaz, M.K. Derbyshire, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, F. Lu, G.H. Marchler, J.S. Song, N. Thanki, Z. Wang, R.A. Yamashita, D. Zhang, C. Zheng, L.Y. Geer, S.H. Bryant, CDD/SPARCLE: Functional classification of proteins via subfamily domain architectures, Nucleic Acids Res. 45 (2017) D200--D203. https://doi.org/10.1093/nar/gkw1129.

[71] A. Marchler-Bauer, M.K. Derbyshire, N.R. Gonzales, S. Lu, F. Chitsaz, L.Y. Geer, R.C. Geer, J. He, M. Gwadz, D.I. Hurwitz, C.J. Lanczycki, F. Lu, G.H. Marchler, J.S. Song, N. Thanki, Z. Wang, R.A. Yamashita, D. Zhang, C. Zheng, S.H. Bryant, CDD: NCBI's conserved domain database, Nucleic Acids Res. 43 (2015) D222--D226. https://doi.org/10.1093/nar/gku1221.

[72] A. Marchler-Bauer, S.H. Bryant, CD-Search: Protein domain annotations on the fly, Nucleic Acids Res. 32 (2004) W327–W331. https://doi.org/10.1093/nar/gkh454.

[73] S. Lu, J. Wang, F. Chitsaz, M.K. Derbyshire, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, G.H. Marchler, J.S. Song, N. Thanki, R.A. Yamashita, M. Yang, D. Zhang, C. Zheng, C.J. Lanczycki, A. Marchler-Bauer, CDD/SPARCLE: the conserved domain database in 2020, Nucleic Acids Res. 48 (2020) D265–D268.

https://doi.org/10.1093/nar/gkz991.

[74] X. Zhang, Y. Peng, J. Zhao, Q. Li, X. Yu, C.G. Acevedo-Rocha, A. Li, Bacterial cytochrome P450-catalyzed regio- and stereoselective steroid hydroxylation enabled by directed evolution and rational design, Bioresour. Bioprocess. 7 (2020) 1–18. https://doi.org/10.1186/s40643-019-0290-4.

[75] C. Noguchi, S. Miyazaki, H. Kawaide, O. Gotoh, Y. Yoshida, Y. Aoyama, Characterization of moss ent-kaurene oxidase (CYP701B1) using a highly purified preparation, J. Biochem. 163 (2018) 69–76. https://doi.org/10.1093/jb/mvx063.

[76] L. Holm, Benchmarking fold detection by DaliLite v.5, Bioinformatics. 35 (2019) 5326–5327. https://doi.org/10.1093/bioinformatics/btz536.

[77] F. Madeira, Y.M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A.R.N. Tivey, S.C. Potter, R.D. Finn, R. Lopez, The EMBL-EBI search and sequence analysis tools APIs in 2019, Nucleic Acids Res. 47 (2019) W636--W641. https://doi.org/10.1093/nar/gkz268.

[78] J.C. Hackett, R.W. Brueggemeier, C.M. Hadad, The final catalytic step of cytochrome P450 aromatase: A density functional theory study, J. Am. Chem. Soc. 127 (2005) 5224–5237. https://doi.org/10.1021/ja044716w.

[79] Y.-C. Kao, K.R. Korzekwa, C.A. Laughton, S. Chen, Evaluation of the mechanism of Aromatase cytochrome P450, Eur. J. Biochem. 268 (2001) 243–251. https://doi.org/10.1046/j.1432-1033.2001.01886.x.

[80] F.K. Yoshimoto, F.P. Guengerich, Mechanism of the third oxidative step in the conversion of androgens to estrogens by cytochrome P450 19A1 steroid Aromatase, J. Am. Chem. Soc. 136 (2014) 15016–15025. https://doi.org/10.1021/ja508185d.

[81] C.A. Helliwell, P.M. Chandler, A. Poole, E.S. Dennis, W.J. Peacock, The CYP88A cytochrome P450, ent-kaurenoic acid oxidase, catalyzes three steps of the gibberellin biosynthesis pathway, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 2065–2070. www.pnas.orgcgidoi10.1073pnas.041588998.

[82] Y. Yang, K. Chen, Molecular evolution of the ent-kaurenoic acid oxidase gene in Oryzeae, Genet. Mol. Biol. 35 (2012) 448–454. https://doi.org/10.1590/S1415-47572012005000020.

6.11 SUPPLEMENTARY MATERIALS

Supplementary Materials 1

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Manage Columns Show 100

select all *100 sequences selected* GenBank Graphics Distance tree of results

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Bacillus cereus strain AR156, complete genome	2228	2228	100%	0.0	99.19%	CP015589.1
Bacillus thuringiensis strain XL6, complete genome	2228	2228	100%	0.0	99.19%	CP013000.1
Bacillus cereus strain BHU2 chromosome	2211	2211	100%	0.0	98.95%	CP023726.1
Bacillus cereus ATCC 4342, complete genome	2194	2194	100%	0.0	98.71%	CP009628.1
Bacillus cereus strain BC-AK genomic sequence	2156	2156	100%	0.0	98.14%	CP020937.1
Bacillus cereus G9241, complete genome	2156	2156	100%	0.0	98.14%	CP009590.1
Bacillus cereus strain 03BB87, complete genome	2156	2156	100%	0.0	98.14%	CP009941.1
Bacillus cereus strain FM1, complete genome	2150	2150	100%	0.0	98.06%	CP009369.1
Bacillus sp. ABP14, complete genome	2111	2111	100%	0.0	97.49%	CP017016.1
Bacillus tropicus strain LM1212-W3 chromosome, complete genome	2089	2089	100%	0.0	97.17%	CP041071.1
Bacillus thuringiensis LM1212 chromosome, complete genome	2089	2089	100%	0.0	97.17%	CP024771.1
Bacillus paranthracis strain CFSAN068816 chromosome, complete genome	2039	2039	100%	0.0	96.44%	CP045777.1
Bacillaceae bacterium C05 chromosome, complete genome	2039	2039	100%	0.0	96.44%	CP045537.1

Figure S1: Sequence homology of *Bacillus cereus* AOA-CPS1 against other CYA genes in the gene bank, using the NCBI Blastn suite.

Supplementary Materials 2a

Figure S2(a): Protein identification spectrum overview.

Supplementary Materials 2b

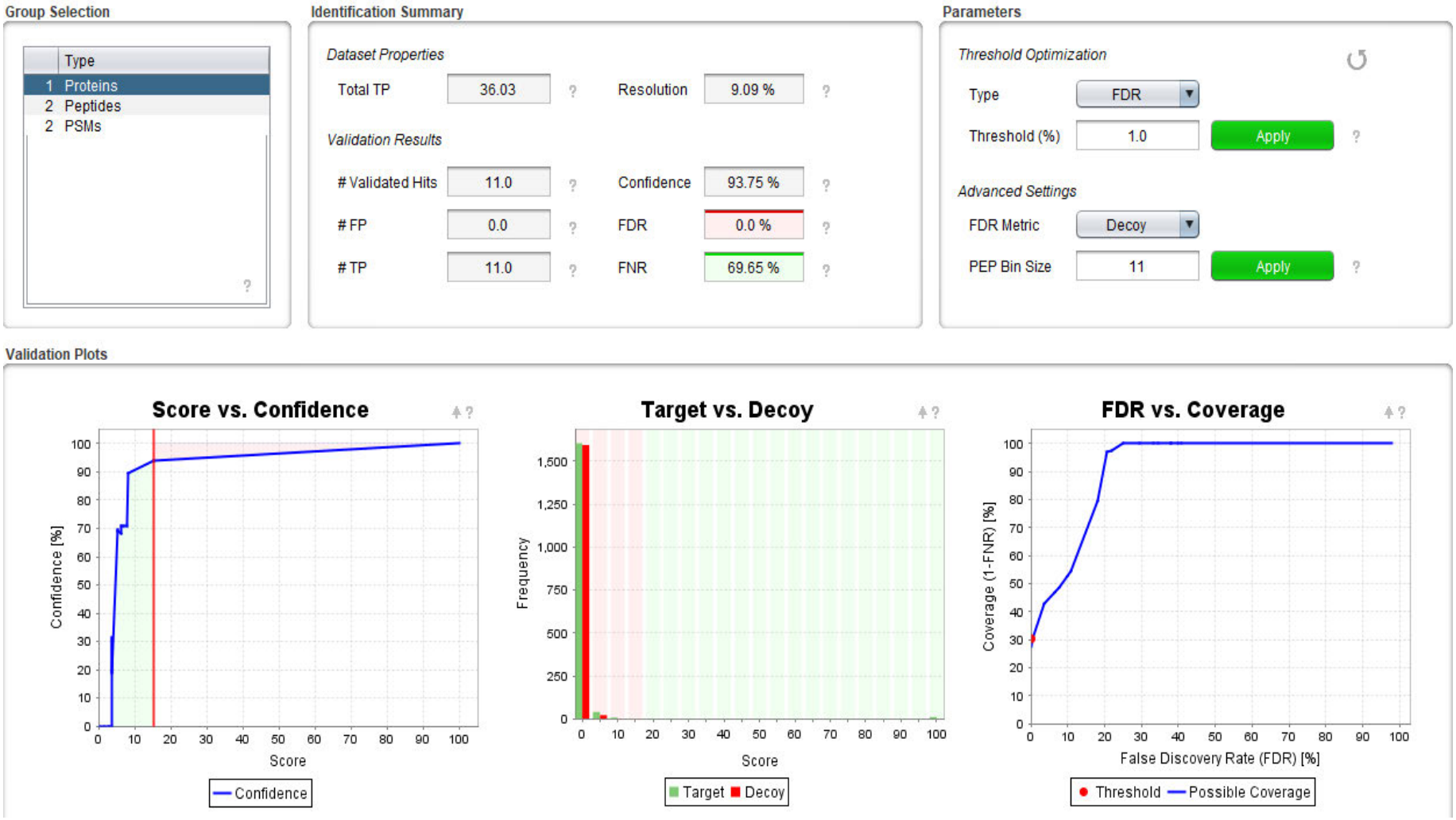


Figure S2(b): Protein validated

Supplementary Materials 2c

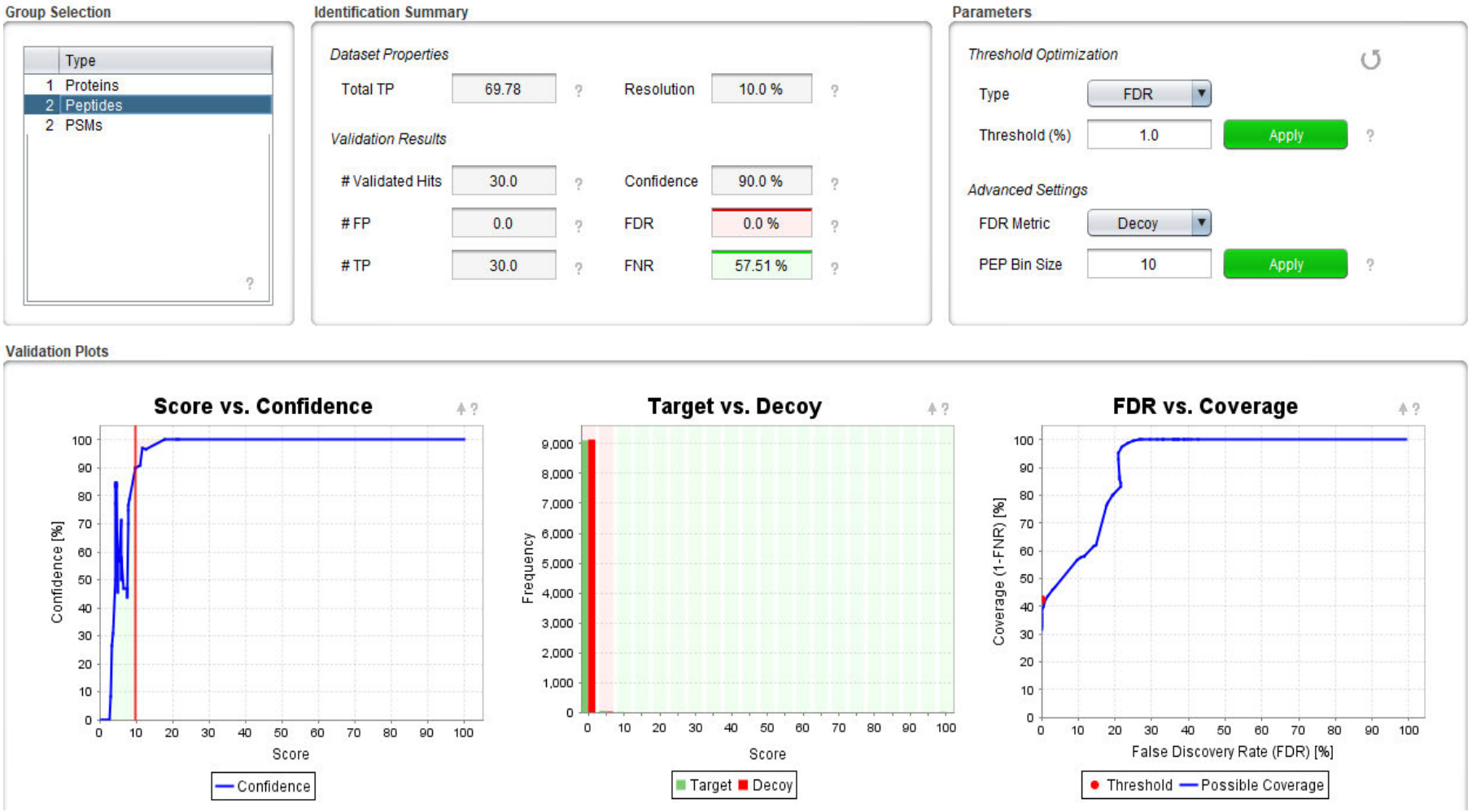


Figure S2(c): Peptide validated

Supplementary Materials 2d

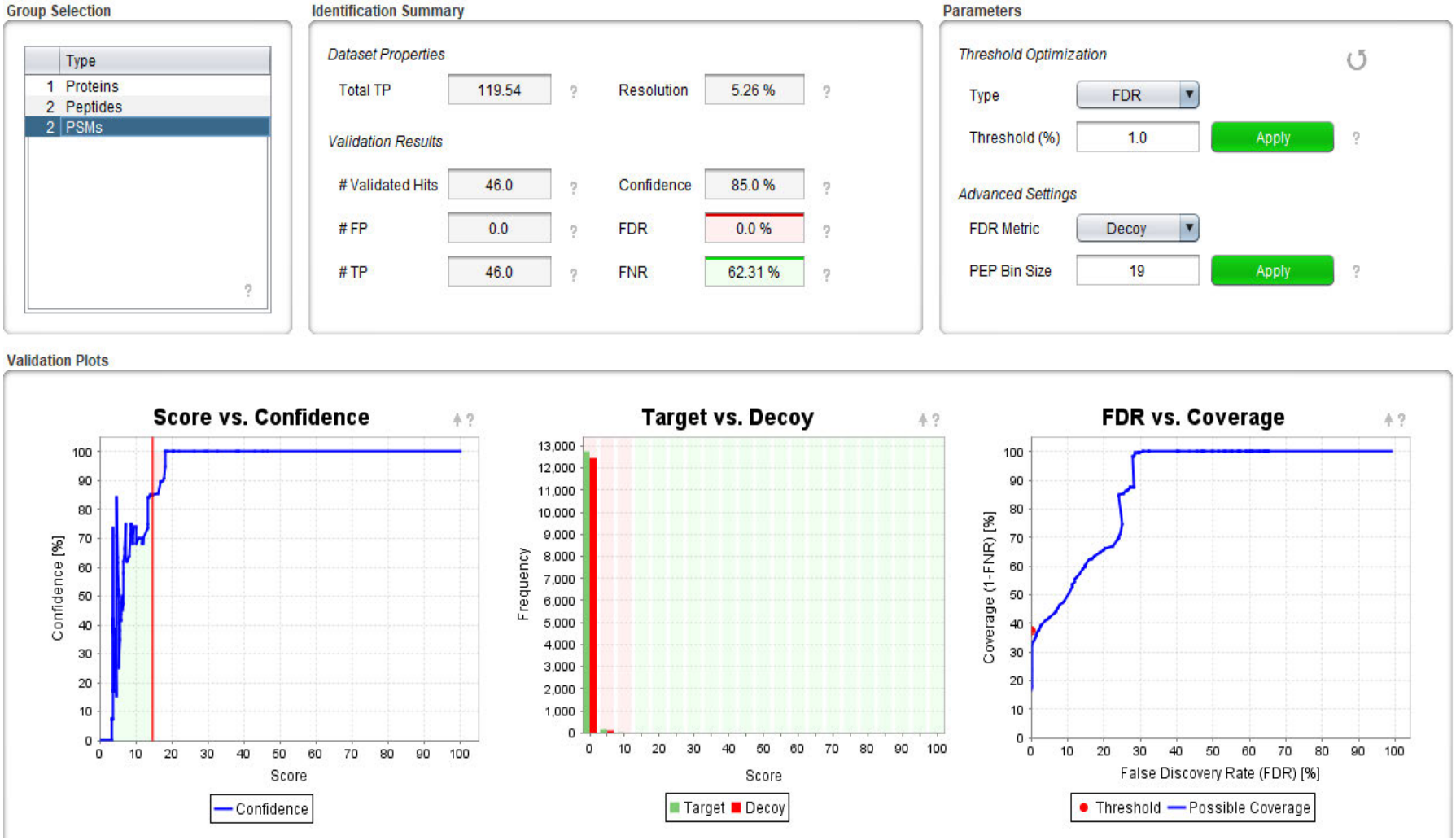


Figure S2(d): PSMs validated

Supplementary Materials 2e

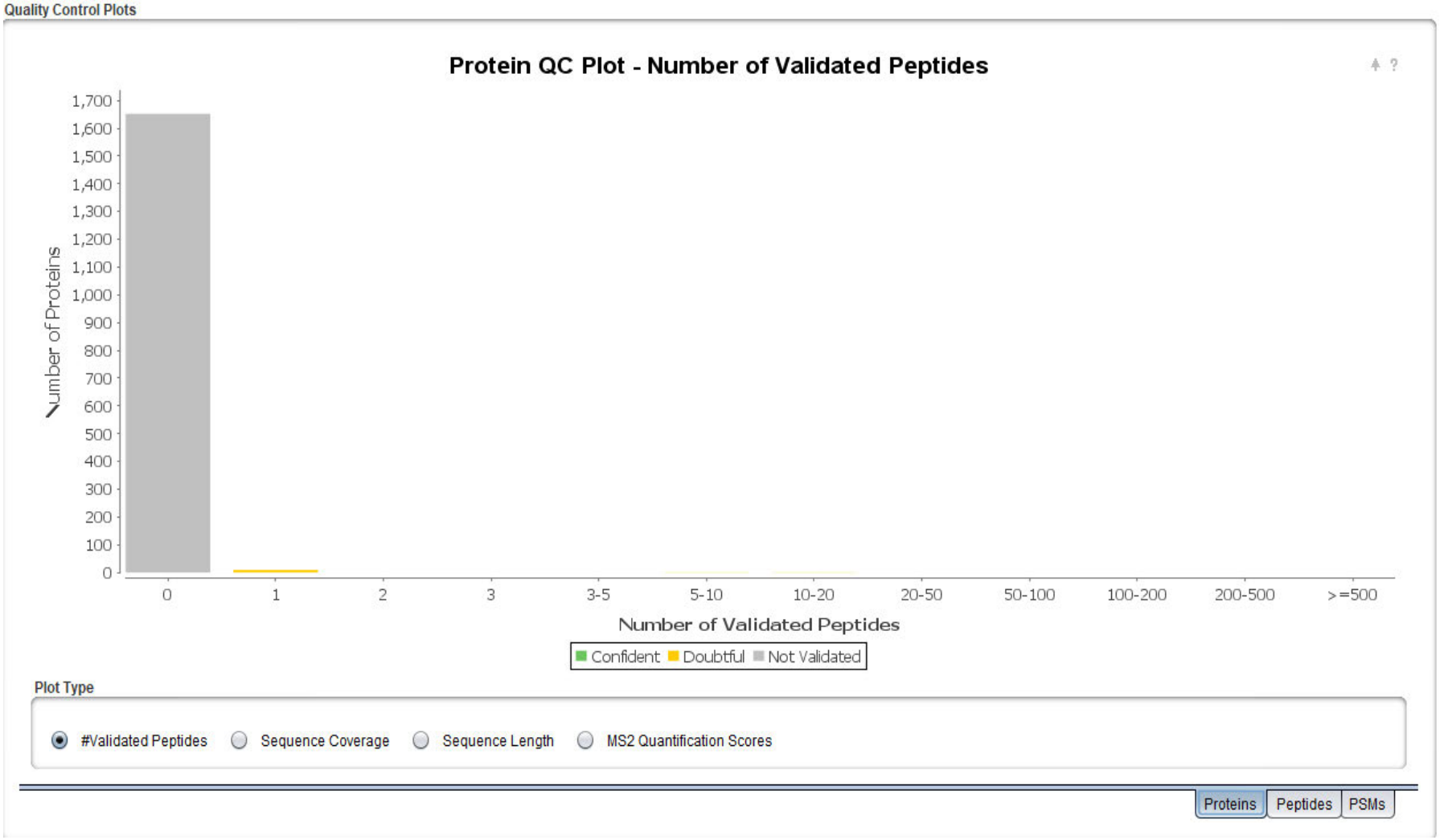


Figure S2(e): Protein quality control plot

Supplementary Materials 2f

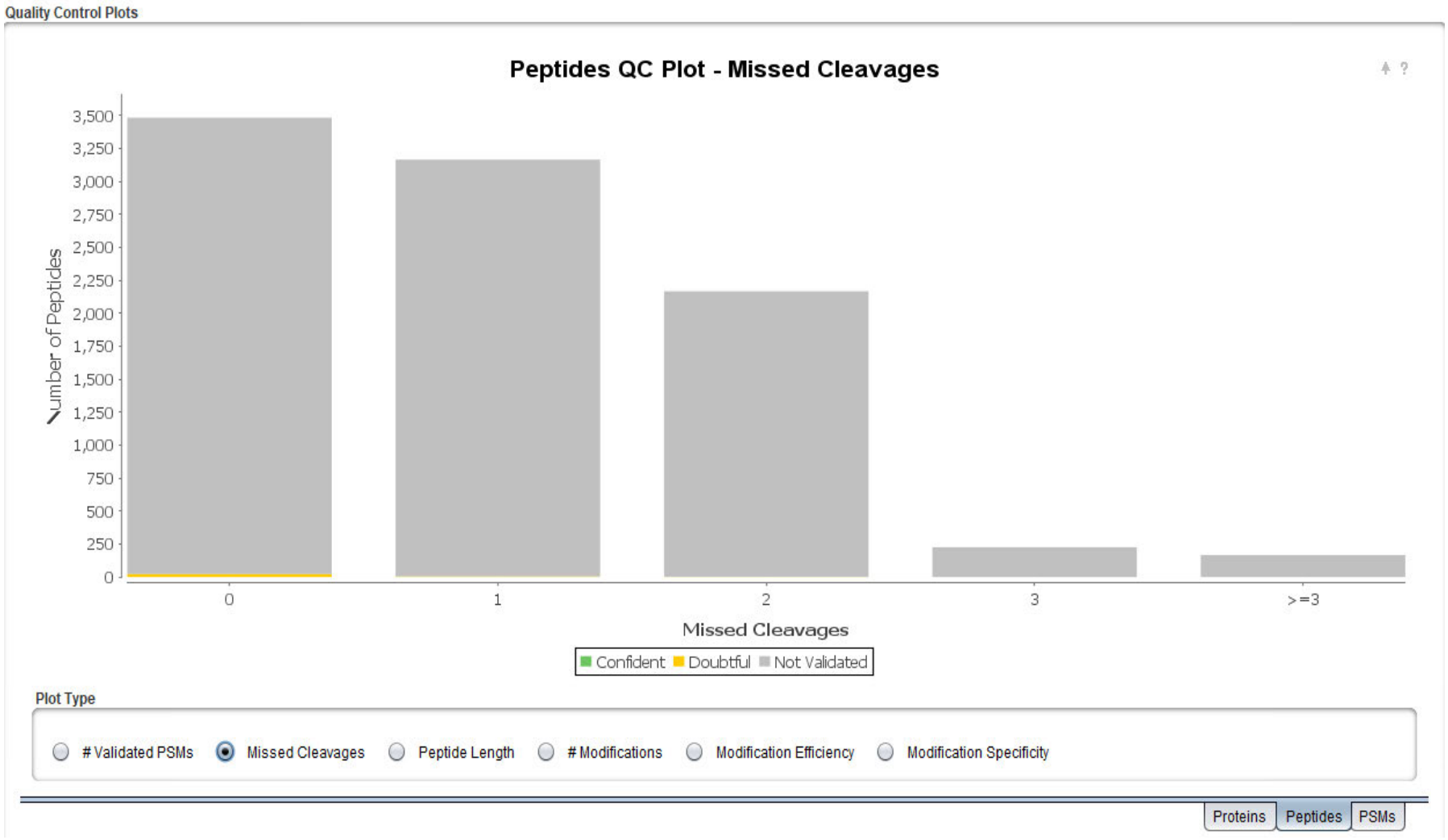


Figure S2(f): Peptide quality control plot

Supplementary Materials 2g

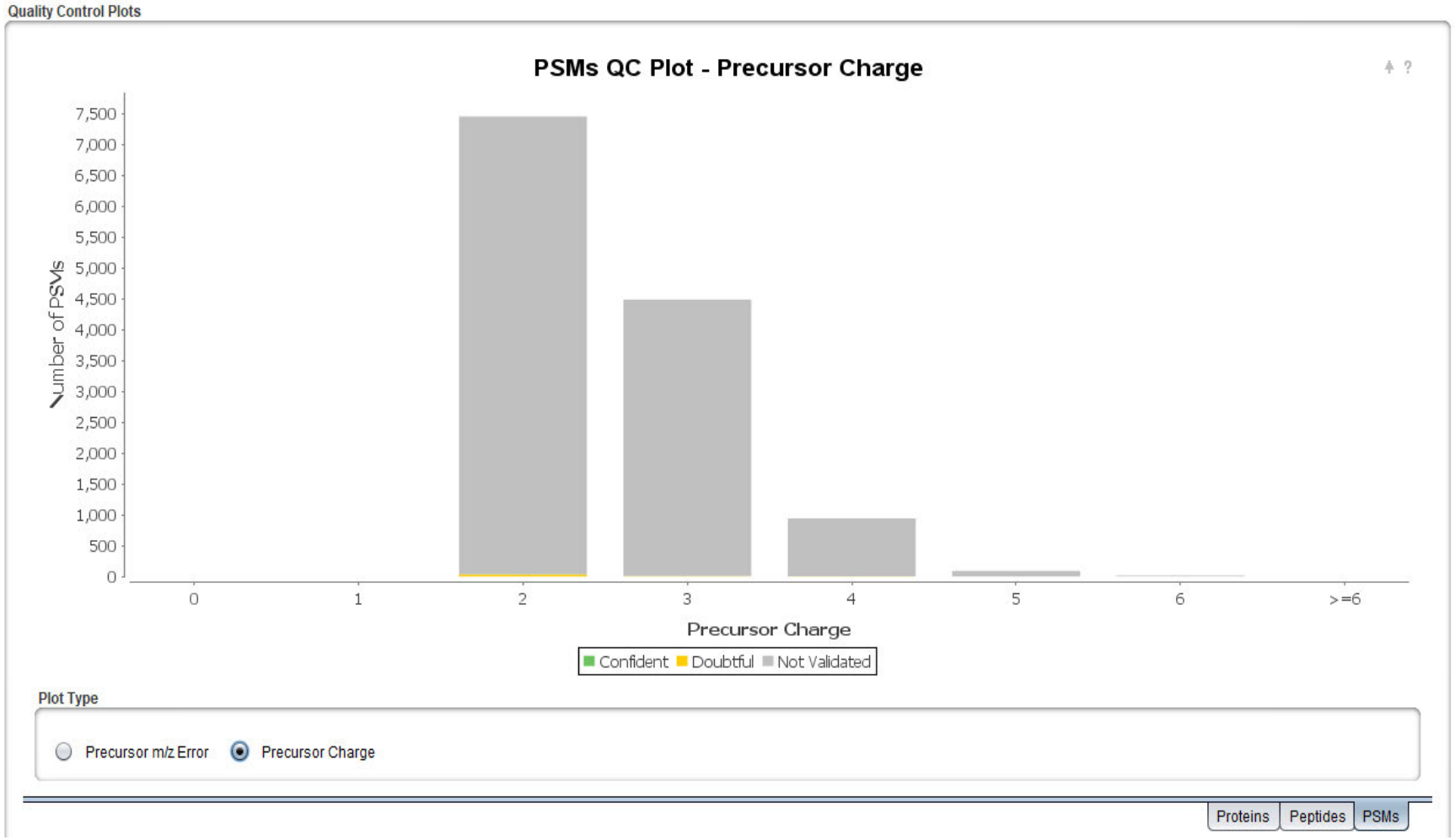


Figure S2(g): PSMs quality control plot

Supplementary Materials 3a

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Manage Columns Show 100

select all *100 sequences selected* GenPept Graphics Distance tree of results Multiple alignment

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
cytochrome P450 [Bacillus cereus]	843	843	100%	0.0	99.76%	WP_074571148.1
MULTISPECIES: cytochrome P450 [Bacillus]	842	842	100%	0.0	99.76%	WP_001980801.1
cytochrome P450 [Bacillus thuringiensis]	842	842	100%	0.0	99.51%	WP_061402872.1
cytochrome P450 [Bacillus cereus]	841	841	100%	0.0	99.27%	WP_074619941.1
cytochrome P450 [Bacillus cereus]	840	840	100%	0.0	99.27%	WP_000062104.1
MULTISPECIES: cytochrome P450 [Bacillus cereus group]	840	840	100%	0.0	99.51%	WP_029438469.1
cytochrome P450 [Bacillus cereus]	840	840	100%	0.0	99.51%	WP_074565691.1
MULTISPECIES: cytochrome P450 [Bacillus]	840	840	100%	0.0	99.27%	WP_000062105.1
MULTISPECIES: cytochrome P450 [Bacillus cereus group]	839	839	100%	0.0	99.27%	WP_098224181.1
cytochrome P450 [Bacillus cereus]	839	839	100%	0.0	99.51%	WP_088298201.1
cytochrome P450 [Bacillus cereus]	838	838	100%	0.0	99.27%	WP_061688238.1
MULTISPECIES: cytochrome P450 [Bacillus cereus group]	838	838	100%	0.0	99.27%	WP_098682387.1
MULTISPECIES: cytochrome P450 [Bacillus cereus group]	838	838	100%	0.0	99.27%	WP_098340411.1
MULTISPECIES: cytochrome P450 [Bacillus cereus group]	837	837	100%	0.0	99.03%	WP_073543955.1
cytochrome P450 [Bacillus cereus]	837	837	100%	0.0	99.03%	WP_098912844.1
MULTISPECIES: cytochrome P450 [Bacillus cereus group]	837	837	100%	0.0	99.03%	WP_071679662.1
MULTISPECIES: cytochrome P450 [Bacillus]	836	836	100%	0.0	99.03%	WP_002027768.1
cytochrome P450 [Bacillus cereus]	835	835	100%	0.0	99.03%	WP_048554593.1

Figure S3(a): NCBI protein BLAST (protein to protein) search of p450MO against other p450 in the protein data bank.

Supplementary Materials 3b

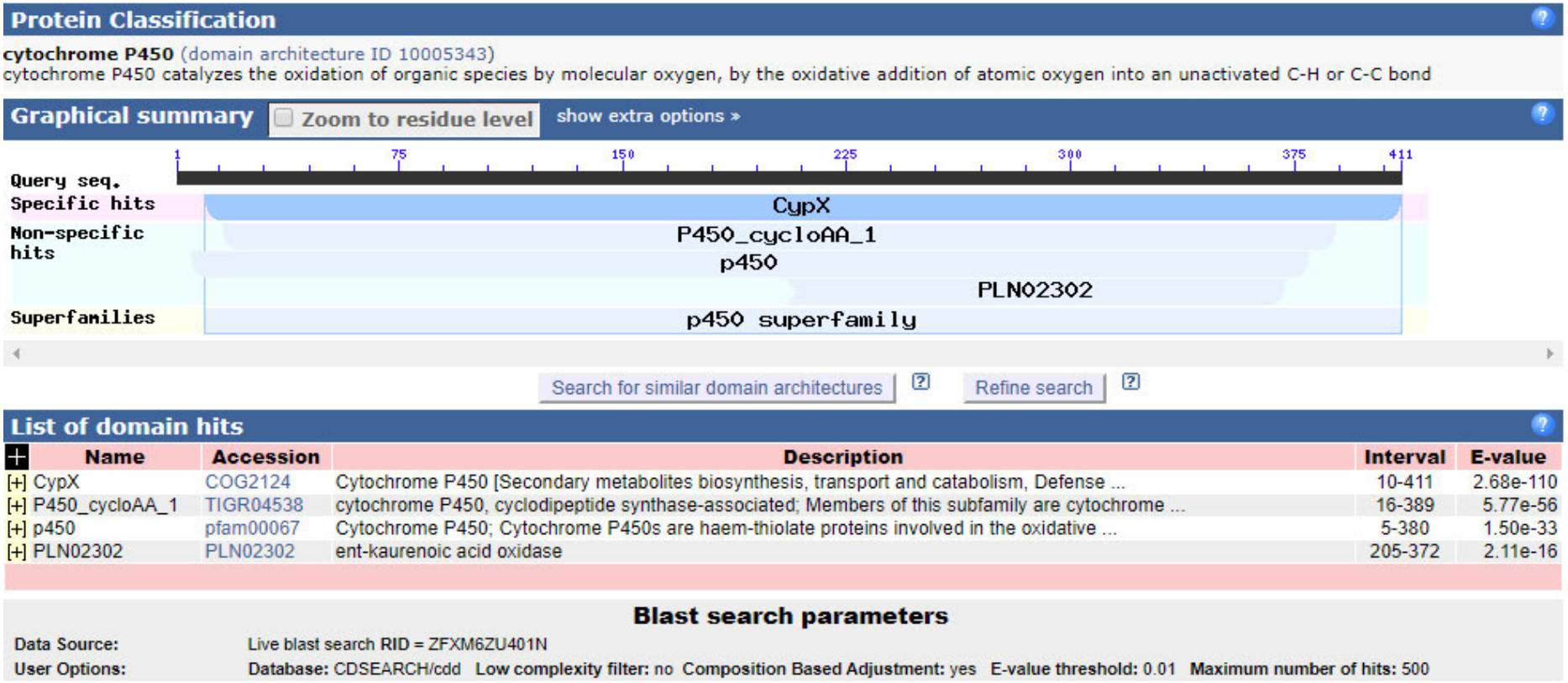


Figure S3(b): Conserved domain in *Bacillus cereus* strain AOA-CPS1 cytochrome p450MO obtain, using the CDD/SPARCLE functional classification of proteins via subfamily domain architectures, in: CDD NCBI's conserved domain database, conserved domain database for the functional annotation of protein and protein domain annotation on the fly.

Supplementary Material 4

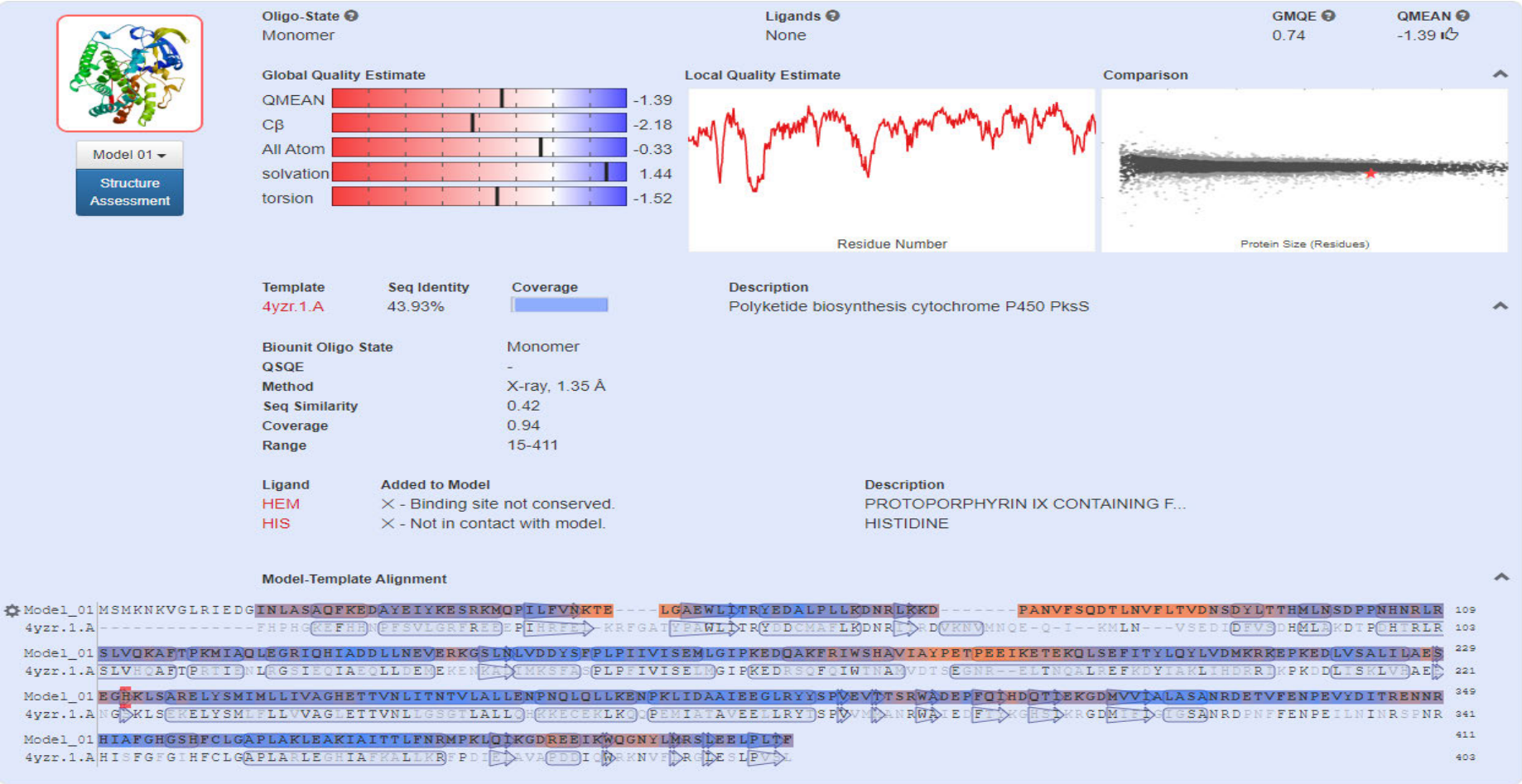


Figure S4: P450AOACPS1 model build using 4yzr as a template.

Supplementary Materials 5a

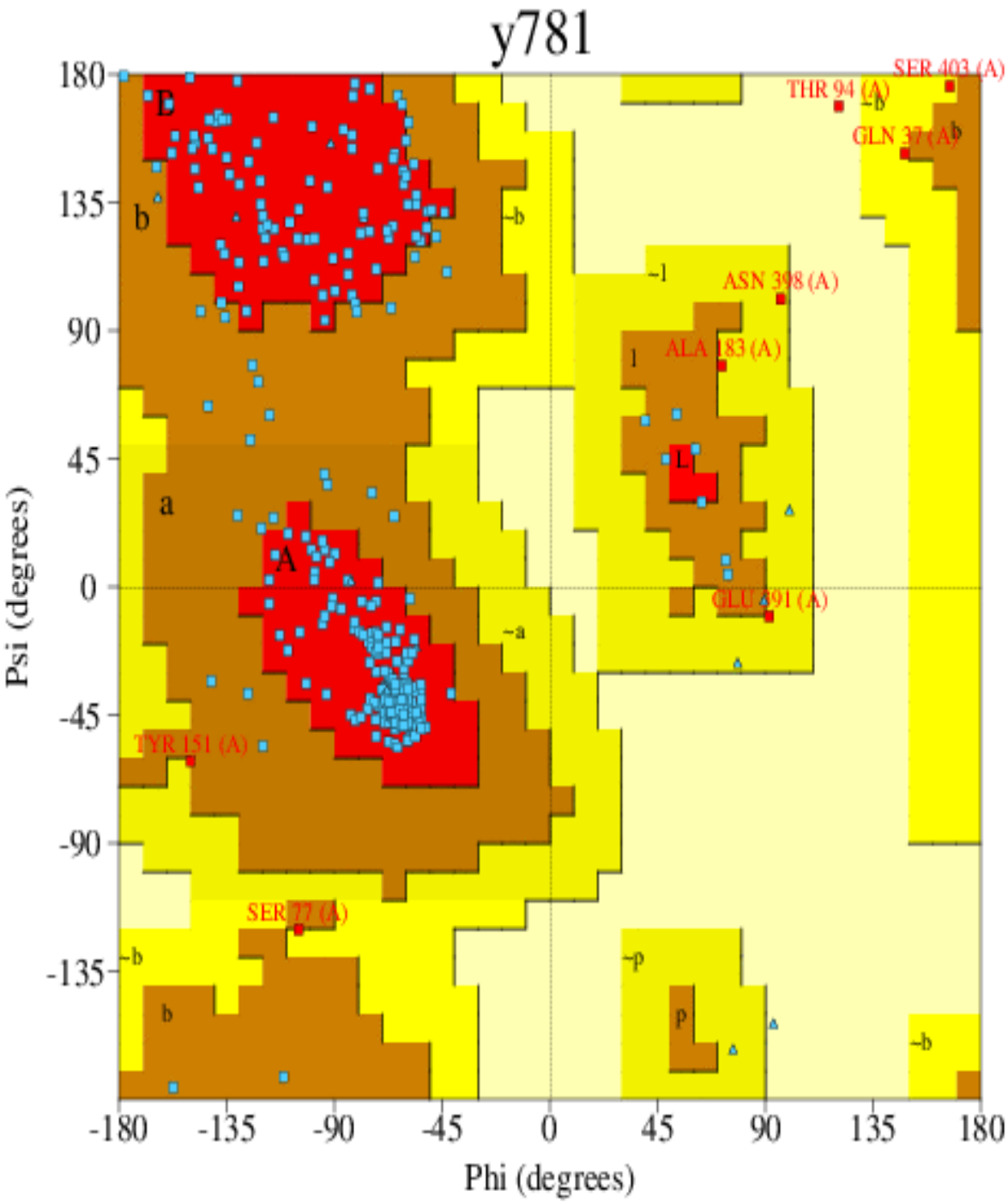


Fig. S5(a): Ramachandra plot statistics of P450CPS1 from *Bacillus cereus* strain AOA-CPS1

Supplementary Materials 5b

Fig. S5(b): Summary of P450CPS1 secondary structure

Strand	Alpha helix	3-10 helix	Other	Total residues
31 (7.9%)	191 (48.6%)	13 (3.3%)	158 (40.2%)	393

Supplementary Materials 5c

Fig. S5(c): Lists of beta sheets in P450CPS1 secondary structure

Sheet	strands	Type	Barrel	Topology
A	3	Mixed	N	1 1X
B	2	Antiparallel	N	1
C	3	Antiparallel	N	2X -1
D	2	Antiparallel	N	1

Supplementary Materials 5d

Fig. S5(d): Lists of hairpins in P450CPS1 structure

Strand 1			Strand 2			Hairpin class
Start	End	Length	Start	End	Length	
Ile39	Asn43	5	Ala49	Ile53	5	3:5
Leu68	Lys69	2	Asn74	Val75	2	4:4 I
Phe308	Ile310	3	Gln313	Ile315	3	2:2 IP
Gln385	Ile386	2	Pro408	Thr410	3	20:22

Supplementary Materials 5e

Fig. S5(e): Beta bulges in P450CPS1 structure

Bulge type	Res X	Res 1	Res 2	Res 3	Res 4
Antiparallel G1	Lys44A	Leu47A	Gly48A		
Antiparallel G1	Lys69A	Ala73A	Asn74A		

Supplementary Materials 5f

Fig. S5(f): Lists of strands found in P450CPS1 secondary structure

S/N	Start	End	Sheet	Number of residues
1	Ile39	Asn43	A	5
2	Ala49	Ile53	A	5
3	Leu68	Lys69	B	2
4	Asn74	Val75	B	2
5	Ser144	Asn146	C	3
6	Phe308	Ile310	D	3
7	Gln313	Ile315	D	3
8	Val321	Ile323	A	3
9	Gln385	Ile386	C	2
10	Pro408	Thr410	C	3

Supplementary Materials 5g

Fig. S5(g): List of helices found in P450CPS1 secondary structure

S/N	Start	End	Type	Number of residues
1	Ala20	Glu24	H	5
2	Tyr27	Met36	H	10
3	Tyr56	Leu63	H	8
4	Thr80	Val83	G	4
5	Asp88	Tyr92	H	5
6	Met97	Asn99	G	3
7	Asn104	Val112	H	9
8	Gln113	Phe116	G	4
9	Pro118	Ile121	H	4
10	Ala122	Leu124	G	3
11	Glu125	Glu140	H	16
12	Leu147	Asp150	H	4
13	Pro154	Leu164	H	11
14	Gln171	Val181	H	11
15	Ile191	Lys214	H	24
16	Leu220	Glu230	H	11
17	Ala236	Glu252	H	17
18	Thr254	Glu267	H	14
19	Pro269	Glu277	H	9
20	Pro279	Leu281	G	3
21	Ile282	Tyr292	H	11
22	Leu325	Ala328	H	4
23	Ile343	Arg345	G	3
24	Ala363	Arg380	H	18

Supplementary Materials 5h

Fig. S5(h): Lists of helix-helix interactions in P450CPS1 secondary structure

S/N	Helices		Helix type		Interaction type		No. of interesting residues	
							Helix 1	Helix 2
1	A1	A2	H	H	I	N	1	2
2	A2	A21	H	H	n	c	1	1
3	A2	A22	H	H	I	I	3	2
4	A3	A22	H	H	n	c	3	2
5	A5	A14	H	H	n	c	1	1
6	A5	A15	H	H	C	I	3	5
7	A7	A8	H	G	I	N	4	4
8	A7	A16	H	H	C	I	2	2
9	A7	A17	H	H	I	I	3	2
10	A8	A9	G	H	C	I	3	3
11	A8	A13	G	H	I	I	1	1
12	A8	A16	G	H	N	I	2	1
13	A8	A17	G	H	N	N	1	1
14	A8	A24	G	H	c	n	1	2
15	A9	A11	H	H	c	n	1	1
16	A9	A24	H	H	n	n	2	3
17	A11	A12	H	H	I	C	2	1
18	A11	A13	H	H	I	I	5	5
19	A11	A24	H	H	I	I	8	7
20	A12	A13	H	H	c	n	3	1
21	A12	A14	H	H	C	N	2	1
22	A12	A18	H	H	C	I	2	4
23	A12	A24	H	H	C	I	1	1
24	A13	A14	H	H	I	I	2	2
25	A13	A16	H	H	c	n	2	2
26	A13	A17	H	H	N	C	5	4
27	A13	A18	H	H	I	I	2	3
28	A13	A24	H	H	I	I	4	3
29	A14	A15	H	H	I	I	5	6
30	A14	A17	H	H	I	I	4	4
31	A15	A16	H	H	I	I	3	4
32	A15	A17	H	H	I	I	6	6
33	A16	A17	H	H	C	N	5	5
34	A17	A18	H	H	c	n	3	3
35	A18	A19	H	H	c	n	3	4
36	A18	A21	H	H	I	I	1	2
37	A18	A24	H	H	I	I	6	7
38	A19	A21	H	H	C	N	1	2
39	A19	A24	H	H	I	I	2	1
40	A21	A24	H	H	I	I	4	5

Supplementary Materials 5i

Fig. S5(i): Lists of beta turns found in P450CPS1 secondary structure

S/N	Turn	Sequence	Turn type	H-bond
1	Lys44-Leu47	KTEL	I	Yes
2	Lys69-Pro72	KKDP	IV	
3	Lys70-Ala73	KDPA	I	
4	Phe76-Asp79	FSQD	IV	
5	Asp101-Asn104	DPPN	VIa1	Yes
6	Ile166-Glu169	IPKE	I	Yes
7	Glu186-Glu189	ETPE	I	Yes
8	His232-Ser235	HKLS	VIII	
9	Ser294-Glu297	SPVE	VIII	
10	Val296-Thr299	VEVT	VIII	
11	Ala304-Pro307	ADEP	IV	
12	Gln309-Asp312	QIHD	IV	
13	Ile310-Gln313	IHDQ	I'	Yes
14	Glu316-Asp319	EKGD	II	Yes
15	Asp331-Val334	DETV	I	
16	Glu332-Phe335	ETVF	I	
17	Thr333-Glu336	TVFE	IV	
18	Phe335-Pro338	FENP	IV	
19	Asn337-Val340	NPEV	I	Yes
20	Asn347-His350	NNRH	I	
21	Ile351-Gly354	IAFG	I	Yes
22	His355-His358	HGSH	IV	
23	Gly356-Phe359	GSHF	I	Yes
24	Cys360-Ala363	CLGA	I	Yes
25	Met381-Leu384	MPKL	I	Yes
26	Asp389-Glu392	DREE	II	Yes
27	Glu391-Lys394	EEIK	IV	
28	Asn398-Met401	NYLM	IV	Yes
29	Leu404-Leu407	LEEL	VIII	

Supplementary Materials 5j

Fig. S5(j): Lists of gamma turns found in P450CPS1 secondary structure

Start	End	Sequence	Turn type
Gly231	Lys233	GHK	INVERSE
Ile323	Leu325	IAL	INVERSE

Supplementary Material 6a

Fig. S6a: Lists of clefts found in P450CPS1 3D structure.

Clefts

	Volume	R1 ratio	Accessible vertices		Buried vertices		Average depth		Residue.type							Ligands
1	6945.75	2.92	75.00	1	13.37	1	17.64	1	13	19	22	47	14	12	1	
2	2381.91	0.00	57.89	8	9.10	3	11.10	2	6	4	6	14	0	5	0	
3	1536.89	0.00	65.46	3	7.97	6	9.27	3	9	5	8	11	4	3	1	
4	1263.52	0.00	65.02	4	8.86	4	7.72	6	2	4	11	7	2	2	0	
5	1252.12	0.00	61.00	6	7.84	7	7.62	7	5	8	4	7	2	2	0	
6	899.86	0.00	61.40	5	9.10	2	7.34	8	5	4	0	2	3	3	0	
7	548.86	0.00	66.69	2	8.47	5	9.05	4	2	0	2	4	1	2	0	
8	526.08	0.00	56.10	9	5.36	8	8.25	5	2	0	2	2	3	2	0	
9	591.89	0.00	58.95	7	4.68	9	5.91	10	2	3	5	2	1	0	0	
10	362.39	0.00	53.76	10	3.19	10	6.63	9	4	2	1	1	1	1	0	

Protein structure

Residue-type_colouring

Positive	Negative	Neutral	Aliphatic	Aromatic	Pro & Gly	Cysteine
H,K,R	D,E	S,T,N,Q	A,V,L,I,M	F,Y,W	P,G	C

Supplementary Material 6b

Fig. S6b: Lists of pores found in P450CPS1 3D structure

Pores

	Radius	Free R	Length	HPathy	HPhob	Polar	Rel Mut	Residue..type							Ligands
1	1.12	2.77	48.9	0.21	-0.13	9.2	85	4	2	3	11	0	1	1	
2	1.17	1.38	49.3	0.98	0.10	5.1	84	2	4	2	17	0	0	1	
3	1.33	2.33	62.9	-0.82	-0.41	7.8	91	0	3	8	4	1	1	0	
4	1.15	2.55	81.8	-0.42	-0.22	8.1	90	4	1	10	11	1	2	0	
5	1.15	1.37	83.2	0.15	-0.10	6.6	88	2	3	7	15	1	1	1	
6	1.35	2.28	99.4	-0.59	-0.43	11.6	91	0	5	8	6	1	0	0	
7	1.13	2.61	114.2	-0.06	-0.21	9.2	87	3	3	8	16	1	1	0	

Residue-type_colouring

Positive	Negative	Neutral	Aliphatic	Aromatic	Pro & Gly	Cysteine
H,K,R	D,E	S,T,N,Q	A,V,L,I,M	F,Y,W	P,G	C

Supplementary Material 7a

```
P450CPS1 ------------------------------------------------------------  0
P14762   ------------------------------------------------------------  0
P38364   MLVDTGLGLISELRARLGW------AAL----------LQIVPVTVVAYNLLWFIY----  40
P17549   ML----ALLLSPYGAYLGL------AL------------------LVLY---YLLP----  25
P28649   MF----LEMLNPMQYNVTI----------------MVPETVTVSAMPLLLIMGLLLLI--  38
P19098   MI----PETLNPLNYF-TS----------------LVPDLMPVATVPIIILICFLFLI--  37
P22680   ------------MMTT-----SLIWGIAI-----------------AACCCLWLILGI--  24
Q02318   -------------MAALGCARLR-WALRGA-GRGLCPHGARAKAAIPA-ALPS---DKAT  41
P05108   ----------------------------ML-AKGLPPRSVLVKGCQTFLSAPREGLGRLR  31
pdb3N9Y  ------------------------------------------------------------  0
P00189   ----------------------------ML-ARGLPLRSALVKACPPILSTVGEGWGHHR  31
pdb3MZS  ------------------------------------------------------------  0
P15539   --------------------------------------MALRVTADVWLARPWQCLHR-T  21
P15150   --------------------------------------MALWAKARVRMAGPWLSLHE-A  21
P27786   -------------MWELV-----------GLLLLI-LAY-----------FFWP------  18
P05185   -------------MWLLL-----------AVFLLT-LAY-----------LFWP------  18
P05093   -------------MWELV-----------ALLLLT-LAY-----------LFWP------  18
pdb3RUK  ------------------------------------------------------------  0
P12394   ------------------M-------PPLAVLLLALALL-----------CAWRLSYS--  22
P30437   ------------MAWFLCM-------CVFSVVGLGLLL----------------LQVK--  23
P10633   -------------MELLNGTGLW--SMAIFTVIFILLVD-----------LMHR------  28
Q01361   -------------MGLLSGDTLG--PLAVALLIFLLLLD-----------LMHR------  28
Q64680   ------------------------------------------------------------  0
P05180   -------------MDFLGL---P--TILLLVCISCLL-I-----------AAWR------  24
pdb3E4E  ----------------------------------------------------MAK------  3
P05181   -------------MSALGV---T---VALLVWAAFLLLV-----------SMWR------  24
P24470   -------------MELLGF---T--TLALVVSVTCLSLL-----------SVWT------  25
P00182   -------------MDLL---------IILGICLSCVVLL-----------SLWK------  21
P19225   -------------MALF---------IFLGIWLSCLVFL-----------FLWN------  21
P00180   -------------MDPV---------VVLGLCLSCLLLL-----------SLWK------  21
P11510   -------------MDPF---------VVLVLSLSFLLLL-----------YLWR------  21
P05179   -------------MDLV---------TFLVLTLSSLILL-----------SLWR------  21
P24903   -------------MDSIST---A--ILLLLLALVCLLLT-----------L---------  22
P11711   -------------MLDTGL---L--LVVILASLSVMLLV-----------SLW-------  24
P24461   -------------MELGGA---F--TIFLALCFSCLLIL-----------IAWK------  25
P00176   ---------------MEPT---I--LLLLALLVG-FLLL-----------LVRG------  22
P37124   -------------MDFFSTSSLSSYYHLIFTILAFVISS-----------IIY--FLS--  32
P37123   ----------------------------IFTAFSLLFSL-----------FIF--LLT--  17
Q04468   -------------MDL----------LLIEKTLVALFAA-----------IIGA-ILI--  23
P37122   -------------MEWE-------WSYVFFSAIIILPAF-----------ILF--FSQ--  25
P37120   -------------MVIL-------PSELIGATIIYIIVY-----------IIIQ-KLI--  26
AJV30500 ---------MSATKSIVGE-ALEYVNIGLSHFLALPLVQRISL----IIIIPFIYNIV--  44
Q08477   -------------MPQLSLSSLGLWPMAASPWLLLLLVGASWLL-----ARILAWTYT--  40
P33274   -------------MSQLSLSWLGLGPEVAFPWQTLLLFGASWIL-----AQILTQIYA--  40
P13584   MV----PSFLS-----LSFSSLGLWASGLILVLG--FLKL------------IHLLLR--  35
pdb5T6Q  ------------------------------MAG--FLKL------------LRLLLR--  13
P20816   MG----FSVFSPTRSLDGVSGFFQGAFLLSLFLV--LFKA------------VQFYLR--  40
P08516   MS----VSALSSTRFTGSISGFLQVASVLGLLLL--LVKA------------VQFYLQ--  40
P29981   ----------------------------MEFITILLSTALFIV---TFLFLFRQGAK--  26
P33269   ----------------------------MFLVIGAILASALFVG---LLLY----HLK--  23
P04800   -------------MDLLSALTLETWV----------LLA--------VVLVLLYGFGT--  27
P11707   ---------------MDLIFSLETWV----------LLA--------ASLVLLYLYGT--  25
Q04552   --------------------MLYLLA----------LVT--------VLAGLLHYYFT--  20
P13527   -------------MDF----GSFLLY----------ALG--------VLASLALYFVR--  23
P33270   --------------------MFVLIY----------LLI--------AISSLLAYLYH--  20
```

Supplementary Material 7a

```
P450CPS1 --------MSMKN---KVGLRIED----------GINLASAQFKED-A-YEIY-KESRKM  36
P14762   ----------MN----KEVIPVTE----------IPKFQSRAEEFF-P-IQWY-KEMLNN  33
P38364   ----------TSFFSSLRKIPGPF-LAR---ISRVWEIKKA-ATGN-I-HEIVMDLHRCH  83
P17549   ----------YLKRAHLRDIPAPG-LAA---FTNFWLLLQT-RRGH-R-FVVVDNAHKKY  68
P28649   -----------WNCESSSSIPGPG-YCL--GIGPLISHGRFLWMGI---GSACNYYNKMY  81
P19098   -----------WNHEETSSIPGPG-YCM--GIGPLISHGRFLWMGV---GNACNYYNKTY  80
P22680   -----------RRRQ-TGEPPLENGLIP--YLGCALQFGAN------P-LEFLRANQRKH  63
Q02318   GAPGAGPG-VRRRQRSLEEIPRLG-QLR--FF---FQLFVQGYALQ-L-HQLQVLYKAKY  92
P05108   VPTGEGAGISTRSPRPFNEIPSPG-DNG--WL-NLYHFWRETGTHK-V-HLHHVQNFQKY  85
pdb3N9Y  ---------STRSPRPFNEIPSPG-DNG--WL-NLYHFWRETGTHK-V-HLHHVQNFQKY  45
P00189   VGTGEGAGISTKTPRPYSEIPSPG-DNG--WL-NLYHFWREKGSQR-I-HFRHIENFQKY  85
pdb3MZS  -------MASTKTPRPYSEIPSPG-DNG--WL-NLYHFWREKGSQR-I-HFRHIENFQKY  47
P15539   RALGTTATLAPKTLQPFEAIPQYS-RNK--WL-KMIQILREQGQEN-L-HLEMHQVFREL  75
P15150   RLLGTRGAAAPKAVLPFEAMPRCP-GNK--WM-RMLQIWKEQSSEN-M-HLDMHQTFQEL  75
P27786   ----------KSKTPNAKFPRSLP-FLP--LVGSLPFLPRR---GH-M-HANFFKLQEKY  60
P05185   ----------KTKHSGAKYPRSLP-SLP--LVGSLPFLPRR---GQ-Q-HKNFFKLQEKY  60
P05093   ----------KRRCPGAKYPKSLL-SLP--LVGSLPFLPRH---GH-M-HNNFFKLQKKY  60
pdb3RUK  ----------MAKKTGAKYPKSLL-SLP--LVGSLPFLPRH---GH-M-HNNFFKLQKKY  42
P12394   ---------QGPTGTGTGRPRSLP-ALP--LVGSLLQLAGH---PQ-L-HLRLWRLQGRY  65
P30437   ---------LRRSLETRGGPPSLP-VFP--LIGSLLSLRSN---QA-P-HVLFQKLQQKY  66
P10633   -----------RHRWTSRYPPGPV-PWP--VLGNLLQVDLS----N-M-PYSLYKLQHRY  68
Q01361   -----------RSRWAPRYPPGPT-PLP--VLGNLLQVDFE----D-P-RPSFNQLRRRF  68
Q64680   ------------------------------------------------------------  0
P05180   -----------STSQRGKEPPGPT-PIP--IIGNVFQLNPW----D-L-MGSFKELSKKY  64
pdb3E4E  -----------KTSSKGKLPPGPF-PLP--IIGNLFQLELK----N-I-PKSFTRLAQRF  43
P05181   -----------QVHSSWNLPPGPF-PLP--IIGNLFQLELK----N-I-PKSFTRLAQRF  64
P24470   -----------KLRTRGRLPPGPT-PLP--IIGNLLQLNLK----D-I-PASLSKLAKEY  65
P00182   -----------KTHGKGKLPPGPT-PLP--VVGNLLQLETK----D-I-NKSLSMLAKEY  61
P19225   -----------QHHVRRKLPPGPT-PLP--IFGNILQVGVK----N-I-SKSMCMLAKEY  61
P00180   -----------QSYGGGKLPPGPT-PFP--ILGNILQIGIQ----D-I-SKSFTKLSEVY  61
P11510   -----------PSPGRGKLPPGPT-PLP--IFGNFLQIDMK----D-I-RQSISNFSKTY  61
P05179   -----------QSSRRRKLPPGPT-PLP--IIGNFLQIDVK----N-I-SQSLTKFSKTY  61
P24903   -----------SSRDKGKLPPGPR-PLS--ILGNLLLLCSQ----D-M-LTSLTKLSKEY  62
P11711   -----------QQKIRGRLPPGPT-PLP--FIGNYLQLNTK----D-V-YSSITQLSERY  64
P24461   -----------RVQKPGRLPPGPT-PIP--FLGNLLQVRTD----A-T-FQSFLKLREKY  65
P00176   -----------HPKSRGNFPPGPR-PLP--LLGNLLQLDRG----G-L-LNSFMQLREKY  62
P37124   ---------KKAESKKLKLPPGPP-GWP--VVGNLLQVARS---GKPF-FQIMRELRQKY  76
P37123   ---------RKPKSKTPNLPPGPP-GWP--IVGNLFQVAGS---GKQF-FEYIRDLKPKY  61
Q04468   ---------SKLRGKKFKLPPGPI-PVP--IFGNWLQVGDD-----LN-HRNLTDLAKRF  65
P37122   ---------KNTTKSSYKFPPGPP-GLP--IFGNMFELGTE------P-YKKMAVLRQKY  66
P37120   ---------ATGSWRRRRLPPGPE-GWP--VIGALPLLGGM------P-HVALAKMAKKY  67
AJV30500 ---------WQLLYSLRKDRPPLV-FYWIPWVGSAVVYGMK-------PYEFFEECQKKY  87
Q08477   ---------FYDNCCRLRCFPQPPKRNW--FLGHLGLIHSS--EEGLL-YTQSLACT--F  84
P33274   ---------AYRNFRRLRGFPQPPKRNW--LMGHVGMVTPT--EQGLK-ELTRLVGT--Y  84
P13584   ---------RQTLAKAMDKFPGPP-THW--LFGHALEIQET---GSLD-KVVSWAHQ--F  77
pdb5T6Q  ---------RQRLARAMDSFPGPP-THW--LFGHALEIQKT---GSLD-KVVTWTQQ--F  55
P20816   ---------RQWLLKALEKFPSTP-SHW--LWGHNL---KD---REFQ-QVLTWVEK--F  79
P08516   ---------RQWLLKAFQQFPSPP-FHW--FFGHKQ-FQGD---KELQ-QIMTCVEN--F  81
P29981   ---------RARFVYLVNKLPGPT-AYP--VVGNAIEAIVP--RNKLF-QVFDRR-AKLY  70
P33269   ---------FKRLIDLISYMPGPP-VLP--LVGHGHHFIGKPPHEMVK-KIFEFMETYSK  70
P04800   ---------RTHGLFKKQGIPGPK-PLP--FFGTVLNYYMGLWKFD-------VECHKKY  68
P11707   ---------STHGLFKKMGIPGPT-PLP--FIGTILEYRKGIWDFD-------IECRKKY  66
Q04552   ---------RTFNYWKKRNVAGPK-PVP--FFGNLKDSVLRRKPQVMVYKSIYDEFP--N  66
P13527   ---------WNFGYWKRRGIPHEE-PHL--VMGNVKGLRSKYHIGEII-ADYYRKFKGS-  69
P33270   ---------RNFNYWNRRGVPHDA-PHP--LYGNMVGFRKNRVMHDFF-YDYYNKYRKSG  67
```

Supplementary Material 7a

```
P450CPS1 QPILFVNKTELGAEWLITRYEDALPLLKDN-RLKKDPANVFSQDTLN--VFLT---VDNS  90
P14762   SPVYFHE---ETNTWNVFQYEHVKQVLSNY-D-------FFSSDGQRTTIFVGDNSKKKS  82
P38364   GPIVRI--GPNRY--DFDTMEALKIIYRIG--------NALPK-ADYY----I----PFG  122
P17549   GKLVRI--APRHT--SIADDGAIQAVYGHG--------NGFLK-SDFY----D----AFV  107
P28649   GEFMRVWISGEET-LIISKSSSMFHVMKHS---------HYISRFGSKR---GL---QCI  125
P19098   GEFVRVWISGEET-FIISKSSSVFHVMKHW---------NYVSRFGSKL---GL---QCI  124
P22680   GHVFTCKLMGKYV-HFITNPLSYHKVLCHG-K--------YFDWKKFHF---A----TSA  106
Q02318   GPMWMSYLGPQMH-VNLASAPLLEQVMRQE-G-------KYPVRNDMEL---WKE-HRDQ  139
P05108   GPIYREKLGNVES-VYVIDPEDVALLFKSE-G-------PNPERFLIPP---WVA-YHQY  132
pdb3N9Y  GPIYREKLGNVES-VYVIDPEDVALLFKSE-G-------PNPERFLIPP---WVA-YHQY  92
P00189   GPIYREKLGNLES-VYIIHPEDVAHLFKFE-G-------SYPERYDIPP---WLA-YHRY  132
pdb3MZS  GPIYREKLGNLES-VYIIHPEDVAHLFKFE-G-------SYPERYDIPP---WLA-YHRY  94
P15539   GPIFRHSVGKTQI-VSVMLPEDAEKLHQVE-S-------MLPRRMHLEP---WVA-HREL  122
P15150   GPIFRYDVGGRHM-VFVMLPEDVERLQQAD-S-------HHPQRMILEP---WLA-YRQA  122
P27786   GPIYSLRLGTTTA-VIVGHYQLAREVLVKK-G------KEFSGRPQMV----TLG-LLSD  107
P05185   GPIYSFRLGSKTT-VMIGHHQLAREVLLKK-G------KEFSGRPKVA----TLD-ILSD  107
P05093   GPIYSVRMGTKTT-VIVGHHQLAKEVLIKK-G------KDFSGRPQMA----TLD-IASN  107
pdb3RUK  GPIYSVRMGTKTT-VIVGHHQLAKEVLIKK-G------KDFSGRPQMA----TLD-IASN  89
P12394   GSLYGLWMGSHYV-VVVNSYQHAREVLLKK-G------KAFAGRPRTV----TTD-LLSR  112
P30437   GHTYSLMMGPHTV-ILVNHHQHAKEVLLKK-G------KIFAGRPRTV----TTD-LLTR  113
P10633   GDVFSLQKGWKPM-VIVNRLKAVQEVLVTH-G------EDTADRPPVP----IFK-CLGV  115
Q01361   GNVFSLQQVWTPV-VVLNGLAAVREALVYR-S------QDTADRPPPA----VYE-HLGY  115
Q64680   --LFSLQLAFESV-VVLNGLPALREALVKY-S------EDTADRPPLH----FND-QSGF  45
P05180   GPIFTIHLGPKKI-VVLYGYDIVKEALIDN-G------EAFSGRGILP----LIE-KLFK  111
pdb3E4E  GPVFTLYVGSQRM-VVMHGYKAVKEALLDY-K------DEFSGRGDLP----AFH--AHR  89
P05181   GPVFTLYVGSQRM-VVMHGYKAVKEALLDY-K------DEFSGRGDLP----AFH--AHR  110
P24470   GPVYTLYFGTSPT-VVLHGYDVVKEALLQQ-G------DEFLGRGPLP----IIE-DTHK  112
P00182   GSIFTLYFGMKPA-VVLYGYEGVIEALIDR-G------EEFSGRGIFP----VFD-RVTK  108
P19225   GPVFTMYLGMKPT-VVLYGYEVLKEALIDR-G------EEFSDKMHSS----MLS-KVSQ  108
P00180   GPVFTVYLGMKPT-VVIHGYDAVKEALVDL-G------EEFSGRIVFP----LTA-KINK  108
P11510   GPVFTLYFGSQPT-VVLHGYEAVKEALIDY-G------EEFSGRGRMP----VFE-KATK  108
P05179   GPVFTLYLGSQPT-VILHGYEAIKEALIDN-G------EKFSGRGSYP----MNE-NVTK  108
P24903   GSMYTVHLGPRRV-VVLSGYQAVKEALVDQ-G------EEFSGRGDYP----AFF-NFTK  109
P11711   GPVFTIHLGPRRV-VVLYGYDAVKEALVDQ-A------EEFSGRGEQA----TYN-TLFK  111
P24461   GPVFTVYMGPRPV-VILCGHEAVKEALVDR-A------DEFSGRGELA----SVE-RNFQ  112
P00176   GDVFTVHLGPRPV-VMLCGTDTIKEALVGQ-A------EDFSGRGTIA----VIE-PIFK  109
P37124   GPIFTLRMGTRTM-IILSNADLVHEALILK-G------QVFATRPRENP---TRT-VFSC  124
P37123   GSIFTLKMGSRTM-IIVASAELAHEALIQK-G------QIFASRPRENP---TRT-IFSC  109
Q04468   GEILLLRMGQRNL-VVVSSPELAKEVLHTQ-G------VEFGSRTRNV----VFD-IFTG  112
P37122   GPVLWLKLGSTYT-MVVQTAQASEELFKNH-D------ISFANRVIPD----VNQ-AHSY  113
P37120   GPIMYLKVGTCGM-VVASTPNAAKAFLKTL-D------INFSNRPPNA----GAT-HMAY  114
AJV30500 GDIFSFVLLGRVM-TVYLGPKGHEFVFNAK-L------ADVSAEAAYA----H----LTT  131
Q08477   GDMCCWWVGPWHAIVRIFHPTYIKPVLFAP-A------AIVPKDKVFY----S----FLK  129
P33274   PQGFLMWIGPMVPVITLCHSDIVRSILNAS-A------AVALKDVIFY----T----ILK  129
P13584   PYAHPLWFGQFIGFLNIYEPDYAKAVYSRG-D---------PKAPDVY----D----FFL  119
pdb5T6Q  PYAHPLWVGQFIGFLNIYEPDYAKAVYSRG-D---------PKAPDVY----D----FFL  97
P20816   PGACLQWLSGSTARVLLYDPDYVKVVLGRS-D---------PKP---Y----Q----SLA  118
P08516   PSAFPRWFWGSKAYLIVYDPDYMKVILGRS-D---------PKANGVY----R----LLA  123
P29981   GPLYRIWAGPIAQ-VGLTRPEHVELILRDT-K--------HIDKSLVY----S----FIR  112
P33269   DQVLKVWLGPELN-VLMGNPKDVEVVLGTL-R--------FNDKAGEY----K----ALE  112
P04800   GKIWGLFDGQMPL-FAITDTEMIKNVLVKECF------SVFTNRRDFG----P------V  111
P11707   GKMWGLFDGRQPL-MVITDPDMIKTVLVKECY------SVFTNRRSFG----P------V  109
Q04552   EKVVGIYRMTTPS-VLLRDLDIIKHVLIKD-F------ESFADRGVEF----S------L  108
P13527   DPLPGIFLGHKPA-AVVLDKELRKRVLIKD-F------SNFANRGLYY----N----EKD  113
P33270   FPFVGFYFLHKPA-AFIVDTQLAKNILIKD-F------SNFADRGQFH----N----GRD  111
```

Supplementary Material 7a

```
P450CPS1 DYLTTH-MLNSDPPNHNRLRSLVQKAF-----TPKMIAQLEGRIQHIADDLLNEVERK--  142
P14762   TSPITN-LTNLDPPDHRKARSLLAAAF-----TPRSLKNWEPRIKQIAADLVEAIQKN--  134
P38364   LPSSPNLFDVQNPARHSAMKKQVASLY-----TMTALLSYEAGVDGQTIILKEQLQRFCD  177
P17549   S-IHRGLFNTRDRAEHTRKRKTVSHTF-----SMKSIGQFEQYIHGNIELFVKQWNRMAD  161
P28649   GMHENGIIFNNNPSLWRTIRPFFMKAL-----TGPGLVRMVEVCVESIKQHLDRLGEV--  178
P19098   GMYENGIIFNNNPAHWKEIRPFFTKAL-----SGPGLVRMIAICVESTIVHLDKLEEV--  177
P22680   KAFGHRSIDPMDGNTTENINDTFIKTL-----QGHALNS-------LTESMMENLQRIMR  154
Q02318   HDLTYGPFTT-EGHHWYQLRQALNQRL----LKPAEAALYTDAFNEVIDDFMTRLDQLRA  194
P05108   YQRPIGVLLK-KSAAWKKDRVALNQEV----MAPEATKNFLPLLDAVSRDFVSVLHRRIK  187
pdb3N9Y  YQRPIGVLLK-KSAAWKKDRVALNQEV----MAPEATKNFLPLLDAVSRDFVSVLHRRIK  147
P00189   YQKPIGVLFK-KSGTWKKDRVVLNTEV----MAPEAIKNFIPLLNPVSQDFVSLLHKRIK  187
pdb3MZS  YQKPIGVLFK-KSGTWKKDRVVLNTEV----MAPEAIKNFIPLLNPVSQDFVSLLHKRIK  149
P15539   RGLRRGVFLL-NGPEWRLNRLRLNRNV----LSPKAVQKFVPMVDMVARDFLETLKEKVL  177
P15150   RGHKCGVFLL-NGPQWRLDRLRLNPDV----LSLPALQKYTPLVDGVARDFSQTLKARVL  177
P27786   Q--GKGVAFADSSSSWQLHRKLVFSTFSLFRD-DQKLEKMICQE---ANSLCDLILT---  158
P05185   N--QKGIAFADHGAHWQLHRKLALNAFALFKDGNLKLEKIINQE---ANVLCDFLAT---  159
P05093   N--RKGIAFADSGAHWQLHRRLAMATFALFKDGDQKLEKIICQE---ISTLCDMLAT---  159
pdb3RUK  N--RKGIAFADSGAHWQLHRRLAMATFALFKDGDQKLEKIICQE---ISTLCDMLAT---  141
P12394   G--GKDIAFASYGPLWKFQRKLVHAALSMFGEGSVALEKIICRE---AASLCETLGA---  164
P30437   D--GKDIAFADYGATWRFHRKTVHGALCMFGEGSASIEKIICRE---ALSLCDTLRE---  165
P10633   KPRSQGVILASYGPEWREQRRFSVSTLRTFGMGKKSLEEWVTKE---AGHLCDAFTA---  169
Q01361   GPRAEGVILARYGDAWREQRRFSLTTLRNFGLGKKSLEQWVTEE---ASCLCAAFAD---  169
Q64680   GPRSQGVVLARYGPAWRQQRRFSVSTFRHFGLGKKSLEQWVTEE---ARCLCAAFAD---  99
P05180   G---TG-IVTSNGETWRQLRRFALTTLRDFGMGKKGIEERIQEE---AHFLVERIRK---  161
pdb3E4E  D---RG-IIFNNGPTWKDIRRFSLTTLRNYGMGKQGNESRIQRE---AHFLLEALRK---  139
P05181   D---RG-IIFNNGPTWKDIRRFSLTTLRNYGMGKQGNESRIQRE---AHFLLEALRK---  160
P24470   G---YG-LIFSNGERWKVMRRFSLMTLRNFGMGKRSLEERVQEE---ARCLVEELQK---  162
P00182   G---LG-IVFSSGEKWKETRRFSLTVLRNLGMGKKTIEERIQEE---ALCLIQALRK---  158
P19225   G---LG-IVFSNGEIWKQTRRFSLMVLRSMGMGKRTIENRIQEE---VVYLLEALRK---  158
P00180   G---YG-IVFSNGKRWKETRRFSLMTLRDFGMGKRSIEDRVQEE---ARCLVEELRK---  158
P11510   G---LG-ISFSRGNVWRATRHFTVNTLRSLGMGKRTIEIKVQEE---AEWLVMELKK---  158
P05179   G---FG-IVFSNGNRWKEMRRFTIMNFRNLGIGKRNIEDRVQEE---AQCLVEELRK---  158
P24903   G---NG-IAFSSGDRWKVLRQFSIQILRNFGMGKRSIEERILEE---GSFLLAELRK---  159
P11711   G---YG-VAFSSGERAKQLRRLSIATLRDFGVGKRGVEERILEE---AGYLIKMLQG---  161
P24461   G---HG-VALANGERWRILRRFSLTILRDFGMGKRSIEERIQEE---AGYLLEEFRK---  162
P00176   E---YG-VIFANGERWKALRRFSLATMRDFGMGKRSVEERIQEE---AQCLVEELRK---  159
P37124   D--KFTVNAAVYGPVWRSLRKNMVQNG--L--SSIRLKEFRAVRKSAMDKMIEKIRAE-A  177
P37123   N--KFSVNAAVYGPVWRSLRRNMVQNM--L--SPSRLKEFREFREIAMDKLIERIRVD-A  162
Q04468   K--GQDMVFTVYGEHWRKMRRIMTVPF--F--TNKVVQQYRYGWEAEAAAVVDDVK---K  163
P37122   Y--QGSLAIAPYGPFWRFQRRICTIEM--F--VHKKISETEPVRRKCVDNMLKWIEKE-A  166
P37120   N--AQDMVFAPYGPRWKLLRKLSNLHM--L--GGKALENWANVRANELGHMLKSMFD---  165
AJV30500 PVFGKGVIYDCPNSRLMEQKKFVKGAL-----TKEAFKSYVPLIAEEVYKYFRDSKNF-R  185
Q08477   PWLGDGLLLS-AGEKWSRHRRMLTPAF-----HFNILKPYMKIFNESVNIMHAKWQLL--  181
P33274   PWLGDGLLVS-AGDKWSRHRRMLTPAF-----HFNILKPYVKIFNDSTNIMHAKWKRL--  181
P13584   QWIGRGLLVL-EGPKWLQHRKLLTPGF-----HYDVLKPYVAVFTESTRIMLDKWEEK--  171
pdb5T6Q  QWIGKGLLVL-DGPKWFQHRKLLTPGF-----HYDVLKPYVAIFADSTRIMLEKWEKK--  149
P20816   PWIGYGLLLL-NGKKWFQHRRMLTPAF-----HYDILKPYVKIMADSVSIMLDKWEKL--  170
P08516   PWIGYGLLLL-NGQPWFQHRRMLTPAF-----HYDILKPYVKNMADSIRLMLDKWEQL--  175
P29981   PWLGEGLLTG-TGAKWHSHRKMITPTF-----HFKILDIFVDVFVEKSEILVKKLQ----  162
P33269   PWLKEGLLVS-RGRKWHKRRKIITPAF-----HFKILDQFVEVFEKGSRDLLRNMEQD--  164
P04800   GIMGKAVSVA-KDEEWKRYRALLSPTF-----TSGRLKEMFPIIEQYGDILVKYLKQE-A  164
P11707   GFMKKAVSIS-EDEDWKRVRTLLSPTF-----TSGKLKEMLPIIAQYGDVLVKNLRQE-A  162
Q04552   DGLGANIFHA-DGDRWRSLRNRFTPLF-----TSGKLKSMLPLMSQVGDRFINSIDEV-S  161
P13527   DPLTGHLVMV-EGEKWRSLRTKLSPTF-----TAGKMKYMYNTVLEVGQRLLEVMYEK-L  166
P33270   DPLTQHLFNL-DGKKWKDMRQRLTPTF-----TSGKMKFMFPTVIKVSEEFVKVITEQ-V  164
```

Supplementary Material 7a

```
P450CPS1 ---------GSLNL-----VDDYSFPLPIIVISEMLGIPKEDQAK---------------  173
P14762   ---------STINI-----VDDLSSPFPSLVIADLFGVPVKDRYQ---------------  165
P38364   QK-------QVIDL-----PQFLQYYAFDVIGVITVGKSMGMMETNSDTNGACGALDAMW  225
P17549   TQRNPKTGFASLDA-----LNWFNYLAFDIIGDLAFGAPFGMLDKGKDFAEMRKTPD PP  216
P28649   -----TDTSGYVDV-----LTLMRHIMLDTSNMLFLGIPLDESAI------VKKI----Q  218
P19098   -----TTEVGNVNV-----LNLMRRIMLDTSNKLFLGVPLDESAI------VLKI----Q  217
P22680   PP--VSSNSKTAAWVTEGMYSFCYRVMFEAGYLTIFGRDLTRRDTQ-----KAHILN-NL  206
Q02318   ES---ASGNQVSDM-----AQLFYYFALEAICYILFEKRIGCLQRS-IPEDTVTFVRSIG  245
P05108   KA---GSGNYSGDI-----SDDLFRFAFESITNVIFGERQGMLEEV-VNPEAQRFIDAIY  238
pdb3N9Y  KA---GSGNYSGDI-----SDDLFRFAFESITNVIFGERQGMLEEV-VNPEAQRFIDAIY  198
P00189   QQ---GSGKFVGDI-----KEDLFHFAFESITNVMFGERLGMLEET-VNPEAQKFIDAVY  238
pdb3MZS  QQ---GSGKFVGDI-----KEDLFHFAFESITNVMFGERLGMLEET-VNPEAQKFIDAVY  200
P15539   QN---ARGSLTMDV-----QQSLFNYTIEASNFALFGERLGLLGHD-LSPGSLKFIHALH  228
P15150   QN---ARGSLTLDI-----APSVFRYTIEASTLVLYGERLGLLTQQ-PNPDSLNFIHALE  228
P27786   ------YDGESRDL-----STLIFKSVINIICTICFNISFENKDPI-----LTTIQTFTE  202
P05185   ------QHGEAIDL-----SEPLSLAVTNIISFICFNFSFKNEDPA-----LKAIQNVND  203
P05093   ------HNGQSIDI-----SFPVFVAVTNVISLICFNTSYKNGDPE-----LNVIQNYNE  203
pdb3RUK  ------HNGQSIDI-----SFPVFVAVTNVISLICFNTSYKNGDPE-----LNVIQNYNE  185
P12394   ------AQDMALDM-----APELTRAVTNVVCSLCFNSSYRRGDPE-----FEAMLEYSQ  208
P30437   ------SGSASLDL-----SPELTRAVTNVVCSLCFSSSYCRGDPE-----FEAMLQFSQ  209
P10633   ------QAGQSINP-----KAMLNKALCNVIASLIFARRFEYEDPY-----LIRMVKLVE  213
Q01361   ------QAGRPFSP-----MDLLNKAVSNVIASLTFGCRFEYNDPR-----IIKLLDLTE  213
Q64680   ------HSGFPFSP-----NTLLDKAVCNVIASLLFACRFEYNDPR-----FIRLLDLLK  143
P05180   ------THEEPFNP-----GKFLIHAVANIICSIVFGDRFDYEDKK-----FLDLIEMLE  205
pdb3E4E  ------TQGQPFDP-----TFLIGCAPCNVIADILFRKHFDYNDEK-----FLRLMYLFN  183
P05181   ------TQGQPFDP-----TFLIGCAPCNVIADILFRKHFDYNDEK-----FLRLMYLFN  204
P24470   ------TKAQPFDP-----TFILACAPCNVICSILFNDRFQYNDKT-----FLNLMDLLN  206
P00182   ------TNASPCDP-----TFLLFCVPCNVICSVIFQNRFDYDDEK-----FKTLIKYFH  202
P19225   ------TNGSPCDP-----SFLLACVPCNVISSVIFQHRFDYSDEK-----FQKFIENFH  202
P00180   ------TNGSPCNP-----TFILGAAPCNVICSVIFQNRFDYTDQD-----FLSLMGKLN  202
P11510   ------TKGSPCDP-----KFIIGCAPCNVICSIIFQNRFDYKDKD-----FLSLIENVN  202
P05179   ------TKGSPCDP-----SLILNCAPCNVICSITFQNHFDYKDKE-----MLTFMEKVN  202
P24903   ------TEGEPFDP-----TFVLSRSVSNIICSVLFGSRFDYDDER-----LLTIIRLIN  203
P11711   ------TCGAPIDP-----TIYLSKTVSNVISSIVFGERFDYEDTE-----FLSLLQMMG  205
P24461   ------TKGAPIDP-----TFFLSRTVSNVISSVVFGSRFDYEDKQ-----FLSLLRMIN  206
P00176   ------SQGAPLDP-----TFLFQCITANIICSIVFGERFDYTDRQ-----FLRLLELFY  203
P37124   D------ANEGVVWV----LKNARFAVFCILLAMCFGVEMDE-KT--IEKIDQMMKS---  221
P37123   K------ENNDVVWA----LKNARFAVFYILVAMCFGVEMDNEEM--IERVDQMMKD---  207
Q04468   NP---AAATEGIVI-----RRRLQLMMYNNMFRIMFDRRFESEDDP----LFLKLKALNG  211
P37122   NS---AEKGSGIEV-----TRFVFLASFNMLGNLILSKDLADLESEEASEFFIAMKRI--  216
P37120   -A---SHVGERIVV-----ADMLTFAMANMIGQVMLSKRVFVEKGKEVNEFKNMVVEL--  214
AJV30500 LN---ERTTGTIDV-----MVTQPEMTIFTASRSLLGKEMRAKL---DTDFAYLYSD---  231
Q08477   -A---SEGSARLDM-----FEHISLMTLDSLQKCVFSFDSH--CQEKPSEYIAAILELSA  230
P33274   -I---SEGSSRLDM-----FEHVSLMTLDSLQKCVFSFDSN--CQEKSSEYIAAILELSA  230
P13584   -A---REGK-SFDI-----FCDVGHMALNTLMKCTFGRG-DTGLGHRDSSYYLAVSDLTL  220
pdb5T6Q  -A---CEGK-SFDI-----FSDVGHMALDTLMKCTFGKG-DSGLNHRDSSYYVAVSELTL  198
P20816   -D---DQDH-PLEI-----FHYVSLMTLDTVMKCAFSHQGSVQLDVNSRSYTKAVEDLNN  220
P08516   -A---GQDS-SIEI-----FQHISLMTLDTVMKCAFSHNGSVQVDGNYKSYIQAIGNLND  225
P29981   -S---KVGGKDFDI-----YPFITHCALDIICETAMGIQMNAQ-EESESEYVKAVYEISE  212
P33269   RL---KHGESGFSL-----YDWINLCTMDTICETAMGVSINAQ-SNADSEYVQAVKTISM  215
P04800   E------TGKPVTM-----KKVFGAYSMDVITSTSFGVNVDSLN-NPKDPFVEKTKK--L  210
P11707   E------KGKPVDL-----KEIFGAYSMDVITGTSFGVNIDSLR-NPQDPFVKNVRR--L  208
Q04552   Q------TQPEQSI-----HNLVQKFTMTNIAACVFGLNLDEGM---LKTLEDLDKH--I  205
P13527   EV------SSELDM-----RDILARFNTDVIGSVAFGIECNSLR-NPHDRFLAMGRK--S  212
P33270   PA---AQNGAVLEI-----KELMARFTTDVIGTCAFGIECNTLR-TPVSDFRTMGQK--V  213
```

Supplementary Material 7a

```
P450CPS1 --------FRIWSHAVIAYPETPE---EIK----E-TE-----K------QLSEFI-TYL  205
P14762   --------FKKWVDILFQPY-DQE---RLEEIEQE-KQ-----R------AGAEYF-QYL  200
P38364   HYSSMMAFIPHMHAWWLRLSSLLPIDVPIKGLTEY-VEQRIIQY----------------  268
P17549   SYVQAVEVLNRRGE----VSATLGCYPALKPFAKY-LPDSFFRDGIQAVEDLAGIAVARV  271
P28649   GY------FNAWQALLIKP-NIFF---KISWLYRK-YERSV--K------DLKDEIAVLV  259
P19098   NY------FDAWQALLLKP-DIFF---KISWLCKK-YEEAA--K------DLKGAMEILI  258
P22680   DN-------------FKQFDKVFP---ALVAGLPIH---MF--R------TAHNAREKLA  239
Q02318   LM------FQNSLYATFLPKWTRP-----VLPFWKRYLDGW--N------AIFSFGKKLI  286
P05108   QM------FHTSVPMLNLPPDLFR---LFRTKTWKDHVAAW--D------VIFSKADIYT  281
pdb3N9Y  QM------FHTSVPMLNLPPDLFR---LFRTKTWKDHVAAW--D------VIFSKADIYT  241
P00189   KM------FHTSVPLLNVPPELYR---LFRTKTWRDHVAAW--D------TIFNKAEKYT  281
pdb3MZS  KM------FHTSVPLLNVPPELYR---LFRTKTWRDHVAAW--D------TIFNKAEKYT  243
P15539   SM------FKSTSQLLFLPKSLTR---WTSTRVWKEHFDAW--D------VISEYANRCI  271
P15150   AM------LKSTVQLMFVPRRLSR---WMSTNMWREHFEAW--D------YIFQYANRAI  271
P27786   GI------VDVLG--HSDLVDIFP---WL-KIFPNKNLEMI--K------EHTKIREKTL  242
P05185   GI------LEVLS--KEVLLDIFP---VL-KIFPSKAMEKM--K------GCVQTRNELL  243
P05093   GI------IDNLS--KDSLVDLVP---WL-KIFPNKTLEKL--K------SHVKIRNDLL  243
pdb3RUK  GI------IDNLS--KDSLVDLVP---WL-KIFPNKTLEKL--K------SHVKIRNDLL  225
P12394   GI------VDTVA--KESLVDIFP---WL-QIFPNRDLALL--K------RCLKVRDQLL  248
P30437   GI------VDTVA--KDSLVDIFP---WL-QVFPNADLRLL--K------QCVSIRDKLL  249
P10633   ES------LTEVSGFIPEVLNTFP---ALL-RIPGLADKVF--Q------GQKTFM-ALL  254
Q01361   DG------LKEEFNLVRKVVEAVP---VLL-SIPGLAARVF--P------AQKAFM-ALI  254
Q64680   DT------LEEESGFLPMLLNVFP---MLL-HIPGLLGKVF--S------GKKAFV-AML  184
P05180   EN------NKYQNRIQTLLYNFFP---TILDSLPGPHKTLI--K------NTETVD-DFI  247
pdb3E4E  EN------FHLLSTPWLQLYNNFP---SFLHYLPGSHRKVI--K------NVAEVK-EYV  225
P05181   EN------FHLLSTPWLQLYNNFP---SFLHYLPGSHRKVI--K------NVAEVK-EYV  246
P24470   KN------FQQVNSVWCQMYNLWP---TIIKYLPGKHIEFA--K------RIDDVK-NFI  248
P00182   EN------FELLGTPWIQLYNIFP---IL-HYLPGSHRQLF--K------NIDGQI-KFI  243
P19225   TK------IEILASPWAQLCSAYP---VL-YYLPGIHNKFL--K------DVTEQK-KFI  243
P00180   EN------FKILNSPWVQMCNNFP---ILIDYLPGSHNKIL--R------NNIYIR-NYV  244
P11510   EY------IKIVSTPAFQVFNAFP---ILLDYCPGNHKTHS--K------HFAAIK-SYL  244
P05179   EN------LKIMSSPWMQVCNSFP---SLIDYFPGTHHKIA--K------NINYMK-SYL  244
P24903   DN------FQIMSSPWGELYDIFP---SLLDWVPGPHQRIF--Q------NFKCLR-DLI  245
P11711   QM------NRFAASPTGQLYDMFH---SVMKYLPGPQQQII--K------VTQKLE-DFM  247
P24461   ES------FIEMSTPWAQLYDMYS---GVMQYLPGRHNRIY--Y------LIEELK-DFI  248
P00176   RT------FSLLSSFSSQVFEFFS---GFLKYFPGAHRQIS--K------NLQEIL-DYI  245
P37124   ----------VLIALDPRLDDYLP---ILSPFFS-KQRKHA--M------DVRKQQIKTI  259
P37123   ----------VLIVLDPRIDDFLP---ILRLFVGYKQRKRV--N------EVRKRQIETL  246
Q04468   ER------SRLAQSFEYNYGDFIP---ILRPFLRNYLKL-C--K------EVKDKRIQLF  253
P37122   ----------NEWSGIANVSDIFP---FLKKFDLQSLRKKM--A------RDMGKAVEIM  255
P37120   ----------MTVAGYFNIGDFIP---QIAWMDLQGIEKGM--K------KLHKKFDDLL  253
AJV30500 -----------LDKGFTPI-NFVF---PNLPLEHY-RKRDHAQK------AISGTYMSLI  269
Q08477   LV------TKRHQQILLYI-DFLY---YLTPDGQR-FRRAC--R------LVHDFTDAVI  271
P33274   LV------AKRHQQPLLFM-DLLY---NLTPDGMR-FHKAC--N------LVHEFTDAVI  271
P13584   LM------QQRLVSFQYHN-DFIY---WLTPHGRR-FLRAC--Q------VAHDHTDQVI  261
pdb5T6Q  LM------QQRIDSFQYHN-DFIY---WLTPHGRR-FLRAC--R------AAHDHTDRVI  239
P20816   LI------FFRVRSAFYGN-SIIY---NMSSDGRL-SRRAC--Q------IAHEHTDGVI  261
P08516   LF------HSRVRNIFHQN-DTIY---NFSSNGHL-FNRAC--Q------LAHDHTDGVI  266
P29981   LT------MQRSVRPWLHP-KVIF---DLTTMGKR-YAECL--R------ILHGFTNKVI  253
P33269   VL------HKRMFNILYRF-DLTY---MLTPLARA-EKKAL--N------VLHQFTEKII  256
P04800   LR------FDFFDPLFLSV-VLFP---FLTPIYEM-LNICMFPK------DSIEFFKKFV  253
P11707   LK------FSFFDPLLLSI-TLFP---FLTPIFEA-LHISMFPK------DVMDFLKTSV  251
Q04552   FT------VNYSAELD-------M---MYPGILKK-LNGSLFPK------VVSKFFDNLT  242
P13527   IE------VPRHNALIMAF---ID---SFPELSRK-LGMRVLPE------DVHQFFMSSI  253
P33270   FT------DMRHGKLLTMF---VF---SFPKLASR-LRMRMMPE------DVHQFFMRLV  254
```

Supplementary Material 7a

```
P450CPS1 QYLVD-M-----------------KRKEPKEDLVSALILAES---------------EGH  232
P14762   YPIVI-E-----------------KRSNLSDDIISDLIQAEV---------------DGE  227
P38364   --RLK----AA-EF----GD---DDALKGENNFLAKLILME-----RQ----------GT  299
P17549   NERLR----PE-VM----AN---N----TRVDLLARLMEGK----DSN----------GE  301
P28649   EKKRH-KVSTA-------------EKLEDCMDFATDLIFAE----------------RRG  289
P19098   EQKRQ-KLSTV-------------EKLDEHMDFASQLIFAQ----------------NRG  288
P22680   ESLRH-E-------------------NLQKRESISELISLRMFLNDT-----------LS  268
Q02318   DEKLE-DMEAQLQAA--------GPDGIQVSGYLHFLLA-------------------SG  318
P05108   QNFYW-EL----RQK--------GSVHHDYRGILYRLLG-------------------DS  309
pdb3N9Y  QNFYW-EL----RQK--------GSVHHDYRGILYRLLG-------------------DS  269
P00189   EIFYQ-DL----RRK--------TE-FRNYPGILYCLLK-------------------SE  308
pdb3MZS  EIFYQ-DL----RRK--------TE-FRNYPGILYCLLK-------------------SE  270
P15539   WKVHQ-EL----RLG--------S--SQTYSGIVAELIS-------------------QG  297
P15150   QRIYQ-EL----ALG--------H--PWHYSGIVAELLM-------------------RA  297
P27786   VEMFE-KCKEK------------FNSE-SLSSLTDILIQAKMNAENNN---TGE-GQDPS  284
P05185   NEILE-KCQEN------------FSSD-SITNLLHILIQAKVNADNNN---AGP-DQDSK  285
P05093   NKILE-NYKEK------------FRSD-SITNMLDTLMQAKMNSDNGN---AGP-DQDSE  285
pdb3RUK  NKILE-NYKEK------------FRSD-SITNMLDTLMQAKMNSDNGN---AGP-DQDSE  267
P12394   QQKFT-EHKEA------------FCGD-TVRDLMDALLQVRLNAENNS---PLEP---GL  288
P30437   QKKYE-EHKSD------------YSDH-EQRDLLDALLRAKRSAENNN---TAEITMETV  292
P10633   DNLLA-ENRTT------------WDPAQPPRNLTDAFLA-EVEKAKGN---P------ES  291
Q01361   DELIA-EQKMT------------RDPTQPPRHLTDAFLD-EVKEAKGN---P------ES  291
Q64680   DELLT-EHKVT------------WDPAQPPRDLTDAFLA-EVEKAKGN---P------ES  221
P05180   KEIVI-AHQES------------FDAS-CPRDFIDAFIN-KMEQEKEN-----------S  281
pdb3E4E  SERVK-EHHQS------------LDPN-CPRDLTDCLLV-EMEKEKHS---A------ER  261
P05181   SERVK-EHHQS------------LDPN-CPRDLTDCLLV-EMEKEKHS---A------ER  282
P24470   LEKVK-EHQKS------------LDPA-NPRDYIDCFLS-KIEEEKDN---L------KS  284
P00182   LEKVQ-EHQES------------LDSN-NPRDFVDHFLI-KMEKEKHK---K------QS  279
P19225   LMEIN-RHRAS------------LNLS-NPQDFIDYFLI-KMEKEKHN---E------KS  279
P00180   LEKIK-EHQET------------LDIN-NPRDFIDCFLI-KMEQEKDN---Q------QS  280
P11510   LKKIK-EHEES------------LDVS-NPRDFIDYFLI-QRCQENGN---Q------QM  280
P05179   LKKIE-EHQES------------LDVT-NPRDFVDYYLI-KQKQANNI---E------QS  280
P24903   AHSVH-DHQAS------------LDPR-SPRDFIQCFLT-KMAEEKED---P------LS  281
P11711   IEKVR-QNHST------------LDPN-SPRNFIDSFLI-RMQEEKN----G------NS  282
P24461   AARVK-VNEAS------------LDPQ-NPRDFIDCFLI-KMHQDKNN---P------HT  284
P00176   GHIVE-KHRAT------------LDPS-APRDFIDTYLL-RMEKEKSN---H------HT  281
P37124   VPFIEQRKKILESPE--------IDKTAASFSYLDTLFDLKIEGRNS-----------T-  299
P37123   VPLIEKRRSVVQNPG--------SDKTAASFSYLDTLFDVKVEGRKS-----------G-  286
Q04468   KDYFVDERKKIGSTK--------KMDNNQLKCAIDHILEAK-----E-----------KG  289
P37122   SMFLK-EREEER-----------KKGTEKGKDFLDVLLEFQGTGKDE-----------PA  292
P37120   TKMFE-EHEAT------------SNERKGKPDFLDFIMANR--DNSE-----------GE  287
AJV30500 KERRK-NN-------------------DIQDRDLIDSLMKNSTYK-------------DGV  297
Q08477   QERRR-TLPSQGVDDFL-----QAKAKSKTLDFIDVLLLSKDED--------------GK  311
P33274   RERRR-TLPDQGLDEFL-----KSKAKSKTLDFIDVLLLTKDED--------------GK  311
P13584   RERKA-ALQDEKVRKK--------IQNRRHLDFLDILLGARDED--------------DI  298
pdb5T6Q  RQRKA-ALQDEKEREK--------IQNRRHLDFLDILLDVRGES--------------GV  276
P20816   KTRKA-QLQNEEELQK--------ARKKRHLDFLDILLFAKMED--------------GK  298
P08516   KLRKD-QLQNAGELEK--------VKKKRRLDFLDILLLARMEN--------------GD  303
P29981   QERKS-LRQMTGMKPTISNEEDELLGKKKRLAFLDLLLEA-SEN--------------GT  297
P33269   VQRRE-ELIREGSSQE-SSNDDADVGAKRKMAFLDILLQS-TVD--------------ER  299
P04800   YRMKE-TRLD--------------SVQKHRVDFLQLMMNAHNDSKDKE---------SHT  289
P11707   EKIKD-DRLK--------------DKQKRRVDFLQLMINSQN-SKEID---------SHK  286
Q04552   KNVLE-MRKG--------------T-PSYQKDMIDLIQELREKKTLELSRKHENEDVKAL  286
P13527   KETVD-YREK--------------N-NIRRNDFLDLVLDLKNNPESI----S-----KLG  288
P33270   NDTIA-LRER--------------E-NFKRNDFMNLLIELKQKGRVTLDNGE-----VIE  293
```

Supplementary Material 7a

```
P450CPS1 KLSARELYSMIMLLIVAGHETTVNLITNTVLALLENPNQ-LQ------------------  273
P14762   TFTDEEIVHATMLLLGAGVETTSHAIANMFYSFLYDDKSLYS------------------  269
P38364   -VTSTETQQAVGLNIGAGSDTTANALSSILYFLYTNPRTLRRLREELDTHVKEDP-----  353
P17549   KLGRAELTAEALTQLIAGSDTTSNTSCAILYWCMRTPGVIEKLHKALDEAIPQDV-----  356
P28649   DLTKENVNQCILEMLIAAPDTMSVTLYFMLLLVAEYPEVEAAILKEIHTVVGDRD-----  344
P19098   DLTAENVNQCVLEMMIAAPDTLSVTLFIMLILIADDPTVEEKMMREIETVMGDRE-----  343
P22680   TFDDLEKAKTHLVVLWASQANTIPATFWSLFQMIRNPEAMKAATEEVKRTLENAGQKVSL  328
Q02318   QLSPREAMGSLPELLMAGVDTTSNTLTWALYHLSKDPEIQEALHEEVVGVVPAGQ-----  373
P05108   KMSFEDIKANVTEMLAGGVDTTSMTLQWHLYEMARNLKVQDMLRAEVLAARHQAQ-----  364
pdb3N9Y  KMSFEDIKANVTEMLAGGVDTTSMTLQWHLYEMARNLKVQDMLRAEVLAARHQAQ-----  324
P00189   KMLLEDVKANITEMLAGGVNTTSMTLQWHLYEMARSLNVQEMLREEVLNARRQAE-----  363
pdb3MZS  KMLLEDVKANITEMLAGGVNTTSMTLQWHLYEMARSLNVQEMLREEVLNARRQAE-----  325
P15539   SLPLDAIKANSMELTAGSVDTTAIPLVMTLFELARNPDVQKALRQESLAAEASIA-----  352
P15150   DMTLDTIKANTIDLTAGSVDTTAFPLLMTLFELARNPEVQQAVRQESLVAEARIS-----  352
P27786   VFSDKHILVTVGDIFGAGIETTSSVLNWILAFLVHNPEVKRKIQKEIDQYVGFSR-----  339
P05185   LLSNRHMLATIGDIFGAGVETTTSVIKWIVAYLLHHPSLKKRIQDDIDQIIGFNR-----  340
P05093   LLSDNHILTTIGDIFGAGVETTTSVVKWTLAFLLHNPQVKKKLYEEIDQNVGFSR-----  340
pdb3RUK  LLSDNHILTTIGDIFGAGVETTTSVVKWTLAFLLHNPQVKKKLYEEIDQNVGFSR-----  322
P12394   ELTDDHLLMTVGDIFGAGVETTTTVLKWAVLYLLHYPEVQKKIQEEMDQKIGLAR-----  343
P30437   GLSEDHLLMTVGDIFGAGVETTSTVLKWAIAYLIHHPQVQQRIQEELDSVVGGDR-----  347
P10633   SFNDENLRMVVVDLFTAGMVTTATTLTWALLLMILYPDVQRRVQQEIDEVIGQVR-----  346
Q01361   SFNDENLRLVVADLFSAGMVTTSTTLAWALLLMILHPDVQRRVQQEIDEVIGQVR-----  346
Q64680   SFNDENLRVVVADLFMAGMVTTSTTLTWALLFMILRPDVQCRVQQEIDEVIGQVR-----  276
P05180   YFTVESLTRTTLDLFLAGTGTTSTTLRYGLLILLKHPEIEEKMHKEIDRVVGRDR-----  336
pdb3E4E  LYTMDGITVTVADLFFAGTETTSTTLRYGLLILMKYPEIEEKLHEEIDRVIGPSR-----  316
P05181   LYTMDGITVTVADLFFAGTETTSTTLRYGLLILMKYPEIEEKLHEEIDRVIGPSR-----  337
P24470   EFHLENLAVCGSNLFTAGTETTSTTLRFGLLLLMKYPEVQAKVHEELDRVIGRHQ-----  339
P00182   EFTMDNLITTIWDVFSAGTDTTSNTLKFALLLLLKHPEITAKVQEEIEHVIGRHR-----  334
P19225   EFTMDNLIVTIGDLFGAGTETTSSTIKYGLLLLLKYPEVTAKIQEEITRVIGRHR-----  334
P00180   EFTIENLMTTLSDVFGAGTETTSTTLRYGLLLLMKHPEVIAKVQEEIERVIGRHR-----  335
P11510   NYTQEHLAILVTNLFIGGTETSSLTLRFALLLLMKYPHITDKVQEEIGQVIGRHR-----  335
P05179   EYSHENLTCSIMDLIGAGTETMSTTLRYALLLLMKYPHVTAKVQEEIDRVIGRHR-----  335
P24903   HFHMDTLLMTTHNLLFGGTKTVSTTLHHAFLALMKYPKVQARVQEEIDLVVGRAR-----  336
P11711   EFHMKNLVMTTLSLFFAGSETVSSTLRYGFLLLMKHPDVEAKVHEEIEQVIGRNR-----  337
P24461   EFNLKNLVLTTLNLFFAGTETVSSTLRYGFLLIMKHPEVQTKIYEEINQVIGPHR-----  339
P00176   EFHHENLMISLLSLFFAGTETSSTTLRYGFLLMLKYPHVAEKVQKEIDQVIGSHR-----  336
P37124   -PTYPELVTLCSEFLNGGTDTTATAIEWAIGRLIENPNIQSQLYEEIKKTVGEN------  352
P37123   -PTNAELVTLCSEFLNGGTDTTATALEWGIGRLMENPTIQNQLYQEIKTIVGDK------  339
Q04468   EINEDNVLYIVENINVAAIETTLWSIEWGIAELVNHPEIQAKLRHELDTKLGPGV-----  344
P37122   KLSEHEIKIFVLEMFLAGTETTSSSVEWALTELLRHPEAMAKVKTEISQAIEPNR-----  347
P37120   RLSITNIKALLLNLFTAGTDTSSSVIEWALTEMMKNPTIFKKAQQEMDQIIGKNR-----  342
AJV30500 KMTDQEIANLLIGVLMGGQHTSAATSAWILLHLAERPDVQQELYEEQMRVLDGGK-----  352
Q08477   KLSDEDIRAEADTFMFEGHDTTASGLSWVLYHLAKHPEYQERCRQEVQELLKDRE-----  366
P33274   ELSDEDIRAEADTFMFEGHDTTASGLSWILYNLANDPEYQERCRQEVQELLRDRD-----  366
P13584   KLSDADLRAEVDTFMFEGHDTTTSGISWFLYCMALYPEHQHRCREEVREILGDQD-----  353
pdb5T6Q  QLSDTDLRAEVDTFMFEGHDTTTSGISWFLYCMALYPEHQQRCREEVREILGDQD-----  331
P20816   SLSDEDLRAEVDTFMFEGHDTTASGISWVFYALATHPEHQERCREEVQSILGDGT-----  353
P08516   SLSDKDLRAEVDTFMFEGHDTTASGVSWIFYALATHPEHQQRCREEVQSVLGDGS-----  358
P29981   KMSDTDIREEVDTFMFEGHDTTSAGICWALFLLGSHPEIQDKVYEELDHIFQGSD-----  352
P33269   PLSNLDIREEVDTFMFEGHDTTSSALMFFFYNIATHPEAQKKCFEEIRSVVGNDK-----  354
P04800   ALSDMEITAQSIIFIFAGYEPTSSTLSFVLHSLATHPDTQKKLQEEIDRALPN-------  342
P11707   ALDDIEVVAQSIIILFAGYETTSSTLSFIMHLLATHPDVQQKLQEEIDTLLPN-------  339
Q04552   ELTDGVISAQMFIFYMAGYETSATTMTYLFYELAKNPDIQDKLIAEIDEVLSR-H-----  340
P13527   GLTFNELAAQVFVFFLGGFETSSSTMGFALYELAQNQQLQDRLREEVNEVFDQFK-----  343
P33270   GMDIGELAAQVFVFYVAGFETSSSTMSYCLYELAQNQDIQDRLRNEIQTVLEE-Q-----  347
```

Supplementary Material 7a

```
P450CPS1 ---------LLKENPKLIDAAIEEGLRYYSPV--EVTTSRWADEP---FQIHD---QTIE  316
P14762   ---------ELRNNRELAPKAVEEMLRYRFHI--SRR-DRTVKQD---NELLG---VKLK  311
P38364   ------IRFQQSQSMPYLQAVIKEALRLHPGV--GTQLTRVVPKGG--LVIEG---QFFP  400
P17549   ----DVPTHAMVKDIPYLQWVIWETMRIHSTS--AMGLPREIPAGNPPVTISG---HTFY  407
P28649   ------IKIEDIQNLKVVENFINESMRYQPVV--DL-VMRRALED---DVIDG---YPVK  389
P19098   ------VQSDDMPNLKIVENFIYESMRYQPVV--DL-IMRKALQD---DVIDG---YPVK  388
P22680   EGNPICLSQAELNDLPVLDSIIKESLRLSSAS--LN--IRTAKEDFT-LHLEDGS-YNIR  382
Q02318   -----VPQHKDFAHMPLLKAVLKETLRLYPVV--PT-NSRIIEKE---IEVDG---FLFP  419
P05108   -----GDMATMLQLVPLLKASIKETLRLHPIS--VT-LQRYLVND---LVLRD---YMIP  410
pdb3N9Y  -----GDMATMLQLVPLLKASIKETLRLHPIS--VT-LQRYLVND---LVLRD---YMIP  370
P00189   -----GDISKMLQMVPLLKASIKETLRLHPIS--VT-LQRYPESD---LVLQD---YLIP  409
pdb3MZS  -----GDISKMLQMVPLLKASIKETLRLHPIS--VT-LQRYPESD---LVLQD---YLIP  371
P15539   -----ANPQKAMSDLPLLRAALKETLRLYPVG--GF-LERILSSD---LVLQN---YHVP  398
P15150   -----ENPQRAITELPLLRAALKETLRLYPVG--IT-LEREVSSD---LVLQN---YHIP  398
P27786   -----TPSFNDRTHLLMLEATIREVLRIRPVA--PLLIPHKANID---SSIGE---FAIP  386
P05185   -----TPTISDRNRLVLLEATIREVLRIRPVA--PTLIPHKAVID---SSIGD---LTID  387
P05093   -----TPTISDRNRLLLLEATIREVLRLRPVA--PMLIPHKANVD---SSIGE---FAVD  387
pdb3RUK  -----TPTISDRNRLLLLEATIREVLRLRPVA--PMLIPHKANVD---SSIGE---FAVD  369
P12394   -----HPHLSDRPLLPYLEATISEGLRIRPVS--PLLIPHVSLAD---TSIGE---YSIP  390
P30437   -----TPQLSDRGSLPYLEATIREVLRIRPVA--PLLIPHVAQTD---TSIGK---FTVR  394
P10633   -----CPEMTDQAHMPYTNAVIHEVQRFGDIA--PLNLPRFTSCD---IEVQD---FVIP  393
Q01361   -----RPEMGDQALMPFTVAVVHEVQRFADIV--PLGLPHMTSRD---IEVQG---FHIP  393
Q64680   -----RPEMADQARMPFTNAVIHEVQRFADIL--PLGVPHKTSRD---IEVQG---FLIP  323
P05180   -----SPCMADRSQLPYTDAVIHEIQRFIDFL--PLNVPHAVIKD---TKLRD---YFIP  383
pdb3E4E  -----IPAIKDRQEMPYMDAVVHEIQRFITLV--PSNLPHEATRD---TIFRG---YLIP  363
P05181   -----IPAIKDRQEMPYMDAVVHEIQRFITLV--PSNLPHEATRD---TIFRG---YLIP  384
P24470   -----PPSMKDKMKLPYTDAVLHEIQRYITLL--PSSLPHAVVQD---TKFRD---YVIP  386
P00182   -----SPCMQDRTRMPYTDAVMHEIQRYVDLV--PTSLPHAVTQD---IEFNG---YLIP  381
P19225   -----RPCMQDRNHMPYTDAVLHEIQRYIDFV--PIPLPRKTTQD---VEFRG---YHIP  381
P00180   -----SPCMQDRSRMPYTDATVHEIQRYINLI--PNNVPRATTCN---VKFRS---YLIP  382
P11510   -----SPCMLDRIHMPYTNAMIHEVQRYIDLA--PNGLLHEVTCD---TKFRD---YFIP  382
P05179   -----SPCMQDRKHMPYTDAMIHEVQRFINFV--PTNLPHAVTCD---IKFRN---YLIP  382
P24903   -----LPALKDRAAMPYTDAVIHEVQRFADII--PMNLPHRVTRD---TAFRG---FLIP  383
P11711   -----QPQYEDHMKMPYTQAVINEIQRFSNLA--PLGIPRRIIKN---TTFRG---FFLP  384
P24461   -----IPSVDDRVKMPFTDAVIHEIQRLTDIV--PMGVPHNVIRD---THFRG---YLLP  386
P00176   -----LPTLDDRSKMPYTDAVIHEIQRFSDLV--PIGVPHRVTKD---TMFRG---YLLP  383
P37124   -----KIDEKDIEKMPYLNAVVKELLRKHPPT--YMSLTHAVTEP---AKLGG---YDIP  399
P37123   -----KVDENDIEKMPYLNAVVKELLRKHPPT--YFTLTHSVTEP---VKLAG---YDIP  386
Q04468   -----QITEPDVQNLPYLQAVVKETLRLRMAI--PLLVPHMNLHD---AKLGG---FDIP  391
P37122   -----KFEDSDIENLPYMQAVLKESLRLHPPL--PFLIPRETIQD---TKFMG---YDVP  394
P37120   -----RFIESDIPNLPYLRAICKEAFRKHPST--PLNLPRVSSDA---CTIDG---YYIP  389
AJV30500 ----KELTYDLLQEMPLLNQTIKETLRMHHPL--HS-LFRKVMKD---MHVPNTS-YVIP  401
Q08477   ---PKEIEWDDLAQLPFLTMCIKESLRLHPPV--PA-VSRCCTQD---IVLPDG--RVIP  415
P33274   ---PEEIEWDDLAQLPFLTMCIKESLRLHPPV--TV-ISRCCTQD---ILLPDG--RTIP  415
P13584   -----FFQWDDLGKMTYLTMCIKESFRLYPPV--PQ-VYRQLSKP---VTFVDG--RSLP  400
pdb5T6Q  -----SFQWEDLAKMTYLTMCMKECFRLYPPV--PQ-VYRQLSKP---VSFVDG--RSLP  378
P20816   -----SVTWDHLDQMPYTTMCIKEALRLYSPV--PS-VSRELSSP---VTFPDG--RSIP  400
P08516   -----SITWDHLDQIPYTTMCIKEALRLYPPV--PG-IVRELSTS---VTFPDG--RSLP  405
P29981   ---RS-TTMRDLADMKYLERVIKESLRLFPSV--PF-IGRVLKED---TK-IGD--YLVP  399
P33269   ---STPVSYELLNQLHYVDLCVKETLRMYPSV--PL-LGRKVLED---CE-ING--KLIP  402
P04800   ---KAPPTYDTVMEMEYLDMVLNETLRLYPIG--NR-LERVCKKD---VEING---VFMP  390
P11707   ---KELATYDTLVKMEYLDMVVNETLRLYPIA--GR-LERVCKKD---VDING---TFIP  387
Q04552   ---DGNITYECLSEMTYLSKVFDETLRKYPVA--DF-TQRNAKTD---YVFPGTD-ITIK  390
P13527   ---EDNISYDALMNIPYLDQVLNETLRKYPVGVGSA-LTRQTLND---YVVPHNPKYVLP  396
P33270   ---EGQLTYESIKAMTYLNQVISETLRLYTLV--PH-LERKALND---YVVPGHEKLVIE  398
```

Supplementary Material 7a

```
P450CPS1 KGDMVVIALASANRDETVF-ENPEVYDITRE-NN------------------------R  349
P14762   KGDVVIAWMSACNMDETMF-ENPFSVDIHRPTNK------------------------K  345
P38364   EGAEVGVNGWALYHNKAIFGNDASVFRPERWLET-KG----------------NLNI-GG  442
P17549   PGDVVSVPSYTIHRSKEIWGPDAEQFVPERWDPA-RL----------------TPRQKAA  450
P28649   KGTNIILNIGRMHR-LEYF-PKPNEFTLENFEKN--V--------------------PYRY  426
P19098   KGTNIILNIGRMHK-LEFF-PKPNEFSLENFEKN--V--------------------PSRY  425
P22680   KDDIIALYPQLMHLDPEIY-PDPLTFKYDRYLDE-NGKTK----TTFYCN---GLKLKYY  433
Q02318   KNTQFVFCHYVVSRDPTAF-SEPESFQPHRWLRN-SQ------------PATPRIQHPFG  465
P05108   AKTLVQVAIYALGREPTFF-FDPENFDPTRWLSK-DK------------NI-----TYFR  451
pdb3N9Y  AKTLVQVAIYALGREPTFF-FDPENFDPTRWLSK-DK------------NI-----TYFR  411
P00189   AKTLVQVAIYAMGRDPAFF-SSPDKFDPTRWLSK-DK------------DL-----IHFR  450
pdb3MZS  AKTLVQVAIYAMGRDPAFF-SSPDKFDPTRWLSK-DK------------DL-----IHFR  412
P15539   AGTLVLLYLYSMGRNPAVF-PRPERYMPQRWLER-KR--------------------SFQ  436
P15150   AGTLVKVLLYSLGRNPAVF-ARPESYHPQRWLDR-QG------------SG-----SRFP  439
P27786   KDTHVIINLWALHHDKNEW-DQPDRFMPERFLDP-TG------------SHLIT--PTPS  430
P05185   KGTDVVVNLWALHHSEKEW-QHPDLFMPERFLDP-TG------------TQLIS--PSLS  431
P05093   KGTEVIINLWALHHNEKEW-HQPDQFMPERFLNP-AG------------TQLIS--PSVS  431
pdb3RUK  KGTEVIINLWALHHNEKEW-HQPDQFMPERFLNP-AG------------TQLIS--PSVS  413
P12394   KGARVVINLWSVHHDEKEW-DKPEEFNPGRFLDE-QG------------QHIHS--PSPS  434
P30437   KGARIIINLWSLHHDEKEW-KNPEMFDPGRFLNE-EG------------TGLCI--PSPS  438
P10633   KGTTLIINLSSVLKDETVW-EKPHRFHPEHFLDA-QG------------NFV----KHEA  435
Q01361   KGTTLITNLSSVLKDETVW-EKPFRFHPEHFLDA-QG------------RFV----KQEA  435
Q64680   KGTTLIINLSSVLKDETVW-EKPLRFHPEHFLDA-QG------------NFV----KHEA  365
P05180   KDTMIFPLLSPILQDCKEF-PNPEKFDPGHFLNA-NG------------TFR----RSDY  425
pdb3E4E  KGTVVVPTLDSVLYDNQEF-PDPEKFKPEHFLNE-NG------------KFK----YSDY  405
P05181   KGTVVVPTLDSVLYDNQEF-PDPEKFKPEHFLNE-NG------------KFK----YSDY  426
P24470   KGTTVLPMLSSVMLDQKEF-ANPEKFDPGHFLDK-NG------------CFK----KTDY  428
P00182   KGTDIIPSLTSVLYDDKEF-PNPEKFDPGHFLDE-SG------------NFK----KSDY  423
P19225   KGTSVMACLTSALHDDKEF-PNPEKFDPGHFLDE-KG------------NFK----KSDY  423
P00180   KGTAVITSLTSMLYNDKEF-PNPDRFDPGHFLDA-SG------------KFR----KSDY  424
P11510   KGTAVLTSLTSVLHARKEF-PNPEMFDPGHFLDE-NG------------NFK----KSDY  424
P05179   KGTKVLTSLTSVLHDSKEF-PNPEMFDPGHFLDE-NG------------NFK----KSDY  424
P24903   KGTDVITLLNTVHYDPSQF-LTPQEFNPEHFLDA-NQ------------SFK----KSPA  425
P11711   KGTDVFPILGSLMTDPKFF-PSPKDFDPQNFLDD-KG------------QLK----KNAA  426
P24461   KGTDVFPLLGSVLKDPKYF-CHPDDFYPQHFLDE-QG------------RFK----KNEA  428
P00176   KNTEVYPILSSALHDPQYF-DHPDSFNPEHFLDA-NG------------ALK----KSEA  425
P37124   TGVNVEIFLPGISDDPNLW-SEPEKFDPDRFYLG-KE------------DADITGVSGVK  445
P37123   MDTNVEFFVHGISHDPNVW-SDPEKFDPDRFLSG-RE------------DADITGVKEVK  432
Q04468   AESKILVNAWWLANNPDQW-KKPEEFRPERFLEE--E------------AKVEANGNDFR  436
P37122   KDTQVLVNAWAIGRDPECW-DDPMSFKPERFLGS----------------KIDVKGQHYG  437
P37120   KNTRLSVNIWAIGRDPDVW-ENPLEFIPERFLSE-KN------------AKIEHRGNDFE  435
AJV30500 AGYHVLVSPGYTHLRDEYF-PNAHQFNIHRWNKD-SASSYSVGEEVDYGFGAISKGVSSP  459
Q08477   KGIICLISVFGTHHNPAVW-PDPEVYDPFRFDPK-NI----------------KERSPLA  457
P33274   KGIICLISIFGIHHNPSVW-PDPEVYNPFRFDPE-NI----------------KDSSPLA  457
P13584   AGSLISMHIYALHRNSAVW-PDPEVFDSLRFSTE-NA----------------SKRHPFA  442
pdb5T6Q  AGSLISLHIYALHRNSDVW-PDPEVFDPLRFSPE-NS----------------SGRHPYA  420
P20816   KGIRVTILIYGLHHNPSYW-PNPKVFDPSRFSPD-S------------------PRHSHA  440
P08516   KGIQVTLSIYGLHHNPKVW-PNPEVFDPSRFAPD-S------------------PRHSHS  445
P29981   AGCMMNLQIYHVHRNQDQY-PNPEAFNPDNFLPE-RV----------------AKRHPYA  441
P33269   AGTNIGISPLYLGRREELF-SEPNSFKPERFDVVTTA----------------EKLNPYA  445
P04800   KGSVVMIPSYALHRDPQHW-PEPEEFRPERFSKE-NK----------------GSIDPYV  432
P11707   KGTIVMMPTYALHRDPQHW-TEPDEFRPERFSKK-NK----------------DNINPYI  429
Q04552   KGQTIIVSTWGIQNDPKYY-PNPEKFDPERFNPE-NV----------------KDRHPCA  432
P13527   KGTLVFIPVLGIHYDPELY-PNPEEFDPERFSPE-MV----------------KQRDSVD  438
P33270   KGTQVIIPACAYHRDEDLY-PNPETFDPERFSPE-KV----------------AARESVE  440
```

Supplementary Material 7a

```
P450CPS1 HIAFGHGSHFCLGAPLAKLEAKIAITTLFNRMPKLQIKGDREE-IKWQG-----NYLMRS  403
P14762   HLTFGNGPHFCLGAPLARLEMKIILEAFLEAFSHIEPFEDFELEPHLTA-----SATGQS  400
P38364   SFAFGAGSRSCIGKNISILEMSKAIPQIVRNFDIEINHGDMT------WKN---ECWWFV  493
P17549   FIPFSTGPRACVGRNVAEMELLVICGTVFRLFEFEMQQEGPM------ETR---EGFLRK  501
P28649   FQPFGFGPRGCAGKYIAMVMMKVVLVTLLRRFQVKTLQKRCIENI---PKK--NDLSLHP  481
P19098   FQPFGFGPRGCVGKFIAMVMMKAILVTLLRRCRVQTMKGRGLNNI---QKN--NDLSMHP  480
P22680   YMPFGSGATICPGRLFAIHEIKQFLILMLSYFELELIEGQAKCPPLDQS-R--AGLGILP  490
Q02318   SVPFGYGVRACLGRRIAELEMQLLLARLIQKYKVVLAPETGE--LKSV-----ARIVLVP  518
P05108   NLGFGWGVRQCLGRRIAELEMTIFLINMLENFRVEIQ-HLSD--VGTT-----FNLILMP  503
pdb3N9Y  NLGFGWGVRQCLGRRIAELEMTIFLINMLENFRVEIQ-HLSD--VGTT-----FNLILMP  463
P00189   NLGFGWGVRQCVGRRIAELEMTLFLIHILENFKVEMQ-HIGD--VDTI-----FNLILTP  502
pdb3MZS  NLGFGWGVRQCVGRRIAELEMTLFLIHILENFKVEMQ-HIGD--VDTI-----FNLILTP  464
P15539   HLAFGFGVRQCLGRRLAEVEMMLLLHHILKTFQVETL-RQED--VQMA-----YRFVLMP  488
P15150   HLAFGFGVRQCLGRRVAEVEMLLLLHHVLKNFLVETL-EQED--IKMV-----YRFILMP  491
P27786   YLPFGAGPRSCIGEALARQELFIFMALLLQRFDFDVSDDKQL--P-CLVGD--PKVVF-L  484
P05185   YLPFGAGPRSCVGEMLARQELFLFMSRLLQRFNLEIPDDGKL--P-SLEGH--ASLVL-Q  485
P05093   YLPFGAGPRSCIGEILARQELFLIMAWLLQRFDLEVPDDGQL--P-SLEGI--PKVVF-L  485
pdb3RUK  YLPFGAGPRSCIGEILARQELFLIMAWLLQRFDLEVPDDGQL--P-SLEGI--PKVVF-L  467
P12394   YLPFGAGIRVCLGEVLAKMELFLFLAWVLQRFTLECPQDQPL--P-SLEGK--FGVVL-Q  488
P30437   YLPFGAGVRVCLGEALAKMEIFLFLSWILQRLTMTVSPGQPL--P-SLEGK--FGVVL-Q  492
P10633   FMPFSAGRRACLGEPLARMELFLFFTCLLQRFSFSVPVGQPR--P-STHGF--FAFPV-A  489
Q01361   FIPFSAGRRACLGEPLARMELFLFFTSLLQHFSFSVPAGQPR--P-SEHGV--FAFLV-T  489
Q64680   FMPFSAGRRACLGEPLARMELFLFFTCLLQRFSFSVPAGQPR--P-SNYGV--FGALT-T  419
P05180   FMPFSAGKRICAGEGLARMEIFLFLTSILQNFSLKPVKDRKD--IDISPII--TSLAN-M  480
pdb3E4E  FKPFSTGKRVCAGEGLARMELFLLLCAILQHFNLKPLVDPKD--IDLSPIH--IGFGC-I  460
P05181   FKPFSTGKRVCAGEGLARMELFLLLCAILQHFNLKPLVDPKD--IDLSPIH--IGFGC-I  481
P24470   FVPFSLGKRACVGESLARMELFLFFTTLLQKFSLKTLVEPKD--LDIKPIT--TGIIN-L  483
P00182   FMPFSAGKRACVGEGLARMELFLLLTTILQHFTLKPLVDPKD--IDPTPVE--NGFVS-V  478
P19225   FMAFSAGRRACIGEGLARMEMFLILTSILQHFTLKPLVNPED--IDTTPVQ--PGLLS-V  478
P00180   FMPFSTGKRVCVGEVLARMELFLFLTAILQNFTPKPLVDPKD--IDTTPLV--SGLGR-V  479
P11510   FMPFSAGKRKCVGEGLASMELFLFLTTILQNFKLKSLSDPKD--IDINSIR--SEFSS-I  479
P05179   FLPFSAGKRACVGEGLARMQLFLFLTTILQNFNLKSLVHPKD--IDTMPVL--NGFAS-L  479
P24903   FMPFSAGRRLCLGESLARMELFLYLTAILQSFSLQPLGAPED--IDLTPLS--SGLGN-L  480
P11711   FLPFSTGKRFCLGDGLAKMELFLLLTTILQNFRFKFPMKLED--INESPKP--LGFTR-I  481
P24461   FVPFSSGKRICLGEAMARMELFLYFTSILQNFSLHPLVPPVN--IDITPKI--SGFGN-I  483
P00176   FMPFSTGKRICLGEGIARNELFLFFTTILQNFSVSSHLAPKD--IDLTPKE--SGIGK-I  480
P37124   MIPFGMGRRICPGLNMATVHVSLMLARLVQEFEWA-DPENTR--VDFTEK---LEFTVVM  499
P37123   MMPFGVGRRICPGLGMATVHVNLMLARMVQEFEWFAYPGNNK--VDFSEK---LEFTVVM  487
Q04468   YLPFGVGRRSCPGIILALPILGITIGRLVQNFELLPPPGQSK--IDTDEKGGQFSLHILK  494
P37122   LIPFGAGRRMCVGLPLGHRMMHFALGSLLREFEWELPDGVSPKSINMDGS---MGVTARK  494
P37120   LIPFGAGRRICAGTRMGIVMVEYILGTLIHSFDWKLPNDVVD--INMEET---FGLALQK  490
AJV30500 YLPFGGGRHRCIGEHFAYCQLGVLMSIFIRTLKWHYPEGKTVPPPDF-T-----SMVTLP  513
Q08477   FIPFSAGPRNCIGQAFAMAEMKVVLGLTLLRFRV--LPDHT-EPRRK-P-----ELVLRA  508
P33274   FIPFSAGPRNCIGQTFAMSEMKVALALTLLRFRL--LPDDK-EPRRQ-P-----ELILRA  508
P13584   FMPFSAGPRNCIGQQFAMSEMKVVTAMCLLRFEF--SLDPSRLPIKM-P-----QLVLRS  494
pdb5T6Q  FIPFSAGPRNCIGQQFAMNEMKVVTALCLLRFEF--SVDPLRLPIKL-P-----QLVLRS  472
P20816   YLPFSGGARNCIGKQFAMNELKVAVALTLLRFEL--LPDPTRIPVPM-P-----RLVLKS  492
P08516   FLPFSGGARNCIGKQFAMSEMKVIVALTLLRFEL--LPDPTKVPIPL-P-----RLVLKS  497
P29981   YVPFSAGPRNCIGQKFATLEEKTVLSSILRNFKVRSIEKRED-LTLM-N-----ELILRP  494
P33269   YIPFSAGPRNCIGQKFAMLEIKAIVANVLRHYEVDFVGDSSEPPVLI-A-----ELILRT  499
P04800   YLPFGNGPRNCIGMRFALMNMKLALTKVLQNFSFQPCKETQ-IPLKLSR----QGLLQ-P  486
P11707   YHPFGAGPRNCLGMRFALMNIKLALVRLMQNFSFKLCKETQ-VPLKLGK----QGLLQ-P  483
Q04552   YLPFSAGPRNCLGMRFAKWQSEVCIMKVLSKYRVEPSMKSS-GPFKFDP----MRLFALP  487
P13527   WLGFGDGPRNCIGMRFGKMQSRLGLALVIRHFRFTVCSRTD-IPMQINP----ESLAWTP  493
P33270   WLPFGDGPRNCIGMRFGQMQARIGLAQIISRFRVSVCDTTE-IPLKYSP----MSIVLGT  495
```

Supplementary Material 7a

```
P450CPS1 L---EELPLTF--------------------- 411
P14762   L---TYLPMTVYR------------------- 410
P38364   K---PEYKAMIKPRAA---------------- 506
P17549   P---LGLQVGMKRRQPGSA------------- 517
P28649   NEDRHLVEIIFSPRNSDKYLQQ---------- 503
P19098   IERQPLLEMVFTQEAQTRIRVTKVDQH----- 507
P22680   P--LNDIEFKYKFKHL---------------- 504
Q02318   N---KKVGLQFLQRQC---------------- 531
P05108   E---KPISFTFWPFNQE----ATQQ------- 521
Pdb3N9Y  E---KPISFTFWPFNQE----ATQQHHHHHH- 487
P00189   D---KPIFLVFRPFNQD----PPQA------- 520
Pdb3MZS  D---KPIFLVFRPFNQD----PPQAHHHH--- 486
P15539   S---SSPVLTFRPVS----------------- 500
P15150   S---TLPLFTFRAIQ----------------- 503
P27786   I---DPFKVKITVRQAWKDAQVEVST------ 507
P05185   I---KPFKVKIEVRQAWKEAQAEGSTP----- 509
P05093   I---DSFKVKIKVRQAWREAQAEGST------ 508
pdb3RUK  I---DSFKVKIKVRQAWREAQAEGSTHHHH-- 494
P12394   V---QKFRVKARLREAWRGEMVR--------- 508
P30437   P---VKYKVNATPRAGWEKSHLQTS------- 514
P10633   P---LPYQLCAVVREQGL-------------- 504
Q01361   P---APYQLCAVPR------------------ 500
Q64680   P---RPYQLCASPR------------------ 430
P05180   P---RPYEVSFIPR------------------ 491
pdb3E4E  P---PRYKLCVIPRSHHHH------------- 476
P05181   P---PRYKLCVIPRS----------------- 493
P24470   P---PPYKLCLVPR------------------ 494
P00182   P---PSYELCFVPV------------------ 489
P19225   P---PPFELCFIPV------------------ 489
P00180   P---PLYQLSFIPA------------------ 490
P11510   P---PTFQLCFIPV------------------ 490
P05179   P---PTYQLCFIPS------------------ 490
P24903   P---RPFQLCLRPR------------------ 491
P11711   I---PKYTMSFMPI------------------ 492
P24461   P---PTYELCLIAR------------------ 494
P00176   P---PTYQICFSAR------------------ 491
P37124   K---NTLRAKIKPRM----------------- 511
P37123   K---NPLRAKVKLRI----------------- 499
Q04468   H---S---TIVAKPRSF--------------- 505
P37122   R---DSLKVIPKKA------------------ 505
P37120   A---VPLEAIVTPRLSFDIYQS--SEPF---- 513
AJV30500 T---GPAKIIWEKRNPE----QKI-------- 530
Q08477   E---GGLWLRVEPLS----------------- 520
P33274   E---GGLWLRVEPLTAG----AQ--------- 524
P13584   K---NGFHLHLKPLGPG----SGK-------- 511
pdb5T6Q  K---NGIHLYLKPLGPK----AEKSTHHHHHH 497
P20816   K---NGIHLRLKKLR----------------- 504
P08516   K---NGIYLYLKKLH----------------- 509
P29981   E---SGIKVELIPRLPA----DAC-------- 511
P33269   K---EPLMFKVRERVY---------------- 512
P04800   T---KPIILKVVPRDEI----ITGS------- 504
P11707   E---KPIVLKVVSRDGI----IRGA------- 501
Q04552   K---GGIYVNLVRR------------------ 498
P13527   K---NNLYLNVQAIRKK----IK--------- 509
P33270   V---GGIYLRVERI------------------ 506
```

Figure S7(a): Multiple sequences alignments of P450CPS1 from *Bacillus cereus* strain AOA-CPS1 and members of the conserved protein domain family p450 i.e., haem-thiolate proteins in

pfam00067 superfamily. *Bacillus cereus* strain AOA-CPS1 P450CPS1 is shaded in yellow. The multiple sequences alignments was constructed using Cluster Omega (1.2.4) multiple sequence alignment (Madeira et al., 2019). Conserved residues were indicated in asterisk (*) and shaded in green. Members of p450 (pfam00067)superfamily include Benzoate 4-monooxygenase (P17549), Erg11p from *Saccharomyces cerevisiae* YJM1307 (AJV30500), cholesterol side-chain cleavage enzyme (pdb: 3N9Y), Steroid 17-alpha-hydroxylase/17, 20 lyase (pdb:3RUK) and cytochrome P450 strains 2e1 (pdb: 3E4E), P-450C27/25 (Q02318), 11B2 (P15539), 11B1 (P15150), scc (P05108), 11A1 (P00189), 19A1 (P28649), P-450AROM (P19098), P450-LTB-omega (Q08477), 2D14 (Q01361), P450-CMF1A (P10633), CYPXVII (P12394), P450-C17 (P30437), P450c17 (P05185), P450c17 (P05093), 17A1 (P27786), BM-1 (P14762), 7A1 (P22680), 57A2 (P38364), 6A1 (P13527), P450 6a2 (P33270), 6B1 (Q04552), 3A1 (P04800), 3A6 (P11707), 4C1 (P29981), 4A2 (P20816), 4A10 (P08516), 4B1 (P13584), 4F1 (P33274), Cyp11a1 (pdb:3MZS), 4B1 (pdb:5T6Q), 4d1 (P33269), 77A2 (P37124), 77A1 (P37123), 73 (Q04468), 76A2 (P37122), 75A2 (P37120), 2C12 (P11510), 2C7 (P05179), 2C1 (P00180), 2C3 (P00182), 2C70 (P19225), 2C23 (P24470), 2E1 (P05181), 2H1 (P05180), 2A1 (P11711), 2G1 (P24461), 2B1 (P00176), 2F1 (P24903), 2D4 (Q64680), and P450CPS1 from *Bacillus cereus* AOA-CPS1.

Supplementary Material 7b

```
P540CPS1       ---------------------------------------------MSM---------KNKV   7
NP001056579    MVMEGMGMAAAWAAGDLWVLAAAVVAGVVLVDAVVRRAHDWVRVAALGAERRSRLPPGEM   60
XP002436354    -MLGEEAALLSWAAALLL-----GAAALVLVDAAARWAHGWYREAPLGAARRARLPPGEM   54
NP180803       ----------MTETGLILMWFPLIILGLFVLKWVLKRVNVWIYVSKLGEK-KHYLPPGDL   49
XP002879389    ----------MTEPGLILMWFPLMVLGLFVLKWVLKSVNVWIYESKLGEK-KHYLPPGDL   49
NP172008       ---------MAETTSWIPVWFPLMVLGCFGLNWLVRKVNVWLYESSLGEN-RHYLPPGDL   50
XP002892270    ---------MAETTSWIPVWFPVMVLGCFVLNWLVKMVNVWLYESSLGEN-RQYLPPGDL   50
XP002264215    -----------MGLASSWVLYTAIFAGALALRWVLLRVNKWVYEGRLKGK-SYHLPPGDL   48
XP002265630    -----------MELGMIWVAFGAILGGVLGVKWVLRRANSWVYEVKLGEK-RYSLPPGDL   48
XP002526524    -----------MEMGFVWVVLIWISGGFWCLKWILKRVNCWLYENQLGEM-QYSLPPGDL   48
XP002321248    -----------MESGSIWVVLAVIFGGLGVGKWILKKVNWWLYEAQLGEK-QYSLPPGDL   48
XP002319447    -----------MGLGSIWVVLVVIFCGLGVGQWILKRVNWWLYEAKLGAK-KDSLPPGDL   48
XP002966472    ---------------------MAVA-LIIVLISCILNFNSWFYAPKLRPG-SPPLPPGSL   37
XP002982048    --------------MDLWLPSIAVA-LIIVLISCILNFNSWFYAPKLRPG-SPPLPPGSL   44
XP002988455    -----------------MILGVAVV-AVTALLTFSSFFNKWYYEPVLKPG-QPPLPPGSL   41
XP002975682    -------------MNLKWAIAIATV-AAATFFELLRNFNRFWYEPKLKPG-QAPLPPGSL   45
XP002993510    -----------------------------------LRNFNRFWYEPKLKPG-QAPLPPGSL   25
                                                            :          ..:

P540CPS1       GLRIEDGINLASAQF--------------KEDAYEIYKESRKMQPILFVNKTELGAEWLI   53
NP001056579    GWPMVGSMWAFLRAFKSGNPDAFIASFIRRFGRTGVYRTFMFSSPTILAVTPEACKQ---   117
XP002436354    GWPVVGGMWAFLRAFKSGKPDAFIASFVRRFGRTGVYRGFMFSSPTVLVTTPEACKQ---   111
NP180803       GWPVIGNMWSFLRAFKTSDPESFIQSYITRYGRTGIYKAHMFGYPCVLVTTPETCRR---   106
XP002879389    GWPIIGNMWSFLRAFKTSDPESFIQSYITRYGRTGIYKAHMFGYPCVLVTTPETCRR---   106
NP172008       GWPFIGNMLSFLRAFKTSDPDSFTRTLIKRYGPKGIYKAHMFGNPSIIVTTSDTCRR---   107
XP002892270    GWPFIGNMLSFLRSFKTSDPDSFTSTLIKRYGPKGIYKAHMFGNPSIIVTTPDTCRR---   107
XP002264215    GWPLIGNMWTFLRAFKTKNPDSFISNIVERYGKGGIYKTFMFGNPSILVTSPEGCRK---   105
XP002265630    GWPLIGNMWSFLRAFKSTDPDSFISSFITRFGQTGMYKVLMFGNPSIIVTIPEACKR---   105
XP002526524    GWPFIGNMWSFLRAFKSNDPDSFMRNFTARYGSGGIYKAFMFGKPSVLVTTSEACKR---   105
XP002321248    GWPFIGNMWSFLRAFKSSDPDSFMRTFINKYGDNGIYKAFMFGNPSVFVTTPEACRR---   105
XP002319447    GWPFIGNMWSFLTAFKSSDPDSFIRSFVNRYGHTGIYKAFMFGNPSVLVTTPEGCRR---   105
XP002966472    GWPVFGNMGDFLQAFKSSNPESFVAGFISKYGCGGLYKAFLFRQPTILATSAEVCKT---   94
XP002982048    GWPVFGNMGDFLQAFKSSNPESFVGGFISKYGCGGLYKAFLFRQPTILATSAEVCKT---   101
XP002988455    GWPVFGNMAAFLRAFKSGRPDTFMAHYVAKYNRVGFYKAFLFWQPTVLAATPEACKF---   98
XP002975682    GWPIFGNMASFLRAFKSHNPDSFITNYLHKYDRTGVYKAFLFWQPTVLATTPETCKV---   102
XP002993510    GWPIFGNMASFLRAFKSHNPDSFITKYLHKYDRTGVYKAFLFWQPTVLATTPETCKV---   82
               *  . ..:      *              : .   .*:      * ::.   :

P540CPS1       TRYEDALPLLKDNR-LKKD-PANVF-SQDTLNVFLTVDNSDYLTTHMLNSDPPNHNRLRS   110
NP001056579    --------VLMDDEGFVTGWPKATV-TLIGPKSFV-------------NMSYDDHRRIRK   155
XP002436354    --------VLMDDDAFVTGWPKATV-ALIGPKSFV-------------AMPYDEHRRLRK   149
NP180803       --------VLTDDDAFHIGWPKSTM-KLIGRKSFV-------------GISFEEHKRLRR   144
XP002879389    --------VLTDDDAFHIGWPKSTM-KLIGRKSFV-------------GISFEEHKRLRR   144
NP172008       --------VLTDDDAFKPGWPTSTM-ELIGRKSFV-------------GISFEEHKRLRR   145
XP002892270    --------VLTDDDAFKPGWPTSTM-ELIGRKSFI-------------GISFEEHKRLRR   145
XP002264215    --------VLTDDDNFKPGWPTSTE-ELIGKKSFV-------------SISYEEHKRLRR   143
XP002265630    --------VLTDDQNFKPGWPTSTM-ELIGRKSFI-------------GITNEEHKRLRR   143
XP002526524    --------VLTDDDAFKPGWPSSTT-ELIGKKSFI-------------GISYEEHKRLRR   143
XP002321248    --------VLSNDDAFKPGWPISTL-KLIGRKSFI-------------DISYEEHKRLRR   143
XP002319447    --------LLTDDNAFKPGWPLATL-KLIGEKSFI-------------DIPYEEHKRLRR   143
XP002966472    --------VLCNHDVFEIGWPERVVKDLLGLKVLS-------------AVTGDDHLKLSK   133
XP002982048    --------VLCNHDVFEIGWPERVVKELLGLKVLS-------------AVTGDDHLKLSK   140
XP002988455    --------VLSKDS-FETGWPESAV-ELMGRNSFA-------------GLTGESHFKLRK   135
XP002975682    --------VLSRDSLFETGWPSSTR-RLIGTRSFA-------------GVTGEEHLKLRR   140
XP002993510    --------VLSRDSLFETGWPSSTR-RLIGTRSFA-------------GVTGEEHLKLRR   120
                          :*  .  :  . *  .       . :                   .* ::
```

Figure S7(b): Multiple sequences alignments of p450AOA-CPS1 from *B. cereus* strain AOA-CPS1 and other members of the conserved protein domain PLN02302 superfamily.

Supplementary Material 7b

```
P540CPS1       LVQKAFTPKMIAQLEGRIQHIADDLLNEVERK-----------GSLNLVD---DYSFPLP  156
NP001056579    LTAAPINGF--DALTTYLSFIDQTVVASLRRWSSPES------GQVEFLTELRRMTFKII  207
XP002436354    LTAAPINGF--DALTAYLPFIDRTVTSSLRAWADESSSGAGAGTGVEFLTELRRMTFKII  207
NP180803       LTSAPVNGP--EALSVYIQFIEETVNTDLEKWSKM--------GEIEFLSHLRKLTFKVI  194
XP002879389    LTSAPVNGP--EALSVYIQFIEETVITDLEKWSKM--------GEIEFLSHLRKLTFKVI  194
NP172008       LTAAPVNGH--EALSTYIPYIEENVITVLDKWTKM--------GEFEFLTHLRKLTFRII  195
XP002892270    LTAAPVNGH--EALSTYIPYIEENVITVLDKWTKM--------GEFEFLTHLRKLTFRII  195
XP002264215    LTSAPVNGH--EALSLYIPYIEKNVISDLEKWSKM--------GNIEFLTGVRKLTFKII  193
XP002265630    LTATPVNGH--EALSIYMQYIEDNVISALNKWAAM--------GEFEFLTALRKLTFKII  193
XP002526524    LTASPVNGH--EALSVYMHYIEDKVKSALEKWSTM--------GEIQFLTQLRKLTFRII  193
XP002321248    LTSAPVNGH--EALSVYIPYIEENVIAMLEKWTTM--------GKIEFLTQVRKLTFKII  193
XP002319447    LTSASVNGH--EALSTYIPYIEQNVIAELEKWTTM--------GQIEFLTKMRKLTFRII  193
XP002966472    LVKPALSSP--KAIQHQMPCIEENVKKLLDEWADR--------GNIVFLDEARMFTLKTI  183
XP002982048    LVKPALSSP--KAIQHQMPCIEENVKKLLDEWADR--------GNIVFLDEARMFTLKTI  190
XP002988455    LTEPAVNSP--KALEQYVPLIVNNIKACLARWSAQ--------DKIVLLTEMRRFTFLTV  185
XP002975682    LTEPALSNP--KALEDYIPRMSSNIKSCLEEWSCQ--------ERTLLLKEMRKYAFRTI  190
XP002993510    LTEPALSNP--KALEDYIPRMSSNIKSCLEEWSCQ--------ERTLLLREMRKYAFRTI  170
               *.   ..      :   :  :   :   :                  ::      ::

P540CPS1       I-IVISEMLGIPKE-DQAKFRIWSHAV----IAYPETPEEIKETEKQLSEFITYLQYLVD  210
NP001056579    VQIFMSGADDATMEALERSYTDLNYGMRAMAINLPGFA--YYRALRARRKLVSVLQGVLD  265
XP002436354    VQIFLGGADEPTTRALERSYTDLNYGMRAMAINLPGFA--YHRALRARRRLVAVLQGVLD  265
NP180803       MYIFLSSESEHVMDSLEREYTNLNYGVRAMGINLPGFA--YHRALKARKKLVAAFQSIVT  252
XP002879389    MYIFLSSESEHVMDALEREYTNLNYGVRAMGINLPGFA--YHRALKARKKLVAAFQSIVT  252
NP172008       MYIFLSSESENVMDALEREYTALNYGVRAMAVNIPGFA--YHRALKARKTLVAAFQSIVT  253
XP002892270    MYIFLSSESENVMDALEREYTALNYGVRAMAVNIPGFA--YHRALKARKTLVAAFQSIVT  253
XP002264215    MYIFLSAESGDVMEALEKEYTILNYGVRALAINIPGFA--FHKAFKARKNLVATLQATVD  251
XP002265630    MYIFLSSESEHVMEALEREYTSLNYGVRSMAINLPGFA--YHKALKARKNLVNIFQSIVN  251
XP002526524    MYIFLSSESHSVMEALEREYTTLNYGVRAMAINLPGFA--YYKALKARKNLVAVLQFVVD  251
XP002321248    MYIFLSSESEVVMEALEKDYTTLNYGVRAMAINLPGFA--YYKALKARKRLVAIFQSIVD  251
XP002319447    IYIFLSKTSERVMEALEKEYTTLNYGIRAMAINLPGFA--YYEALKARKKLVAIFQSIVD  251
XP002966472    HEILVGEDTGIDFKQVSGLFHTMNKGLRALPLKFPGTA--YSNAVKARATLANDFWRIFY  241
XP002982048    HEILVGEDTGIDFKQVSGLFHTMNKGLRALPLNFPGTA--YSNAVKARATLANDFWRIFY  248
XP002988455    LHILYGKDSSLDVDETFSLYYIVNQGIRALPINFPGTA--YNKALKARRKLIKLIQDVIN  243
XP002975682    HDILFSKDSGLDVEEVSSIYYEVNQGIRSLPINLPGTS--YNRALKARKKLDVLLHRVLN  248
XP002993510    HDILFSKDSGLDVEEVSSLYYEGNQGIRSLPINLPGTS--YNRALKARKKLDVLLHRVLN  228
                 *. .            :   . .:    :  *       .: :    :   :   .

P540CPS1       MKRKE---------PKEDLVSALIL-AESEGHKLSARELYSMIMLLIVAGHETTVNLITN  260
NP001056579    GRRAAAAKGFK-RSGAMDMMDRLIEAEDERGRRLADDEIVDVLIMYLNAGHESSGHITMW  324
XP002436354    ERRAAKAKGVS-GA-GVDMMDRLIEAEDERGRRLDDDEIIDVLIMYLNAGHESSGHITMW  323
NP180803       NRRNQRKQNI--SSNRKDMLDNLIDVKDENGRVLDDEEIIDLLLMYLNAGHESSGHLTMW  310
XP002879389    NRRNQRKQNI--SSNRKDMLDNLIDVKDENGRVLDDEEIIDLLLMYLNAGHESSGHLTMW  310
NP172008       ERRNQRKQNI--LSNKKDMLDNLLNVKDEDGKTLDDEEIIDVLLMYLNAGHESSGHTIMW  311
XP002892270    ERRNQREQNI--LPNKKDMLDNLLNVKDEDGRTLDDEEIIDVLLMYLNAGHESSGHTIMW  311
XP002264215    ERRQRERENS--SAREKDMLDALLHVEDENGRKLTDEEIIDLLIMYLNAGHESSGHVTMW  309
XP002265630    ERRDRKKGNS--QTMKKDMMDALLDIEDENGRKLSDEEIIDILVMYLNAGHESSAHVTMW  309
XP002526524    ARRNQKKGAEPNSSKKKDMMDALLDVEDEKGRKLSDEEIVDVLLMYLNAGHESSGHITMW  311
XP002321248    ERRNLRKNSAR-NAKKKDMMDSLLGVEDENGRKLTDEEIIDVILMYLNAGHESSGHITTW  310
XP002319447    GRRNLKKDDVT-NTKKKDMMDSLLDVEDENGRKLTDEEVIDIMLMYLNAGHESSGHITTW  310
XP002966472    ERKKSR-K------RGGDTLSMLLDATDEGGQPLEDDQIVDLIMSFMNAGHESTAHLVTW  294
XP002982048    ERKESK-K------RGGDTLSMLLDATDEGGQPLEDDQIVDLIMSFMNGGHESTAHLVTW  301
XP002988455    QRRASGKP------QETNILSLLMDQLDDKGEALEDAQIIDVLNMYMNAGHDSTAHVIMW  297
XP002975682    KRRFSEKP------EKTDTLSLLMDATDENGKHLDDKQIVDLLVMYLNAGHDSTAHLILW  302
XP002993510    KRRFSEKP------EKTDTLSLLMDATDENGKHLDDKQIVDLLVMYLNAGHDSTAHLILW  282
                ::              : :. *:   :. *. *   :: .::   : .**::: :
```

Supplementary Material 7b

```
P540CPS1       TVLALLENPNQLQLLKENPK-LIDAAI-----------EEGLRYYSPVEVTTSRW-----  303
NP001056579    ATVFLQENPDIFARAKAEQE-EIMRSIPATQNGLTLRDFKKMHFLSQVVDETLRCVNISF  383
XP002436354    ATVFLQENPDIFAKAKAEQE-AIMRSIPASQQGLTLRDFRKMEYLSQVIDETLRLVNISF  382
NP180803       ATILMQEHPMILQKAKEEQE-RIVKKRAPGQ-KLTLKETREMVYLSQVIDETLRVITFSL  368
XP002879389    ATILMQEHPEILQKAKEEQE-RIVKNRALGQ-KLTLKETREMEYLSQVIDETLRVITFSL  368
NP172008       ATVFLQEHPEVLQRAKA-EQEMILKSRPEGQKGLSLKETRKMEFLSQVVDETLRVITFSL  370
XP002892270    ATIFLQEHPEFLQRAKVNEQEMILKNRPEGQKGLTLKETRQMEFLSQVVDETLRVITFSL  371
XP002264215    ATLLLQGHPEIFQRAKAEQE-EIVKNRPPTQKGLTLREVRKMEYLSQVIDETLRWLTFSL  368
XP002265630    ATVKLQENPEFFQRAKAEQE-EIIRKRPPNQKRLTLKEIREMEYLPKVIDETLRWITFSF  368
XP002526524    ATVFLQEHPEFLQKARQEQE-EIIKRRPPTQKGLTLKEVRDMEYLSKVIDETLRLITFSL  370
XP002321248    ATIFLQEHPEFLQKAKEEQE-QIVKRRPPAQNGLSLKEVREMDYLSKVIDETLRLITFSL  369
XP002319447    ATIFLQDHPEYFQKAKEEQE-QIIKRRPLTQKRLSLKEVREMKYLSKVIDETLRMVTFSL  369
XP002966472    LAILLKEHPAVYQRLKAEQD-EIALKKLPGE-SLTLADIRSMTYMSRVIDETLRLINISP  352
XP002982048    LAILLKEHPAVYQRLKAEQD-EIALKKMPGE-SLTLADMRSMTYMSRVIDETLRLINISP  359
XP002988455    LMIFLKRNPDVLEKVKTEQD-GIAKCISEGE-MLNLSDIKRMRYLSSVVDETLRLANISP  355
XP002975682    LLIFLLKHEIVYDKVKEEQE-LIASQRPVGE-SLSLSDVKKMSYLSRVINETLRVANISP  360
XP002993510    LLIFLLKHEIVYDKVKEEQE-LIASQKPLGD-SLSLSDVKKMSYLSRVINETLRVANISP  340
                 : :  :       :   .  *                . : :   *   * *

P540CPS1       -----ADEPFQIHDQTIEKGDMVVIALASANRDETVFENPEVYDITREN-------NRHI  351
NP001056579    VSFRQATRDIFVNGYLIPKGWKVQLWYRSVHMDDQVYPDPKMFNPSRWE-GPPPKAGTFL  442
XP002436354    VSFRQATKDVFVNGYLIPKGWKVQLWYRSVHMDPQVYPDPTKFNPSRWE-GHSPRAGTFL  441
NP180803       TAFREAKSDVQMDGYIIPKGWKVLTWFRNVHLDPEIYPDPKKFDPSRWE-GYTPKAGTFL  427
XP002879389    TAFREAKSDVQIDGYIIPKGWKVLTWFRNVHLDPEIYPDPKKFDPSRWE-GYTPKAGTFL  427
NP172008       TAFREAKTDVEMNGYLIPKGWKVLTWFRDVHIDPEVFPDPRKFDPARWDNGFVPKAGAFL  430
XP002892270    TAFREAKTDVEMNGYLIPKGWKVLTWFRDVHIDPEVYPDPRKFDPSRWDNGFVPKAGAFL  431
XP002264215    MVFREAKADVNIGGYLFPKGWKVLVWFRAVHYDPEIYPNPEVFNPSRWD-NFTPKAGTFL  427
XP002265630    VVFREAKADINICGYTIPKGWKVLVWFRSLHFDPETYPDPKEFNPCRWD-DYTAKPGTFL  427
XP002526524    VVFREAKTNVNISGYVIPKGWKILVWFRSVHLDPEIYPNPREFNPSRWD-NHTAKAGTFL  429
XP002321248    TVFREAKTDFSINGYIIPKGWKVLVWFRTVHLDPEIYQNPKEFNPSRWD-NYTPKAGTFL  428
XP002319447    TVFREAKTDFCMNGYTIPKGWKVLAWFRTIHLDPEVYPNPKEFNPSRWD-DYTPKAGTFL  428
XP002966472    FVFRKVLSDVQLNGYTIPRGWFVEAWLRQVHMDPLVHKNPREFDPDRWI-NEKPQPHTYV  411
XP002982048    FVFRKVLSDVQLNGYTIPRGWFVEAWLRQVHMDPLVHKNPREFDPDRWI-NEKPQPHTYV  418
XP002988455    MVFRRALVDVEFNGFTIPKGWHAEAWLRQVHMDPHVHPDPEKFDPERWE-KYGASPFTFM  414
XP002975682    MVFRRAVTDVEVNGFTIPKGWYVEPWLRQVHMDPAVHSNPQNFDPDRWA-RNEVRPFTHL  419
XP002993510    MVFRRAVTDVEVNGFTIPKGWYVEPWLRQVHMDPAVHSNPQNFDPDRWA-RNEVRPFTHL  399
                    .   . . .  : :*          : *   . :*  ::  *           .:

P540CPS1       AFGHGSHFCLGAPLAKLEAKIAITTLFNRMPKLQIKGDREEIKWQGNYLM--RSLEELPL  409
NP001056579    PFGLGARLCPGNDLAKLEISVFLHHFLLGYKLKR---ANPKCRV--RYLPHPRPVDNCLA  497
XP002436354    AFGLGARLCPGNDLAKLEISVFLHHFLLGYRLAR---TNPRCRV--RYLPHPRPVDNCLA  496
NP180803       PFGLGSHLCPGNDLAKLEISIFLHHFLLKYRVER---SNPGCPV--MFLPHNRPKDNCLA  482
XP002879389    PFGLGSHLCPGNDLAKLEISIFLHHFLLRYRVER---SNPGCPV--MFLPHNRPKDNCLA  482
NP172008       PFGAGSHLCPGNDLAKLEISIFLHHFLLKYQVKR---SNPECPV--MYLPHTRPTDNCLA  485
XP002892270    PFGAGSHLCPGNDLAKLEISIFLHHFLLKYQVKR---SNPKCPV--MYLPHTRPTDNCLA  486
XP002264215    PFGAGSRLCPGNDLAKLEISIFLHYFLLNYRLER---VNPGCEL--MYLPHPRPVDNCLA  482
XP002265630    PFGLGSRLCPGNDLAKLEISVFLHHFLLNYQLER---LNPGCPR--MYLPHSRPRDNCLA  482
XP002526524    PFGAGSRMCPGNDLAKLEIAIFLHHFLLNYELER---LNPGSSM--MYLPHSRPKDNCLA  484
XP002321248    PFGAGSRLCPGNDLAKLEISIFLHYFLLDYRLER---ENPECRW--MFLPHTRPTDNCVA  483
XP002319447    PFGAGSRLCPGNNLAKLEISIFLHYFLLDYR------QNPECSW--RFLPHTRPIDNCLA  480
XP002966472    AFGLGNRKCPGSNLSKIQSSIIIHHLITKYNWEP---LNPHYKL--VYLPHPRPADHYPV  466
XP002982048    AFGLGNRKCPGSNLSKIQSSIIIHHLITKYNWEP---LNPHYKL--VYLPHPRPADHYPV  473
XP002988455    PFGMGNRTCPGNELAKLQIFIVVHYFVTGYRWTA---LNPNSKV--SYLPHPRPRDFYSV  469
XP002975682    PFGLGSRTCPGNELAKLEACIIVHHLVLGYDMKP---LNPDCEV--TFLPHPRPKDYFPV  474
XP002993510    PFGLGSRTCPGNELAKLEACIIVHHLVLGYDVKP---LNPDCEV--TFLPHPRPKDYFPV  454
                ** * : * *  *:*::  : :  :.           .        :*   *  :
```

Fig. S7(b): Multiple sequences alignments of P450AOA-CPS1 from *B. cereus* strain AOA-CPS1 and other members of the conserved protein domain PLN02302 superfamily.

```
P540CPS1       TF-------- 411
NP001056579    TITKVSDEH- 506
XP002436354    KITRVSEEY- 505
NP180803       RITRTMP--- 489
XP002879389    RITKTTP--- 489
NP172008       RISYQ----- 490
XP002892270    RISYQ----- 491
XP002264215    RVRKVA---- 488
XP002265630    IVRKVAAESE 492
XP002526524    RIKKIPSA-- 492
XP002321248    RIKKVSSTSV 493
XP002319447    RIKKVSSESV 490
XP002966472    KITKRALV-- 474
XP002982048    KITKRALV-- 481
XP002988455    RVSKLL---- 475
XP002975682    QVRRRR---- 480
XP002993510    QVRR------ 458
```

Figure S7(b): Multiple sequences alignments of P450CPS1 from *B. cereus* strain AOA-CPS1 and all other members of the conserved protein domain family PLN02302 superfamily (accession: cl12078) i.e., ent-kaurenoic acid oxidase *Bacillus cereus* strain AOA-CPS1 p450MO is shaded in green. The multiple sequences alignments was constructed using Cluster Omega (1.2.4) multiple sequence alignment (Madeira et al., 2019). Conserved residues were indicated in asterisk (*) and shaded in green. Members of PLN02302 superfamily (cl12078) include Os06g0110000 from *Oryza sativa Japonica* Group (NP001056579), cytochrome P450, family 88, subfamily A, polypeptide 3 from *Arabidopsis thaliana* (NP_172008), ent-kaurenoic acid hydroxylase 2 *A. thaliana* (NP180803), cytochrome P450, probable ent-kaurenoic acid oxidase from *Populus trichocarpa* (XP_002321248), ent-kaurenoic acid oxidase family protein from *P. trichocarpa* (XP_002319447), ent-kaurenoic acid oxidase 2 (predicted) from *Vitis vinifera* (XP_002265630), ent-kaurenoic acid oxidase 1 (predicted) from *V. vinifera* (XP_002264215), cytochrome P450 88A1 from *Sorghum bicolor* (XP_002436354), ent-kaurenoic acid oxidase 2 from *Ricinus communis* (XP_002526524), ent-kaurenoic acid hydroxylase 1 from *Arabidopsis lyrata* subsp. *Lyrata* (XP_002892270), ent-kaurenoic acid hydroxylase from *A. lyrata* subsp. *Lyrata* (XP_002879389), kaurenoic acid oxidase from *Selaginella moellendorffii* (XP_002966472), ent-kaurenoic acid oxidase 2 isoform X1 from *S. moellendorffii* (XP_002975682), hypothetical protein SELMODRAFT_1840 from *S. moellendorffii* (XP_002993510), hypothetical protein SELMODRAFT_115926 from *S. moellendorffii* (XP_002982048), kaurenoic acid oxidase from *S. moellendorffii* (XP_002988455) and p450MO from *Bacillus cereus* AOA-CPS1.

CHAPTER SEVEN

The chapter has been ***published in International Journal of Biological Macromolecules***

(2020 Impact Factor 5.162)

International Journal of Biological Macromolecules 161 (2020) 875–890

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: http://www.elsevier.com/locate/ijbiomac

Role of tetrachloro-1,4-benzoquinone reductase in phenylalanine hydroxylation system and pentachlorophenol degradation in *Bacillus cereus* AOA-CPS1

Oladipupo A. Aregbesola, Ajit Kumar, Mduduzi P. Mokoena, Ademola O. Olaniran *

Discipline of Microbiology, School of Life Sciences, College of Agriculture, Engineering and Science, University of KwaZulu-Natal (Westville Campus), Private Bag X54001, Durban 4000, South Africa

ARTICLE INFO

Article history:
Received 22 April 2020
Received in revised form 9 June 2020
Accepted 9 June 2020
Available online 11 June 2020

Keywords:
Phenylalanine hydroxylating system
Tetrachloro-1,4-benzoquinone reductase
Ping-Pong mechanism
Phenylalanine 4-monooxygenase
Pterin-4α-carbinolamine dehydratase

ABSTRACT

This study reports a ≅12.5 kDa protein tetrachloro-1,4-benzoquinone reductase (CpsD) from *Bacillus cereus* strain AOA-CPS1 (*Bc*AOA). CpsD is purified to homogeneity with a total yield of 35% and specific activity of 160 U·mg^{-1} of protein. CpsD showed optimal activity at pH 7.5 and 40 °C. The enzyme was found to be functionally stable between pH 7.0–7.5 and temperature between 30 °C and 35 °C. CpsD activity was enhanced by Fe^{2+} and inhibited by sodium azide and SDS. CpsD followed Michaelis-Menten kinetic exhibiting an apparent v_{max}, K_m, k_{cat} and k_{cat}/K_m values of 0.071 μmol·s^{-1}, 94 μmol, 0.029 s^{-1} and 3.13 × 10^{-4} s^{-1}·μmol^{-1}, respectively, for substrate tetrachloro-1,4-benzoquinone. The bioinformatics analysis indicated that CpsD belongs to the PCD/DCoH superfamily, with specific conserved protein domains of pterin-4α-carbinolamine dehydratase (PCD). This study proposed that CpsD catalysed the reduction of tetrachloro-1,4-benzoquinone to tetrachloro-*p*-hydroquinone and released the products found in phenylalanine hydroxylation system (PheOHS) via a Ping-Pong or atypical ternary mechanism; and regulate expression of phenylalanine 4-monooxygenase by blocking reverse flux in *Bc*AOA PheOHS using a probable Yin-Yang mechanism. The study also concluded that CpsD may play a catalytic and regulatory role in *Bc*AOA PheOHS and pentachlorophenol degradation pathway.

1. Introduction

Aromatic organic compounds are the second most abundant environmental organic pollutants [1], some of which are bioremediated via many degradation pathways [2]. However, overcoming the resonance energy that stabilizes the aromatic ring of these compounds remains the major rate-limiting factor [2]. The aromatic ring skeleton makes many aromatic organic compounds unreactive to neither oxidation nor reduction and requires a complex catabolic system such as phenylalanine hydroxylating system (PheOHS). Phenylalanine (Phe) is an essential amino acid that cannot be produced in the body and must be included in the diet [3]. Metabolism of Phe is an irreversible hydroxylation to tyrosine in the PheOHS [4].

The typical mammalian PheOHS (Fig. S1) involve phenylalanine 4-monooxygenase (Phe4MO), pterin-4α-carbinolamine dehydratase/dimerization of hepatocyte nuclear factor-1 (PCD/DCoH), dihydropteridine reductase, tetrahydrobiopterin (BH4) and molecular oxygen [5]. Phe4MO hydroxylates Phe to Tyr while PCD/DCoH and Dihydropteridine reductase convert the intermediate back to BH4 [4]. PCD/DCoH is a protein with two contrasting functions (as PCD and as cofactor hepatocyte nuclear factor (HNF-1) dimerization) in mammals, based on a change in oligomeric states of the protein [6]. Unlike mammalian and other higher animals, PheOHS in most bacteria do not produce BH4, the primary role of PCD/DCoH in some bacterial PheOHS is to convert dihydroneopterin triphosphate to a coenzyme (folate) for one-carbon transfer reactions [7].

NADH dependent quinone oxidoreductase (NDQR) detoxifies quinone-based substrates to hydroquinone (HQ) via two-electrons reduction [8–10], which protect cells against the toxicity of quinones [11]. The hydroquinone produced by NDQR is more stable and readily conjugates with autogenous ligands such as glutathione (GSH) to facilitate further detoxification [8]. Based on the stability, some hydroquinones may flux back to a semiquinone free radical and/or parent compound with the molecular oxygen reduced to a reactive species (ROS), which can elicit cytotoxicity and DNA damage [8–10]. For instance, hydroxylation of pentachlorophenol (PCP) produces a toxic tetrachloro-1,4-benzoquinone (Tet-CBQ) intermediate [12], which abrogated several cellular functions [13–17]. Also, Tet-CBQ converted to tetrachloro-*p*-hydroquinone (Tet-CHQ) in PCP degradation pathway of *Sphingobium chlorophenolicum*, reverses flux [18,19] and the activity of PCP hydroxylase is hindered by its promiscuity with Tet-CHQ [18].

The presence of PheOHS in bacteria suggests that bacteria can utilize the system in folate synthesis and xenobiotics degradation. Moreover, enzymes associated with intermediary catabolism (such as PheOHS)

* Corresponding author.
E-mail address: olaniranna@ukzn.ac.za (A.O. Olaniran).

https://doi.org/10.1016/j.ijbiomac.2020.06.083
0141-8130/

may exhibit unrecognised alternative catalytic function(s) to protect the organism from potential hazardous and toxic compounds [20]. Previously, some studies have reported the involvement of PheOHS in xenobiotics degradation [1,4,7,20–23], although, no report on the roles of PheOHS nor its enzymes on the biotransformation of persistent organochlorine pesticides such as PCP and its derivatives exist. Additionally, the bi/multi-functional role of PCD/DCoH and biochemical characterization of any putative PheOHS system in the *Bacillus cereus* group has not been reported previously. Hence, this study reports on the detection, amplification, cloning, overexpression, purification, biochemical characterization and structural homology modelling of a Tet-CBQ reductase (CpsD), a homolog of human PCD/DCoH and NADH quinone oxidoreductase from *Bacillus cereus* strain AOA-CPS1 recently isolated from contaminated wastewater in Durban, South Africa. The possible role of CpsD in phenylalanine hydroxylation system and PCP degradation pathway in AOA-CPS1 was also elucidated.

2. Materials and methods

2.1. Materials

Tetrachloro-1,4-benzoquinone (Tet-CBQ, 99.0%), sodium azide (≥99.0%), β-nicotinamide adenine dinucleotide reduced disodium salt hydrate (NADH, ≥97.0%), isopropyl β-D-1-thiogalactopyranoside (IPTG, ≥99.0%), Luria Bertani (LB) agar and broth (Vegitone), imidazole ACS (≥99.0%), *N*-methyl-*N*-trimethylsilyl-trifluoroacetamide (TMS, ≥98.5%), sodium dodecyl sulfate (SDS), phenylmethylsulfonyl fluoride (PMSF, ≥98.5%), ampicillin sodium salt (Amp), 4-morpholineethanesulfonic acid hemisodium salt (MES, ≥98.0%) and 2-mercaptoethanol (2-ME, ≥99.0%) were purchased from Merck (Merck, MA, USA). Dithiothreitol (DTT, ≥99.5%) molecular biology grade, ethylenediaminetetraacetic acid ACS (EDTA, 99.4%), *N,N,N',N'*-tetramethylethylenediamine (TEMED, ≥99.0%), restriction endonuclease (*Nde*I and *Bam*HI) FastDigest, T4 DNA ligase, DNA ladders (100 bp plus, 1 kb and 1 kb plus), page ruler plus prestained protein ladder (10–250 kDa), Phusion high-fidelity DNA polymerase and PCR reaction mix were purchased from ThermoFisher Scientific (Waltham, MA, USA). Bradford reagent and 40.0% acrylamides/bis-acrylamide solution was purchased from Bio-Rad (Bio-Rad, CA, USA). *Escherichia coli* strains DH5α and BL21 (DE3) (ThermoFisher Scientific, Waltham, MA, USA) were used as cloning and protein overexpression hosts, respectively. Expression vector pET15b (Merck, MA, USA) was used for cloning and protein expression. All chemicals and reagents used in this study were of analytical grade standards. Tet-CBQ was dissolved in acetone and diluted with 50 mmol $NaPO_4$ buffer pH 7.0. Chemically competent cells of *E. coli* strains DH5α and BL21(DE3) were prepared as previously described [24].

2.2. Isolation and characterization of Bacillus cereus strain AOA-CPS1 (BcAOA)

*Bc*AOA was isolated from an activated sludge of a wastewater treatment plant in Durban, South Africa, via culture enrichment and purified through successive sub-culturing on sterile nutrient agar plates until distinct colonies were obtained. Briefly, after three successive sub-culturing, 0.1 mL of the enriched culture was spread inoculated on minimal salt agar plates supplemented with 50 $mg \cdot L^{-1}$ of PCP. The plates were incubated at 30 °C until visible growths were observed. The isolate was purified via successive sub-culturing on sterile nutrient agar (NA) plates until distinct colonies were obtained. The isolate (*Bc*AOA) was screened in a PCP-fortified minimal salt medium (PCP-MSM) and observed to demonstrate good potential for PCP transformation (data not shown), and therefore selected for further studies. The pure isolate was preserved in glycerol with and without 10 ppm pentachlorophenol (PCP) in the bio-freezer at −80 °C. Further, the genomic DNA of the isolate was extracted using a Quick-DNA™ fungal/bacterial miniprep kit (Zymo Research Corporation, CA, USA) from 5 mL of Luria-Bertani (LB) grown culture of the pure single colony of the isolate. The 16S rDNA fragment was amplified from the isolated genomic DNA via PCR using 63F and 1387R universal primers pair [25], sequenced (Inqaba Biotech, Pretoria, South Africa) and submitted at NCBI BLASTn server [26] for the identification of the pure culture. The bacterium show >99% sequence homology with *Bacillus cereus* strains and named as *Bacillus cereus* strain AOA-CPS1 (unpublished data). The 16S rDNA sequence was deposited to the NCBI GenBank with an accession number MH504118.1.

2.3. Detection, PCR amplification and cloning of cpsD

To detect *cps*D from *Bc*AOA, PCR was performed using whole genomic DNA isolated from *Bc*AOA (purified using Quick-DNA Fungal/Bacteria Kit) as template and the primer pair 1 (forward: 5'-TAAATGATGCT AAGATTAACTGAA-3'), and (reverse: 5'-GTATTATTTTCTTATAATT GC-3'). Primer pair 1 was designed using *Bacillus* spp. entries (from the NCBI BLASTn for 16S rDNA) with whole-genome sequences. The whole genomes were screened for the presence of pterin-4α-carbinolamine dehydratase (PCD). The whole-genomes of *B. thuringiensis* strain L-7601 and *B. thuringiensis* strain ATCC 10792 showed the presence of a putative PCD gene. The gene sequences were aligned using DNAMAN v10.0.2.100 software (Lynnon, Biosoft, USA) and the primer pair 1 were designed. The primer pair 1 successfully amplified ≅315 bp fragment from the genomic DNA of *Bc*AOA. The amplified gene fragment was sequenced at Inqaba Biotechnical Industry (Pty) Ltd. (Pretoria, South Africa) and again submitted at NCBI BLASTn. The gene showed a 100% homology with a 315 bp nucleotide sequences from many species of *Bacillus cereus* group (data not shown). Subsequently, the primer pair 2, forward: 5'-TAACATA TGATGATGCTAAGATTAACTGAA-3' and reverse: 5'-GTAGGATCCTTAT TTTCTTATAATTGC-3' targeting expression vector pET15b for cloning purpose. Underlines are the restriction site sequences to use *Nde*I and *Bam*HI restriction enzymes. The gene was amplified in a 10 μL PCR mix containing 20 μmol of dNTPs, 1.5 mmol of $MgCl_2$, 1.0 μmol of each primer (primer pair 2), 1 μL of genomic DNA, 1.25 U of high-fidelity DNA polymerase, 1 μL of 10× buffer and 5.15 μL of ddH_2O. The PCR conditions were used as followed: initial denaturation at 95 °C for 5 min, 95 °C for 1 min, annealing at 54 °C for 1 min, extension at 72 °C for 1 min with 25 cycles, and a final extension at 72 °C for 10 min.

The amplified PCR product was extracted from agarose gel using GeneJet gel purification kit (Thermo-Scientific) and cloned into pET15b expression vector using standard gene cloning protocol [27]. The recombinant pET15b-cpsD plasmid was transformed into *E. coli* BL21(DE3) for the overexpression of protein. The in-frame cloning of *cps*D and the correct sequence was confirmed by sequencing (Inqaba Biotechnical Industry (Pty) Ltd, Pretoria, South Africa) using the universal T7 promoter and T7 terminator primer pair.

2.4. Overexpression and purification of CpsD

The transformed *E. coli* BL21(DE3) harbouring pET15b-cpsD was pre-grown overnight at 37 °C in 100 mL LB broth supplemented with of 100 $\mu g \cdot mL^{-1}$ ampicillin (LB-amp) followed by inoculating in 2 L LB-amp and incubated at 37 °C until optical density (OD) at wavelength 600 nm reached 0.6. The culture was induced with 1 mmol IPTG and incubated at 20 °C for 24 h, shaking at 200 rpm. The biomass was harvested by centrifugation at 8000 rpm for 15 min and the supernatant was decanted. The pellets were re-suspended in 50 mmol sodium phosphate ($NaPO_4$) buffer (pH 7.0) containing 0.5 mmol each of DTT and PMSF, washed twice and re-suspended in 100 mL of the same buffer. The re-suspended cells were sonicated at Psi 40 and 40 V for 5 min with 10 s pulse (Omni Sonic Ruptor 400, Omni International, GA, USA). The lysate was centrifuged for 30 min at 15,000 rpm and 4 °C

and CpsD was purified from the supernatant. The total protein concentration in the supernatant was determined using the Bradford method [28].

To purify CpsD, 5 mL (20.83 mg total protein) of the supernatant was loaded in 5 mL Pierce HisPur Cobalt Chromatography Cartridge (#90094, Thermo Scientific, IL, USA) connected to ÄKTA purifier system (GE Healthcare Life Sciences, IL, USA) and the recombinant 6 × His-tagged CpsD protein was purified following the manufactural specifications. The eluted fractions showing a single protein peak of interest at 280 nm were pulled and concentrated using 10 kDa cut-off Amicon Ultra-15 centrifugal filter unit (Merck, MA, USA). The expression of CpsD, level in the supernatant, homogeneity of the purified protein and molecular weight was determined by loading 25 μL (200 μg total protein) of the sample in 12% SDS-PAGE [29]. The enzymatic activity of CpsD at each purification steps was determined as described below.

2.5. Enzyme activity assay

Enzyme activity of CpsD was determined in a 1 mL reaction mixture by continuous spectrophotometric method [12], with some modifications. The 1 mL reaction mixture contained 100 μmol Tet-CBQ, 150 μmol NADH, 50 mmol $NaPO_4$ (pH 7.0), 350 μmol ascorbate and 125 μL (30 μg) CpsD. Ascorbate was added to the reaction mixture to protect Tet-CBQ to be formed as Tet-CHQ [18] without the enzymatic reaction. The decrease in absorbance at wavelength 340 nm which measures the depletion of NADH. The oxidation of NADH due to the enzymatic reaction was measured by determining its concentration at the start and end of the reaction (using molar extinction coefficient $\varepsilon 340 = 6200\ mol^{-1} \cdot cm^{-1}$) [18]. One unit of enzyme activity was defined as the amount of CpsD that caused 1 μmol of NADH oxidised per minute under standard assay conditions. All experiment were performed in triplicate and the results presented were the means and standard deviations of a triplicate experiment, unless otherwise stated.

2.6. Determination of optimum pH and pH stability

To determine the optimum pH for the CpsD activity, 10 μL (30μg) of purified CpsD was assayed as described above except the reactions were performed at different pH. The buffers for different pH used were sodium acetate-acetic acid (pH 4.0–5.0), NaH_2PO_4-Na_2HPO_4 (pH 6.0–8.0), glycine-NaOH (pH 9.0–10.0), Na_2HPO_4-NaOH (pH 11–12) and KCl-NaOH (pH 13.0). To determine the functional stability of purified CpsD, the enzyme (30 μg) was pre-incubated in buffers ranging from pH 7–13. The aliquots were withdrawn at 15 min interval for 120 min and enzyme activity was measured at optimum conditions as described above. The activity obtained at optimum pH 7.5 was considered as 100% residual activity.

2.7. Determination of optimum temperature and temperature stability

To determine the optimum temperature for the CpsD activity, 10 μL (30 μg) of the purified enzyme was assayed as described above except the reactions were performed at different temperatures, ranging from 25 °C to 60 °C at the optimal pH 7.5. The functional stability of CpsD at different temperatures was investigated by incubating the enzyme (30 μg) at temperatures (40, 50, 60 and 60 °C). The aliquots were withdrawn at 20 min interval for 120 min and enzyme activity was measured at optimum conditions as described above. The residual activity of CpsD at optimum conditions was considered as 100% residual activity.

2.8. Determination of kinetic parameters

The purified CpsD (30 μg, 2.4 μmol final concentration) was incubated with 20, 40, 80, 120, 160 and 200 μmol of substrates in the presence of 240 μmol NADH and 350 μmol ascorbate. CpsD activity was determined at optimum conditions as described above. The K_m and v_{max} values were calculated by applying the Michaelis-Menten equation in Lineweaver-Burk plot. The catalytic constant of the enzyme-substrate reaction (k_{cat}), also referred to as the turnover number was determined using the equation, $k_{cat} = v_{max} / [E]t$, where v_{max} is the maximum velocity and [E]t, is the total enzyme concentration. Catalytic efficiency was calculated by the equation, k_{cat}/K_m.

2.9. CpsD activity in the presence of metal ions and inhibitors

To determine the CpsD activity in the presence of metal ions and inhibitors, 30 μg of purified CpsD was pre-incubated with the different concentrations of metal ions and inhibitors for 10 min and assays were performed as described above. The enzyme activity determined without the addition of any metal ion or inhibitor was considered as 100% residual activity [30].

2.10. In-gel trypsin digestion and identification of the purified CpsD in ES-MS

The pure CpsD (50 μg) was loaded onto 12% SDS-PAGE and stained with Coomassie blue R250. The protein band was excised and digested with trypsin and fragments analysed by electrospray mass spectrometry (ES-MS) (Central Analytical Facility, Stellenbosch University, Stellenbosch, South Africa). The data analysis and protein identification were done as described previously [31].

2.11. Template-based structure prediction and homology modelling for CpsD

Three-dimensional structure and homology modelling of the protein were predicted by submitting the amino acid sequence at the SWISS-MODEL tool at *ExPASy* bioinformatics resource portal workspace (https://swissmodel.expasy.org/interactive). The default parameters used for performing the automated SWISS-MODEL were as explained previously [32,33] and elaborated at (https://swissmodel.expasy.org/docs/examples) webpage. The modelled PDB files were submitted to an online tool (PDBsum) for determining the structural summary [34].

2.12. Evolutionary relationships of CpsD with other dioxygenases

To study the evolutionary origin of CpsD with other Tet-CBQ reductases (PcpD) and PCD of human, animals, fungi and bacteria origin, the phylogenetic based evolutionary analysis was performed using Neighbor-Joining method [35]. The evolutionary history, bootstrap consensus tree, percentage of replicate trees [36] and the evolutionary distance tree [37] were performed using MEGAX software [38].

3. Results and discussion

3.1. Cloning, overexpression and purification of 6xHis-tagged CpsD

Bacillus cereus strain AOA-CPS1 (*Bc*AOA) was isolated from wastewater activated sludge and identified via PCR amplification, sequencing and analysis of 16S rDNA (data not shown). A 315 bp gene encoding tetrachloro-1,4-benzoquinone (Tet-CBQ) reductase (*cpsD*) was amplified from the genomic DNA of *Bc*AOA (Fig. 1A). The *cpsD* gene shared 100% sequence homology with putative PCD/DCoH from *Bacillus* sp. strain COPE52 (CP031292.1) and *Bacillus* sp. FDAARGOS_235 (CP020437.2). The *cpsD* gene is located on the same operon with an aromatic amino acid hydroxylase gene in the chromosomes of some *Bacillus* spp. that harbours the biological system. The *cpsD* gene was successfully cloned into an expression vector pET15b, and the recombinant C-terminal 6 × His-tagged CpsD was overexpressed in *E. coli* BL21 (DE3) (Fig. 1B). The recombinant CpsD protein was purified to homogeneity and showed M_r of ≅12,500 (Fig. 1C). The CpsD exists in a

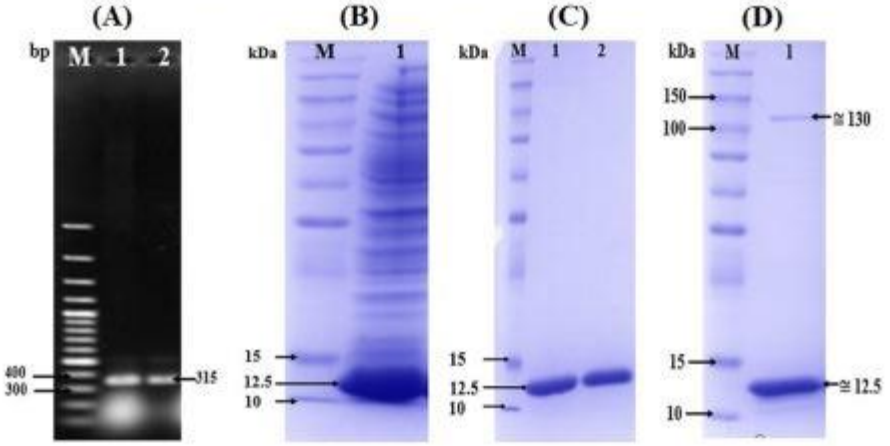


Fig. 1. PCR amplification, overexpression and purification of CpsD. (A) Amplification of *cpsD* gene from *BcAOA* genomic DNA; DNA marker (Lane M), amplified *cpsA* gene (Lane 1 and 2). (B) 12% SDS-PAGE showing overexpression of CpsD; protein marker (Lane M), induced CpsD (Lane B). (C) AKTA fractions of purified recombinant 6 × His-tagged CpsD; protein marker (Lane M), purified CpsD (Lane 1 and 2). (D) protein marker (Lane M), Monomeric form ≅12.5 kDa and multimer form ≅130 kDa of CpsD (Lane 1). Molecular weight ≅12.5 kDa indicated is based on the biophysical properties calculated at ProtPram tool at ExPASy for CpsD amino acid sequence.

monomeric form similar to PCD/DCoH from *Martierella alpina* [4] and multimeric form with M_r of ≅130,000 (Fig. 1D) in accordance to PCD/DCoH from mammalian and other higher animals [1,4,23,39–50].

3.2. Enzyme activity

A single-step purification strategy (affinity chromatography) purified the enzyme 10.41-fold with a total yield of 35% (Table 1). The specific activity of the purified enzyme was estimated as 160 U · mg^{-1} of the protein. To make sure of the accuracy of the enzyme assay, the absorbance of each component of the reaction mixture was recorded using the wavelength scanning facility in the spectrophotometer. The purified CpsD absorbed at wavelength 280 nm with a long shoulder at 220 nm (Fig. S2a). Pure Tet-CBQ dissolved in acetone absorbed at 270 nm with a short peak at wavelength 321 nm (Fig. S2b), while NADH dissolved in 50 mmol sodium phosphate buffer (pH 7.0) absorbed at 340 nm with higher peaks at 260 nm and 210 nm (Fig. S2c). The absorption of Tet-CBQ at 270 nm could not interfere with the reduction in absorbance of NADH at 340 nm. Therefore, the activity of CpsD was measured by the reduction of NADH monitored at wavelength 340 nm. The activity of the enzyme was enhanced in the presence of NADH with a noticeable shifting of the peak at 330 nm (Fig. S2d) in accordance to a previous report [12,51]. A statistically significant non-enzymatic reduction of NADH absorbance in a reaction mix that contained Tet-CBQ without the enzyme also occurred (Fig. S2e) as previously reported [12,51]. However, the rate of reduction of NADH absorbance (0.01 min^{-1}) was found to be faster than reduction of NADH absorbance of the non-enzymatic (0.016 min^{-1}) reaction (Fig. 2A). Therefore, change in absorption in non-enzymatic reaction (control experiment) was deducted from all enzymatic assay as earlier suggested [12,51]. Additionally, continuous spectrophotometry reaction of CpsD with Tet-CBQ in the presence of NADH and excess ascorbate leads to an increase in absorption at 327 nm (Fig. 2B), with a stoichiometric increase in absorption at 340 nm (Fig. S2f). Increase in absorption at 327 nm represents the formation and accumulation of Tet-CHQ in the reaction mixture.

3.3. Optimum pH and pH stability of CpsD

The optimum pH of 7.5 obtained for CpsD is in agreement with the optimum temperature of PCD from human [49,52] and bovine liver [53], but contrary to those of *Chromobacterium violaceum* [54], fungi [55] and Tet-CBQ reductase (PcpD) from *S. chlorophenolicum* [12]. The CpsD activity increased gradually from pH 5.0 to pH 7.5 and drastically decreased at pH 8.0, with no activity observed at pH[53] 9.5 (Fig. 3A), similar to the reported activity for PcpD [12]. CpsD was found to be functionally stable between pH 7.0–7.5, retaining 95% of its activity between pH 6.5 and pH 7.5, with above 80% residual activity at pH 6.0 relative to activity at pH 7.5 (Fig. 3B).

The stability of CpsD in alkaline pH may be due to the adaptability of the PCP-degrading *BcAOA* to alkaline environment. Many PCP impacted sites are marred by acidic or alkaline pH and high salt concentrations and PCP-degrading microbes that thrived in these environments, play prominent role in bioremediation. This also corroborates earlier report that >80% of PCP was used as wood preservation in combination with sodium hydroxide, which can turn most PCP-impacted sites into alkaline environments [56]. Furthermore, PCP is mostly dissolved in sodium hydroxide for biodegradation assay since the compound is sparingly soluble in water. This also buttresses the report that bacterial PCP

Table 1
Purification of Tet-CBQ reductase (CpsD) from *Bacillus cereus* strain AOA-CPS1.

Purification step	Total volume (mL)	Total protein (mg · mL^{-1})	Total activity (U · mL^{-1})	Total activity (U)	Specific activity (U · mg^{-1})	Fold purification	Yield (%)
Crude	100[a]	4.17	64.66[c,e]	6466	15.38	1.00	100
HisTrap	16[b]	0.89	142.5[d,e]	2280	160	10.41	35

Units defined explained in enzyme activity assay section.
[a] The cell lysate volume after sonication.
[b] The total volume of eluted fractions.
[c] 8 μL crude enzyme extract used for assay.
[d] 12 μL pure enzyme fraction used for assay.
[e] Calculated using molar extinction coefficient ε340 = 6200 M^{-1} · cm^{-1} for NADH.

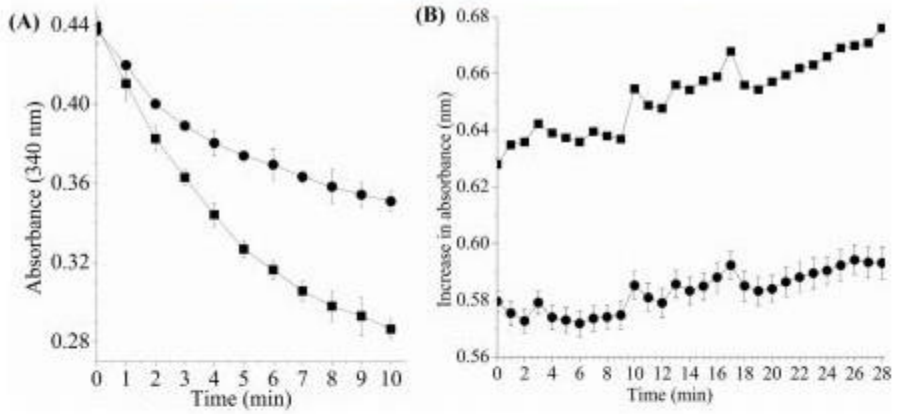


Fig. 2. CpsD reaction with substrate Tet-CBQ in the presence of NADH and ascorbate. (A) 1.0 mL reaction mixture contained 30 μg CpsD, 150 μmol Tet-CBQ, 160 μmol NADH in 50 mM $NaPO_4$ buffer (pH 7.0) (■), and without enzyme (●). (B) Continuous spectrophotometry reaction of CpsD with Tet-CBQ in the presence (■) and absence (●) of 350 μmol ascorbate.

transformation occurred more rapidly at neutral to alkaline conditions than acidic conditions [57].

Previous studies on the optimization of biodegradation process parameters for effective and efficient transformation of PCP by *Bc*AOA show that the optimum pH for the growth of *Bc*AOA and maximum transformation of PCP is between pH 7.0–7.4 (data not shown), which means that catabolic enzymes of *Bc*AOA involved in the PCP transformation might work optimally at an alkaline pH, which is consistent with the finding of this study. Plausible reason for the rapid loss of Tet-CBQ reductase activity at alkaline pH have been attributed to the fact that the hydroxyl (OH^-) ion of Tet-CBQ might decrease the positive electrostatic potential of the enzyme at its active site thereby decreasing its binding capacity [12].

3.4. Optimum temperature and temperature stability of CpsD

CpsD showed optimum activity between 30 °C and 40 °C in consonance with human PCD [49,52], but contrary to that of *Chromobacterium violaceum* [54]. The activity of CpsD at 40 °C is slightly higher than 30 °C and 35 °C similar to that of human PCD [58]. The enzyme showed >80% residual activity between 35 °C and 55 °C and decrease to about 75% at 60 °C (Fig. 4A). Moreover, CpsD showed more functional stability between 30 °C-35 °C than at 40 °C. The enzyme retained >90% of its activity at 30 °C–35 °C for 120 min (Fig. 4B). CpsD also has about 90% residual activity between 40 °C and 55 °C for 40 min, however, the activity decreased to about 78% within 80 min and <10% at 60 °C within 120 min.

The optimum temperature of CpsD activity showed that the enzyme can work maximally if used in wastewater remediation without any additional cost to mimic environmental temperature to suit enzyme activity. The stability of CpsD between 25 °C and 30 °C and its ability to retain about 90% of its activity between 35 °C and 40 °C may be attributed to the fact that the organism is adapted to water. This stability/mesophilic nature of CpsD and other pterin-4α-carbinolamine dehydratases (PCD's) may also be due to the fact that most PCD's enzymatic reaction takes place at physiological body temperature. The activity of CpsD was attenuated by about 90% after incubation at 60 °C for 120 min, which is not surprising based on the source of the isolate.

3.5. CpsD activity in the presence of metal ions and inhibitors

CpsD exhibited maximum activity in the presence of Fe^{2+} ion compared to activity in buffer only. Cobalt ion also slightly stimulated CpsD

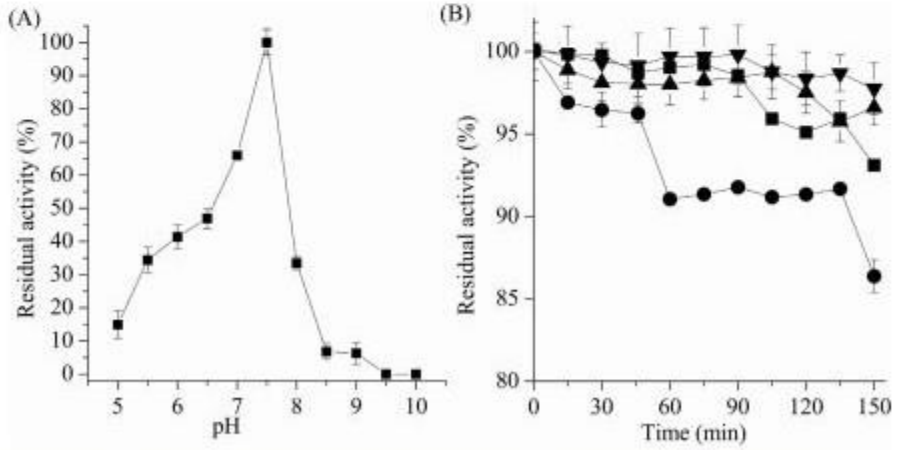


Fig. 3. Optimum pH and pH stability of CpsD. (A) % residual activity of CpsD activity at pH 5–10. (B) Functional stability of CpsD incubated at pH 6.0 (■), 6.5 (●), 7.0 (▲), and 7.5 (▼) for 150 min.

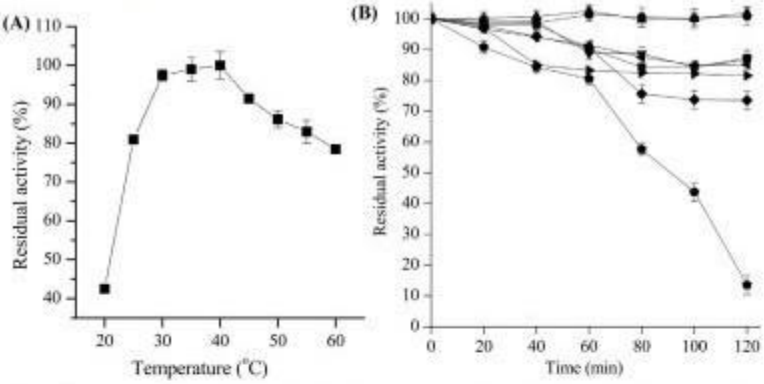


Fig. 4. Optimum temperature and temperature stability of CpsD. (A) % residual activity of CpsD at temperature from 20 to 60 °C. (B) Functional stability of CpsD incubated at 25 (■); 30 (●); 35 (▲); 40 (▼); 45 ( ), 50 (►), 55 (♦) and 60 °C ( ) for 120 min.

activity while the other metal ions (Zn^{2+}, Mn^{2+}, Fe^{3+}, Ca^{2+}, Cu^{2+}, Mg^{2+}, Pb^{2+}, Ni^{2+} and Hg^{2+}) could not enhance the enzyme activity (Table 2). Previously, Fe^{2+} was reported to enhance the oxidation of BH4 in the biopterin synthesis system [59]. Metal ions are present at the active sites of some enzymes to coordinate and reduce dioxygen (O_2) to electrophilic activated oxygen species needed to attack carbon-hydrogen bond on substrate [60–62].

CpsD exhibited maximum activity in the presence of 1.5 mmol EDTA compared to activity in buffer only (Table 3). About 2-fold and 3.4-fold increase in CpsD activity was obtained in the presence of 1.0 mmol and 1.5 mmol EDTA, respectively. 0.5 mmol EDTA resulted in about 10% loss in CpsD activity while 2.0 mmol EDTA completely inhibited the activity. Furthermore, the activity of CpsD was also enhanced in the presence of DTT (1.0 mmol), similar to previous report [63]. On the contrary, sodium azide and SDS showed inhibitory effects on CpsD activity at all the concentrations tested.

3.6. Steady-state kinetic parameters of CpsD

The increase in Tet-CBQ concentration in the reaction mixture resulted in increased CpsD activity (Fig. 5A) and rate of initial velocity (Fig. 5B) in the presence of 240 μmol NADH contrary to earlier report [12]. The reaction followed a typical Michaelis-Menten kinetics, with apparent v_{max}, K_m, k_{cat} and k_{cat}/K_m calculated from the Lineweaver-Burk plots using the reciprocal of the initial reaction velocity and Tet-CBQ concentration found to be 0.071 μmol·s^{-1}, 94 μmol, 0.029 s^{-1} and 3.13×10^{-4} s^{-1}·μmol^{-1}, respectively (Fig. 5C). The steady state kinetic parameters for the reaction between CpsD and Tet-CBQ were found to be higher than those reported for CpsD reaction with other substrate in *Pseudomonas aeruginosa* [64,65].

Table 2
Effect of metal ions on CpsD activity.

Metal ions	Residual activity (%)
Zn^{2+}	89.65 ± 0.07
Mn^{2+}	90.67 ± 4.01
Fe^{3+}	90.01 ± 0.08
Fe^{2+}	120.21 ± 2.82
Ca^{2+}	99.27 ± 0.05
Cu^{2+}	0.00 ± 0.00
Mg^{2+}	93.74 ± 2.01
Pb^{2+}	0.00 ± 0.00
Ni^{2+}	0.00 ± 0.00
Hg^{2+}	0.00 ± 00
Co^{2+}	107.06 ± 1.5
Buffer	100[a]

[a] Enzyme activity in a buffer without metal was taken as 100% residual activity.

3.7. Confirmation of the identity of purified CpsD

The excision of pure SDS-PAGE followed by tryptic digestion resulted in the generation of small peptides of CpsD. The LC-MS analysis of the peptides and resolution with a multiple search engine (SearchGui v3.3.15) followed by virtualisation on PeptideShaker (version 1.16.45), showed the peptide structure matched with multi-species PCD (UniProt accession number Q818B3) from *B. cereus* strains ATCC 14579/DSM 31/JCM 2152/NBRC 15305/NCIMB 9373/NRRL B-3711, with 57.69% coverage, precursor charges of 2 and a 100% peptide matching confidence. The protein Q818B3, containing 104 amino acids residue, was reported only at the prediction level and no experimental evidence existed. The present study, however, fills the gap and report CpsD at an experimental level. The spectrum overview; protein, peptides and PSMs identification summary; peptides structure matches and spectrum identification results, as well as protein, peptides and PSMs validation and quality control plots, are shown in Fig. S3a–h. Furthermore, the NCBI protein BLAST (protein to protein) search indicated that CpsD shared 100% sequence homology with multispecies: PCD from *Bacillus* spp. (WP_000979552.1) and *B. toyanensis* (WP_098203324.1) and >97% sequence identity with *B. cereus* group (WP_000979535.1) (Fig. S4a). This confirmed that CpsD expressed by the cloned *cpsD* gene fragment from *Bc*AOA is the protein of interest.

Putative conserved domains were detected in CpsD structural model by using the CDD/SPARCLE functional classification of proteins via subfamily domain architectures [66], in NCBI's conserved domain database [68], conserved domain database for the functional annotation of protein [67] and protein domain annotation on the fly [68]. CpsD is classified as PCD and belongs to PCD/DCoH superfamily. The putative conserved domains were detected for 6 substrates (chemicals) binding sites (at residues His55, His56, His57, Ser72, Trp74 and Asn75); DCoH

Table 3
Residual CpsD activity (%) in presence of different inhibitors concentrations.

Inhibitors	Inhibitor concentration (mmol)			
	0.5	1.0	1.5	2.0
EDTA	90.21 ± 5.4	202.44 ± 8.8	339.20 ± 5.2	0
DTT	89.23 ± 4.6	278.98 ± 2.8	108.04 ± 3.0	87.88 ± 1.9
Na-Azide	0	0	0	0
SDS	0	0	0	0

The control experiment without enzyme inhibitor was considered 100% residual activity.

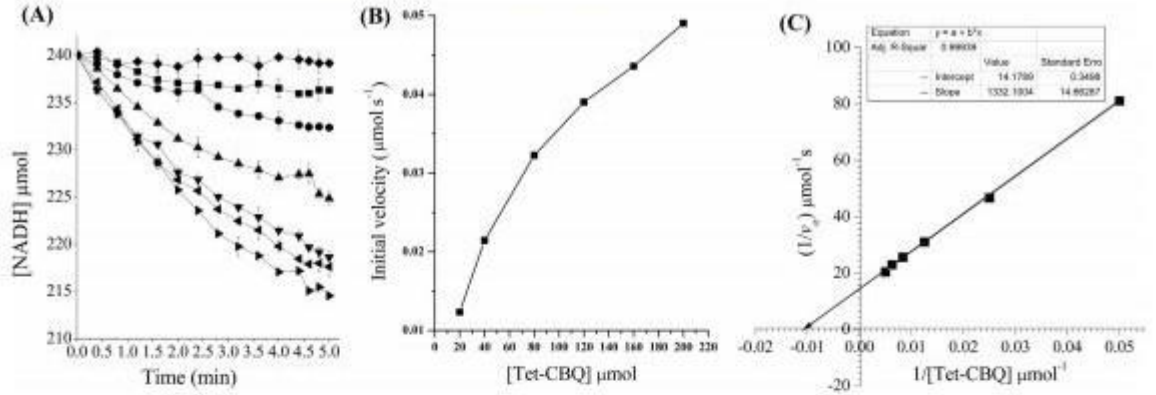


Fig. 5. Steady state kinetics of CpsD enzymatic reaction. (A) Continuous spectrophotometric reaction monitored at wavelength 340 nm in the presence of 20 (■) 40 (●) 80 (▲) 120 (▼), 160 ( ) and µmol 200 (►) Tet-CBQ and 240 µmol NADH. Decrease in absorption at 340 nm represents NADH oxidation due to the reaction. The control experiment (•) represents the reaction in the absence of enzyme and presence of 200 µmol Tet-CBQ and 240 µmol NADH. (B) Initial velocities (v_0) of the reaction at increasing concentration of Tet-CBQ as calculated from the curves shown in 5A. (C) double reciprocal plot of data shown in 5B.

dimer interaction site (at residues Tyr37, Val41, Ser45, Glu52, His57, Pro58, Phe59, Ile60, Leu61, Ile62, Gln63, and Tyr64); DCoH/HNF-1 dimer interaction (with Leu38, Lys39, Ser45, Glu46, Lys49, Glu52 and Glu53); DCoH tetramer interaction (at residues Leu38, Ser45, Glu46, Lys49 and Glu53); aromatic arch (on residues Tyr37, Val41, His57, Ile60 and Trp74) sites (Fig. S4b).

The specific conserved domains hits observed were PCD (E-value 1.14e−28); PCD/DCoH (E-value 4.86e−27); phhB (PRK00823, E-value 1.14e−26) and a non-specific conserved domains hit was PhhB (COG2154). The conserved domains pterin-4α- (accession: pfam01329, a member of the superfamily cl00942), is a PCD/DCoH [70]; PCD/DCoH (accession: cd00488, a member of the superfamily cl00942), is a bifunctional protein PCD/DCoH for transcription activation and metabolic enzyme [69]; phhB (PRK00823, a member of the superfamily cl00942) is a validated PCD [70], while PhhB (accession: COG2154, a member of the superfamily cl00942) is for coenzyme transport and metabolism [70]. CpsD has been added to PCD/DCoH and pterin-4α super family.

As reported previously [69], four of the conserved substrates (chemical) binding sites within the homo-tetramer are located adjacent to a saddle motif. The aromatic arch is a continuous arch of aromatic residues within the DCoH dimer. It had also been shown [41,56] that the hydrophobic aromatic arch connects the DCoH dimer interface from one active site to the other in accordance with the TATA box-binding protein. The conserved binding and interaction sites are interwoven for direct interactions probably to create an enzyme-substrates interface for probable regulation, interaction and catalysis.

3.8. Sequence homology comparison and structural modelling of CpsD

In quest for genuine sequence homologues and comprehensive structural characterization of CpsD, the protein was compared to the available structures in the protein data bank (PDB) via the Dali server [71]. The Dali search indicated that CpsD is most similar to PCD/DCoH (PDB: 2EBB) from *Geobacillus kaustophilus* strain HTA426 [72] with a *Z*-score of 21.8 and sequence identity of 59%, followed by PCD/DCoH

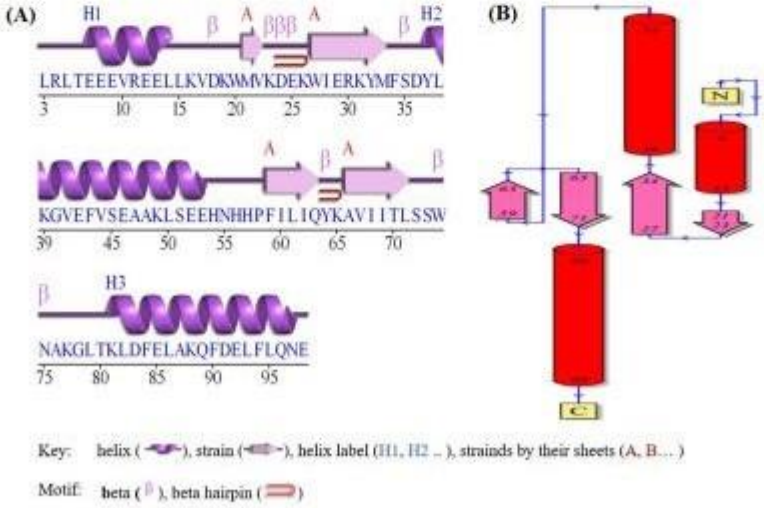


Fig. 6. The predicted secondary and 3D structural framework of CpsD. (A) Secondary structural elements of CpsD drawn using PDBsum indicating βαεαβαεαβ motif, 1 β-sheet, 2 β-hairpins, 1 β-bulges, 4 strands, 3 α-helices, 3 helix-helix interacts and 8 β-turns. (B) Topology of CpsD obtained using PDBsum showing the motif.

(PDB: 3JST) from *Brucella melitensis*, and PCD/DCoH (PDB: 1F93; 1USO and 4IOW) with Z-scores of 17.2, 16.8, 16.4, 13.8 and 12.3, and sequences identities of 32%, 27%, 33%, 25% and 26%, respectively (Fig. S4d).

Models were built on the Swiss-Model server [32], using PDB entry 2EBB as a template (Fig. S4e). The CpsD structural model summary, including the secondary structure, model validation, interactions and other structural properties of the protein model (Fig. S5a–h) were evaluated using the PDBsum database [34]. The Ramachandran plot statistics of the CpsD structure indicated that 91.2% of CpsD residues were at the most favoured regions, with 5.5% additional allowed regions, 2.2% generously allowed regions and only 1.1% disallowed regions (Fig. S5a). The predicted CpsD secondary structure comprises of 96 residues of 1 β-sheets, 2 β-hairpins, 1 β-bulges, 4 strands, 3 α-helixes, 3 α-helixes-helix interacts and 8 β-turns (Fig. 6A, B).

In addition, a 3D superimposition of PDB entries 2EBB, 3JST, 1F93, 2V6U and CpsD sequences showed conserved sequences and structures in the selected proteins (Fig. 7A, B). The 3D graphical structures and models of CpsD structure generated at the PDBsum database by submitting CpsD PDB file provided significant insight into its structure (Fig. 7C). The model predicted the protein to be a homodimer (Fig. 7C) but in vitro confirmed CpsD to be a monomer and multimer based on the 12% SDS-PAGE of the pure protein (Fig. 1C, D). There were 10 clefts found in the CpsD structure (Fig. 7D, E, F), most of which were large enough to enhance the effectiveness of the protein catalytic function. The clefts were buried vertices with an average depth between 11.36 and 4.98 and were accessible (Fig. S6).

3.9. Residue interactions and hydrogen bonds networks

Docking of atoms in contact with each other with their bond lengths revealed a total of 4 H-bonds and 67 non-bonded contacts (Fig. 8). The H-bonds occurred between residues 60 (Ile) and 62 (Ile) of chains A and B with bond distances ranging from 2.73 to 3.20. The number of H-bond lines between any two residues indicates the number of potential hydrogen bonds between them. For nonbonded contacts, which can be plentiful (Fig. S7a–b), the width of the striped line is proportional to the number of atomic contacts.

3.10. The active site of CpsD

Unlike other proteins, CpsD does not have any metal-binding ion centre. Previous studies on human and rat PCD 3D structure revealed the possible active site where dehydration activity occurs. The conserved residues that were believed to be involved in catalyses were His61, His62, Pro63 and His79 [44]. Other charged residues that contributed to the active hydroxylation from the neighbouring subunits include Glu57, Asp88, Arg87 and Tyr69 [44]. The locations of these residues were further verified [73]

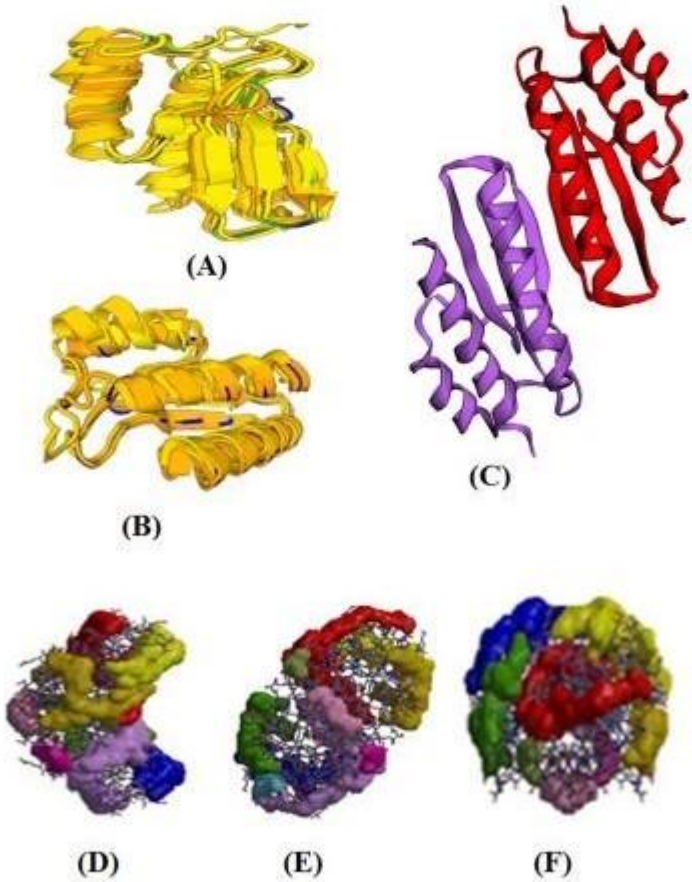


Fig. 7. Graphics of the 3D structure (PDBSum) and clefts of CpsD (MOLE 2.0 program v2.5.13.11.08 and visualized on pymol 0.97rc) found in CpsD structure and 3D superimposition of 2EBB, 3JST, 1F93, 2V6U and CpsD showing: (A) the conserved and (B) conserved structures; (C) 3D structure; and (D, E, F) cleft of CpsD.

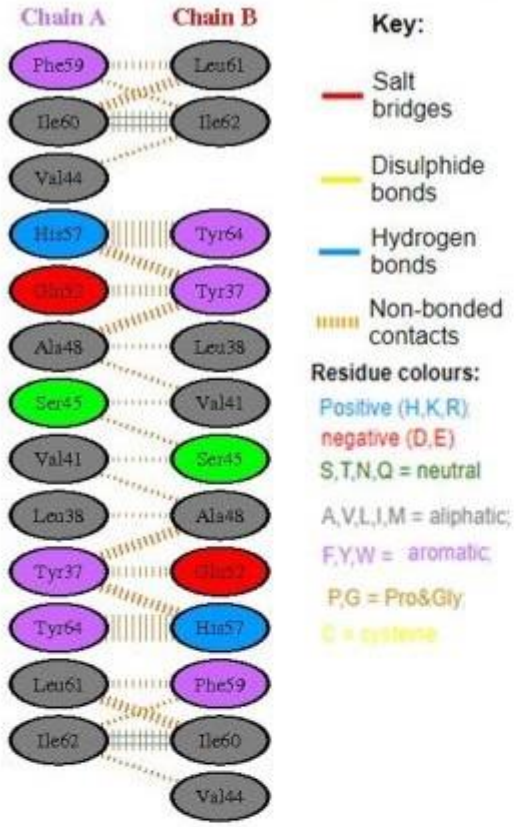


Fig. 8. Graphics of residues interactions across interface coloured by residue type (MOLE 2.0 program v2.5.13.11.08 and visualized on pymol 0.97rc) found in CpsD structure. The number of H-bond lines between any two residues indicates the number of potential hydrogen bonds between them. For non-bonded contacts, which can be plentiful, the width of the striped line is proportional to the number of atomic contacts. Residue colours: Positive (H, K, R); negative (D, E); S, T, N, Q = neutral; A, V, L, I, M = aliphatic; F, Y, W = aromatic; P, G = Pro & Gly; C = cysteine spacer.

and His61, His62 and His79 were found to be crucial for substrates binding and catalysis [74]. Furthermore, His62 specific role is primarily to bind substrate [52].

The multiple sequences alignments of mammalian PCD (human (P61457) and rat (P61459)), fungi (P38744) and their homolog (CpsD) showed that there are 15 conserved residues within the structures, whereas there are 26 conserved residues between the alignments of mammalian (P61457, P61459) PCD and CpsD, meaning that CpsD is more related to mammalian PCD/DCoH as compared to fungi. The conserved His61, His62, Pro63 residues in mammalian and fungal PCD/DCoH were also conserved in CpsD (His56, His57, Pro58). However, the conserved His79 in mammalian and fungi PCD/DCoH has been swapped with tryptophan (W74) in CpsD (Fig. 9).

Furthermore, analyses of the multiple structural alignments of bacterial PCD's indicated that residues E52, H56, H57, P58, D83, A87 (i.e. Glu52, His56, His57, Pro58, Asp83, and Ala87) were conserved in bacterial PCD (Fig. 10). The His79 is swapped in most of the bacterial PCD, in accordance with the previous report that the catalytic active sites of human and rat PCD were not conserved in bacteria [75]. In addition, protein-protein interaction between chain A and B of the theoretical CpsD 3D structural model (Fig. 8), showed that His57 of chain A and B formed salt bridges that connected Glu52, Ala48, Ser45, Val41, Leu38, Tyr37 and Try64. These residues could also form part of the CpsD active sites residues in accordance with earlier reports [39,74,76]. Since these residues are parts of conserved domains of the protein structure, this study proposed that the conserved residues (Glu52-His56-His57-Pro58-Asp83-Ala87) might be the catalytic active site of CpsD from *BcAOA* and might be responsible for the different roles that this protein plays in bacteria. Cluster analysis of CpsD structure and those of the other members of PCD/DCoH Superfamily (Fig. S8a) showed that only 2 residues (H57 and D83) were conserved in structure and mutational events or deletion of genes seems to have occurred in similar manners. However, multiple sequence alignments between CpsD and PCD/DCoH from other members of the Pterin-4α- superfamily (Fig. S8b) indicated more conserved residues among members compared to those of PCD/DCoH superfamily.

3.11. Evolutionary relationships of CpsD

The phylogenetic tree (Fig. 11A) showed that CpsD is in the same monophyletic node with PCD's from other *Bacillaceae*, human (H0YA5) and *Schizosaccharomyces pombe* (C3GP9B). The tree indicated that CpsD is not related to PCD's from other bacterial species or animals. *Bacillaceae* PCD may be of human origin based on the considerable conserved domain in their structures (Fig. 9) and their evolutionary relatedness to H0YA52 and C3GP98. Additionally, the phylogenetic analysis of the optimal tree between CpsD and other members of Pterin-4α- superfamily (Fig. 11B) showed that CpsD is evolutionarily related to PCD/

```
P61457      MAGKAHRLSAEERD--QLLPNLRAVGWNELEGRD------AIFKQFHEKDFNRAFGFMTR
P61459      MAGKAHRLSAEERD--QLLPNLRAVGWNELEGRD------AIFKQFHEKDFNRAFGFMTR
P38744      MHNKIVRIASSALTGGKLLEKLKPLTRWEVQWDPNKTKCLGITREVTFKDYETTWAFLTR
CpsD        ---MMLRLTEEEVR-----EELLKVDKWMVKDEK------WIERKYMFSDYLKGVEFVSE
                  *:: .          :*  :    ::           * ::  *.*:     *::.

P61457      VALQAEKLDHHPEWFNVYNKVHITLSTHECAG-------LSERDINLASFIEQVAVSMT-
P61459      VALQAEKLDHHPEWFNVYNKVHITLSTHECAG-------LSERDINLASFIEQVAVSMT-
P38744      VSMRSHLWGHHPLIHTSYTWYKLELHTHDIDPKDGAHSQLSDIDVRMAKRIDSYIDEMTT
CpsD        AAKLSEEHNHHPFILIQYKAVIITLSSWNAKG-------LTKLDFELAKQFDELFLQNEN
            .:  :.  .***     *. * : * : :          *:. *..:*. ::.   .

P61457      -----
P61459      -----
P38744      -----
CpsD        AIIRK
```

Fig. 9. Multiple sequences alignments of mammalian (human (P61457), *Rat* (P61459)) and fungi Fungi (O42658) PCD with their homolog (CpsD). Conserved residues were indicated in green and asterisks. Where CpsD sequence is similar to those of human (P61457) and rat (P61459) were indicated in yellow in addition to those shaded in green.

```
G0GXA2      -------------------------------------MTVLSEKRCVPCEGGVPSLEKKE  23
Q475Y4      ------------------------------------MSEKLESQTCTPCRGGIPPLERAE  24
A0A428VDB9  ------------------------------------MSELASQQCEACHADAPKVSEQE  23
P43335      ------------------------------------MTALTQAHCEACRADAPHVSDEE  23
A0A1R3FLM8  ---------------------------------------MLNEQKCEACSIDAIALSKDE  21
G7VH31      ------------------------------------------------------------  0
Q0BU27      MQRRRGSDGMDGTACMTHVPLSPGGGQDHARFSCACLTGGEANRREMPMTARETALAEAE  60
A0A094WQ23  ----------------------------------------------------MKKLNEQE  8
CpsD        ----------------------------------------------------MMLRLTEEE  9
A0A3E5DUB5  ----------------------------------------------------MMLRLTEEQ  9
C3GP98      ------------------------------------------------------------  0
A0A4P6KMX2  ---------------------------------------------------MARNRLTESE  10
A0A045J5V7  ----------------------------------------------------MAVLTDEQ  8
A0A231GVD6  ----------------------------------------------------MSTPLLSDSE  10

G0GXA2      IDKLL-VQLQSVWQVNA---LGPLYKKYKFSNFIK--AFANKINDIAEQEEHHPDLSISW  77
Q475Y4      AEALL-VETPG-WTLADDAG--RLERSFTFRNFAQALEFVSGVGRLAEEQGHHPEISFGW  80
A0A428VDB9  LSELM-HLIPD-WQPLVVKGELQLRREFTFRNFKEALAFTNRLGELAEAEFHHPAILTEW  81
P43335      LPVLL-RQIPD-WNIEVRDGIMQLEKVYLFKNFKHALAFTNAVGEISEAEGHHPGLLTEW  81
A0A1R3FLM8  QQSLL-LELSD-WQIMERDGIPQLEKVFKFKNYKQAWAFSNKVSELAEEEFHHPSILLEW  79
G7VH31      -------------MKRE----ECIVYKLRFKTYIDAVRFLTHLAEIAERHGHHPDVELKY  43
Q0BU27      IQQRLQDGLSH-WRYEN--G--WIRRTYRTASWKATLMVINTVGHLAEAAWHHPDITASY  115
A0A094WQ23  ITEQL-TLVPK-WKRKE--T-KWIERRYRFNEFLDGITFVNNVSTIAERENHHPFIQIQY  63
CpsD*       VREEL-LKVDK-WMVKD--E-KWIERKYMFSDYLKGVEFVSEAAKLSEEHNHHPFILIQY  64
A0A3E5DUB5  VQEEL-LKLDK-WMIKD--E-KWIERKYMFSDYLKGVEFVSEAAKLSEEHNHHPFILIQY  64
C3GP98      ----------------------------MFSDYLKGVEFVSEAAKLSEEHNHHPFILIQY  32
A0A4P6KMX2  MNEAL-RALDG-WQKVD--GREAITRSFKFKDFSTAFGFMAQAALYAEKLDHHPEWFNAY  66
A0A045J5V7  VDAAL-HDLNG-WQRAG--G--VLRRSIKFPTFMAGIDAVRRVAERAEEVNHHPDIDIRW  62
A0A231GVD6  IATGL-RELPA-WTHTG--S--SISRTVEAPSFLAGIELVRRVADSAEKVNHHPDIDIRW  64
                                          :           :*   ***     :

G0GXA2      GACSMEIWTNKID-GLMENDFILAAKIESKI----------------------------  107
Q475Y4      GHATVSWRTKKIK-GLHRNDFVMAAKTSELAAGMTQG----------------------  116
A0A428VDB9  GRVTVSWWTHKIG-GLHRNDFIMAARTDQLLSDPLHK----------------------  117
P43335      GKVTVTWWSHSIK-GLHRNDFIMAARTDEVAKTAEGRK----------------------  118
A0A1R3FLM8  GKVTVTWWSHSIK-GLHKNDFICASRCDVFAASE-------------------------  112
G7VH31      TNLVLRLTTHDAGNKITDRDLALAKEIDRLIEAHRDAISSAE-----------------  85
Q0BU27      AWVEVRLQTHTAK-GITEKDFALARMIEQVVQWQPGKDDGPLEGTPSGDQRFAYVKYDT  173
A0A094WQ23  KMVILSLSSWNEN-GLTDLDFQLAQEMDELFVEQTKTSNH-------------------  102
CpsD        KAVIITLSSWNAK-GLTKLDFELAKQFDELFLQNENAIIRK------------------  104
A0A3E5DUB5  KAVIITLSSWNAK-GLTKLDFELAKQFDELFVQNEKAIIRK------------------  104
C3GP98      KAVIITLSSWNAK-GLTKLDFELAKQFDELFVQNEKAVIRK------------------  72
A0A4P6KMX2  NRVDVILATHSEN-GVTELDIKMARKMNAIAG---------------------------  97
A0A045J5V7  RTVTFALVTHAVG-GITENDIAMAHDIDAMFGA--------------------------  94
A0A231GVD6  RRVTFTLSTHSEG-GITRKDLALAGQIDRLADQS-------------------------  97
                .   :      :   *:  *   .
```

Fig. 10. Multiple sequences alignment of CpsD with other bacteria PCD/DCoH. The secondary structures of CpsD and PCD from *Mycobacterium tuberculosis* (A0A045J5V7), *Caprionidus mecator* JMP 134 (Q475Y4), *P. aeruginosa* ATCC 15692 (P43335), *Nocardia cerradoenis* (A0A231GVD6), *Vibrio* sp. (A0A1R3FLM8), *Aeromonas salmonicida* (A0A428VDB9), *Granulibacter bethesdensis* ATCC BAA-1260 (Q0BU27), *Brucella melitensis* (A0A4P6KMX2), *B. alcalophilus* ATCC 27617 (A0A094WQ23), *Rickettsia heilongjiangensis* ATCC VR-1524 (G0GXA2), *B. cereus* (A0A3E5DUB5), *B. thuringiensis serovar pondicheriensis* BGSC 4BA1 (C3GP98), Bacillaceae (WP000979552), *B. toyonensis* (WP098203324), *B. cereus* group (WP000979535), *B. thuringiensis* (WP109737504), *B. anthracis* (WP098345270), *Bacillus cereus* (WP074625113), *Pyrobaculum ferrireducens* (G7VH31) were aligned and compared.

DCoH from *Prochlorococcus marinus* subsp. *marinus* str. CCMP1375 (AAP99535), and shared the same branch with PCD/DCoH from PHS_STRCO (O86722), *Nocardioides* sp. JS614 (ZP00658229), *Nocardia farcinica* IFM 10152 (BAD59613), *Hydrogenovibrio crunogenus* XCL-2 (ABB41227) and *Corynebacterium glutamicum* ATCC 13032 (BAB97857). Furthermore, the evolutionary relatedness of CpsD with other members of PCD/DCoH superfamily (Fig. 11C) showed that CpsD is closely related to PHS/DICDI (Q54RY8), *Pseudanabaena* sp. PCC 7367 (WP015165790), *Advenella kashmirensis* (WP014751602) and *Parvularcula bermudensis* (WP013299844).

3.12. The proposed role of CpsD in BcAOA PheOHS

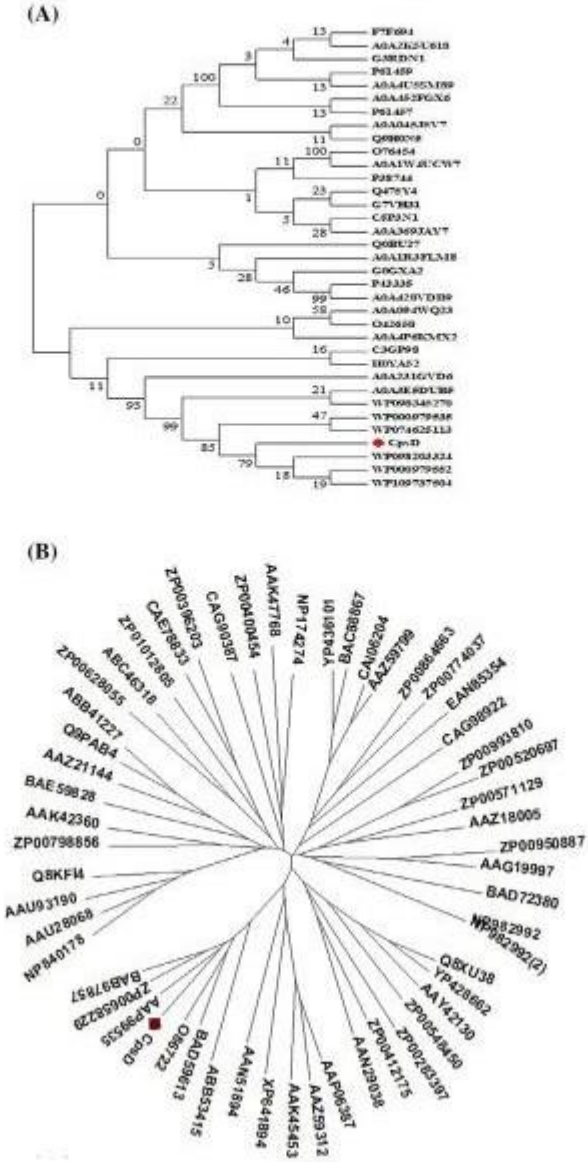
In human and other mammalian PheOHS; Phe4MO hydroxylates phenylalanine (phe) to tyrosine (tyr) and released the tyr from PheOHS while PCD/DCoH recycled the BH4 oxidised during the process [3]. Unlike humans and other higher animals, most bacteria do not produce BH4, but the primary role of PCD/DCoH in bacteria is to convert dihydroneopterin triphosphate (H2-NTP) to a coenzyme (folate) for one-carbon transfer reactions [7]. The proposed roles of CpsD (a homolog of PCD/DCoH) in BcAOA PheOHS may include regulation of the expression of Phe4MO and the entire activity of the PheOHS, in addition to working as Tet-CBQ reductase, being a homolog of an enzyme with many regulatory properties [77]. As a Tet-CBQ reductase, CpsD converts Tet-CBQ (benzoquinone) to Tet-CHQ (hydroquinone) and release the hydroquinone from PheOHS (Scheme 1).

CpsD may also block the release of benzoquinone (to prevent cytotoxic and cell damage/death) from the PheOHS after hydroxylation of PCP to Tet-CBQ (by Phe4MO), by recruiting Tet-CBQ into its own cycle as its main substrate (Scheme 1). CpsD may also down- or up-regulate the expression of Phe4MO (using a possible Yin-Yang mechanisms which maintained Tet-CBQ at a level that would neither significantly decrease the hydroxylation of PCP nor cause cytotoxicity in the organism cells) to control the amounts of PCP hydroxylated (by Phe4MO) and Tet-CBQ available in the PheOHS per time to prevent reverses flux (of Tet-CBQ ↔ PCP and Tet-CHQ ↔ PCP) otherwise known as a redox/

futile cycling (see blue path of Scheme 1) in the PheOHS. CpsD may also prevent reverses flux of Tet-CHQ to Tet-CBQ by releasing the Tet-CHQ out of PheOHS into an autologous ligand such as glutathione to form a conjugate for further detoxification by other enzymes in the pathway. The reverse fluxes observed in *S. chlorophenolicum* PCP degradation pathway [18,19] may have been well managed in *BcADA* PheOHS. This study, therefore, proposed that CpsD play a catalytic and regulatory role in *BcADA*.

3.13. Proposed mechanism of Tet-CBQ catalysis by CpsD

Based on the putative conserved domains, multiple catalytic sites, arrangements of the active sites on the CpsD structure, coupled with the ability of the enzyme to perform multiple roles, the mechanisms of catalysis of the proposed catalytic role of CpsD (Tet-CBQ reductase) in *BcADA* PheOHS and PCP degradation pathway is currently not clear. However, since the protein is a homolog of NADH dependent quinone oxidoreductase and can exist in different oligomeric states to perform multiple tasks (6, 50, 66), its catalytic reaction mechanism as benzoquinone reductase can follow one of the three stochastic mechanisms. The proposed reaction catalysed by CpsD (a homolog of PCD/DCoH) may follow a Ping-Pong or atypical ternary and alternatively, the ternary mechanism of reaction, as the case of NAD(P)H:quinone oxidoreductase [8–10]. The enzyme may follow a Ping-Pong mechanism [78], in which NADH binds (to the enzyme), react (by transferring 2^{e^-} to the enzyme to become oxidised) and dissociate sequentially, either in the same binding site (a one site Ping-Pong mechanism) or in discrete binding sites (a two-site ping-pong mechanism). Tet-CBQ binds and become



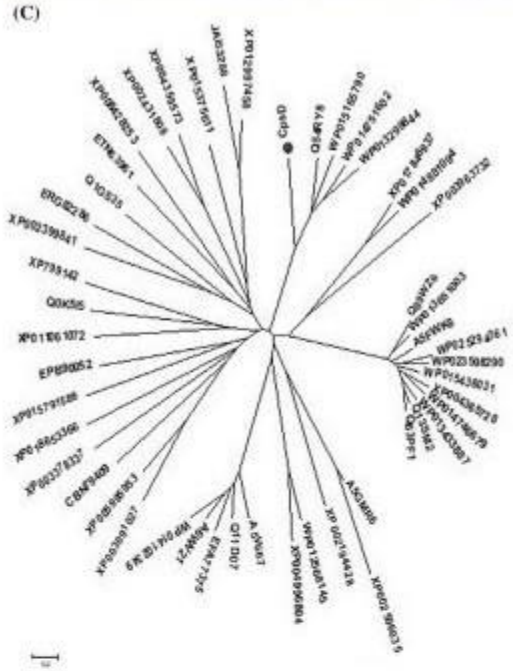
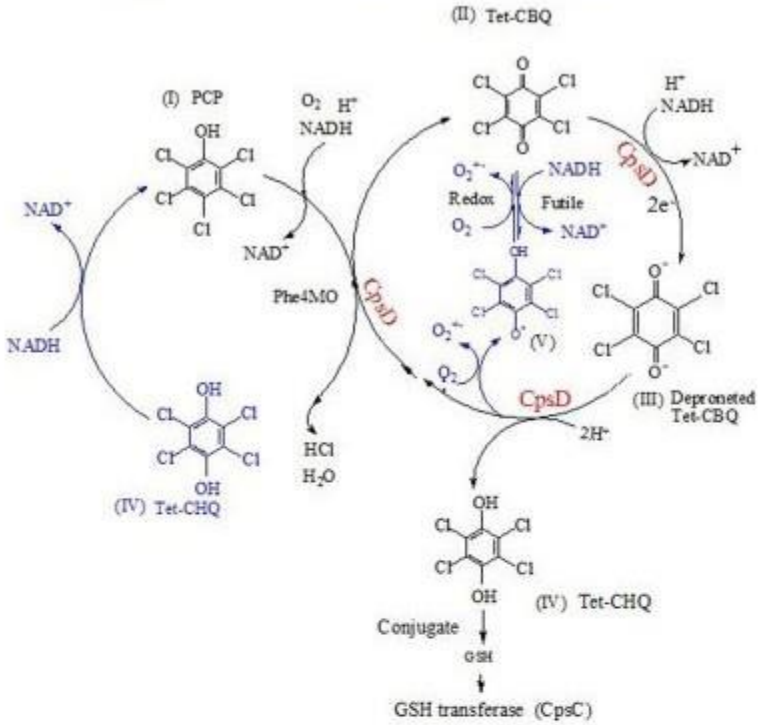


Fig. 11. (A) Evolutionary relatedness of CpsD to other PCD/DCoH. A phylogenetic tree was constructed using: CpsD and PCD/CHoH from human (P61457), (Q9H0N5) (H0YA52); animals: *Rattus norvegicus* (P61459), *Macaca mulatta* (F7F694), *Capra hircus* (A0A452FGX6), *Camelus dromedaries* (A0A4U5SM89), *M. fascicularis*, (A0A2K5U618), *Gorilla gorilla* (G3RDN1), *Drosophila melanogaster* (O76454), *D. ficusphila* (A0A1W4UCW7), fungi: *S. pombe* ATCC 24843 (O42658), *Saccharomyces cerevisiae* ATCC 204508 (P38744), *Coccidioides posadasii* C735 (C5P3N1), *Hypsizygus marmoreus* (A0A369JAY7) and bacteria: *M.tuberculosis*(A0A045J5V7), *C. necator* JMP 134 (Q47SY4), *P. aeruginosa* ATCC 15692 (P43335), *Nocardia cerradoensis* (A0A231GVD6), *Vibrio sp.* (A0A1R3FLM8), *Aeromonas salmonicida* (A0A428VDB9), *G. bethesdensis* ATCC BAA-1260 (Q0BUZ7), *Brucella melitensis* (A0A4P6KMX2), *B. alcalophilus* ATCC 27647 (A0A094WQ23), *Rickettsia heilongjiangensis* ATCC VR-1 524 (G0GXA2), *B. cereus* (A0A3E5DUB5), *B. thuringiensis* serovar *pondicheriensis* BGSC 4BA1 (C3GP98), *Bacillaceae* (WP000979552), *B. toyonensis* (WP098203324), *B. cereus* group (WP000979535), *B. thuringiensis* (WP109737504), *B. anthracis* (WP098345270), *B. cereus* (WP074625113) and *Pyrobaculum ferrireducens* (G7VH31). (B) Evolutionary relationship of CpsD with other members of PCD/DCoH superfamily. The optimal tree with the sum of branch length (50.56) is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved 55 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 88 positions in the final dataset. PCD/DCoH from other members of PCD_DCoH superfamily were from *Schistosoma japonicum* (AAP06387), *Peptoclostridium acidaminophilum* (AAN86540), PHS_CHLTE (Q8KF4), PHS_RALSO (Q8XU38), PHS_XYLFA (Q9PAB4), *Nitrosomonas europaea* ATCC 19718 (NP840178), *Arthrobacter* sp. FB24 (ZP00412175), *Janibacter* sp. HTCC2649 (ZP00993810), *Nocardia farcinica* IFM 10152 (BAD59613), *Frankia* sp. EAN1pec (ZP00571129), *Frankia* sp. Ccl3 (ZP00548450), *Nocardioides* sp. JS614 (ZP00658229), *Corynebacterium glutamicum* ATCC 13032 (BAB97857), PHS_STRCO (Q86722), *Leptospira interrogans* serovar *Lai* str. 56601 (AAN51894), *Aspergillus oryzae* RIB40 (BAE59828), *Mycobacterium tuberculosis* CDC1551 (AAK47768), *Nitrobacter hamburgensis* X14 (ZP00628055), *Arabidopsis thaliana* (AAM66965), *Dictyostelium discoideum* AX4 (XP641894), *Oryza sativa japonica* Group (BAD72380), *Arabidopsis thaliana* (NP174274), *Brucella suis* 1330 (AAN29038), *Candidatus Pelagibacter ubique* HTCC1062 (AAZ21144), *Hahella chejuensis* KCTC 2396 (YP436101), *Alkaliphilus metalliredigenes* QYMF (ZP00798856), *Hydrogenovibrio crunogenus* XCL-2 (ABB41227), *Anaeromyxobacter dehalogenans* 2CP-C (ZP00400454), *Alkalilimnicola ehrlichei* MLHE-1 (ZP00864663), *Maritimibacter alkaliphilus* HTCC2654 (ZP01012805), *Legionella pneumophila* subsp. *pneumophila* str. Philadelphia 1 (AAU28068), *Gloeobacter violaceus* PCC 7421 (BAC88867), *Aromatoleum aromaticum* EbN1 (CAI06204), *Methylococcus capsulatus* str. Bath (AAU93190), *Debaryomyces hansenii* CBS767 (CAG90387), *Cupriavidus pinatubonensis* JMP134 (AAZ59799), *Salinibacter ruber* DSM 13855 (ABC46318), *Trypanosoma cruzi* (EAN85354), *Solibacter usitatus* Ellin6076 (ZP00520697), *Prochlorococcus marinus* str. NATL2A (AAZ59312), *Eremothecium gossypii* ATCC 10895 (NP982992), *Kluyveromyces lactis* (CAG98922), *Pseudoalteromonas atlantica T6c* (ZP00774037), *Rhodothermus marinus* (AAY42130), *Prochlorococcus marinus* subsp. *marinus* str. CCMP1375 (AAP99535), *Psychrobacter arcticus* 273-4 (AAZ18005), *Croceibacter atlanticus* HTCC2559 (ZP00950887), *Bdellovibrio bacteriovorus* HD100 (CAE78833), *Plasmodium falciparum* (ABB53415), *Deinococcus geothermalis* DSM 11300 (ZP00396203), *Burkholderia fungorum* LB400 (ZP00283397), *Halobacterium salinarum* NRC-1 AAG19997), *Mycobacterium tuberculosis* CDC1551 (AAK45453), *Rhodospirillum rubrum* ATCC 11170 (YP428662), *Saccharolobus solfataricus* P2 (AAK42360). (C) Evolutionary relationships of CpsD with other members of Pterin-4α- superfamily. The optimal tree with the sum of branch length (38.26) is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The analysis involved 49 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 93 positions in the final dataset. Members of the Pterin-4α- were PHS_SYNPW (A5GM86), PHS_CHESB (Q11DX0), *Cupriavidus necator* H16 (Q0K585), PHS_PARXL (Q13SM2), PHS_BURPS (Q63PF1), PHS_DICDI (Q54RY8), PHS_BRADU (Q89WZ6), PHS_OCHA4 (A6WV21), Q1GS35_SPHAL (Q1GS35), PHS_SPHWW (A5V667), PHS_ACICJ (A5FWK8), PHS_ACICJ (A5FWK8), *Caenorhabditis remanei* (XP003091027), *Ixodes scapularis* (XP002399841), *Ascaris suum* (ERG82286), *Taeniopygia guttata* (XP002194428), *Capsaspora owczarzaki* ATCC 30864 (XP004365728), *Anopheles darling* (ETN63961), *Polymorphum gilvum* (WP013651003), *Poecilia reticulate* (XP008428253), *Trichinella spiralis* (XP003378337), *Ectocarpus siliculosus* (CBN79409), *Branchiostoma floridae* (XP002596035), *Achemella kashmirensis* (WP014751602), *Strongylocentrotus purpuratus* (XP799142), *Pediculus humanus corporis* (XP002431808), *Diuraphis noxia* (XP015375611), *Micavibrio aeruginosavorus* (WP014102349), *Paraburkholderia rhizoxinica* (WP013433887), *Drosophila busckii* (XP017846937), *Latimeria chalumnae* (XP005995953), *Pseudanabaena* sp. PCC 7367 (WP015165790), *Schistosoma mansoni* (XP018653366), *Tetranychus urticae* (XP015791586), *Acromyrmex echinatior* (XP011061072), *Mucor circinelloides f. circinelloides* 1006PhL (EPB90052), *Rhodnius neglectus* (JAI53288), *Parvularcula bermudensis* (WP013299844), *Azoarcus* sp. KH32C (WP015436031), *Takifugu rubripes* (XP003963732), *Cavia porcellus* (XP012997458), *Lonodonta africana* (XP010592093), *Tistrella mobilis* (WP014746679), *Rhodospirillum centenum* (WP012568145), *Phaeobacter inhibens* (WP014881094), *Sphingomonas saxaniigenens* (WP025294361), *Cavenderia fasciculata* (XP004359573), *Heterostelium album* PN500 (EFA77375), *Salpingoeca rosetta* (XP004996804), *Pandoraea* (WP023598290).

deprotonated by the reduced enzyme, then the reduced enzyme releases the product (Tet-CHQ) from the PheOHS and become oxidised to start a new reaction (Scheme 2).

In an atypical ternary mechanism, the cofactor (NADH) may first bind and react with the enzyme (in its oxidised form but the oxidised product (NAD^+) does not dissociate and the enzyme becomes



Scheme 1. Proposed phenylalanine hydroxylating system in *Bacillus cereus* strain AOA-CPS 1. CpsD shaded in red denotes its benzoquinone reductase activity steps. The reactions shaded in blue indicate the proposed abrogated steps by CpsD in PheOHs. The reactant (V) represents an intermediate semiquinone formed by reverses flux between (IV) and (II).

reduced); then, Tet-CBQ binds and become deprotonated (in a ternary complex with enzyme-product complex), the reduced enzymes may then release both products (NAD^+ and Tet-CHQ) and become oxidised to start a new reaction (Scheme 3). Alternatively, the reaction may follow a classical ternary complex formation mechanism where both substrates (NADH and Tet-CBQ) are bound by the enzyme in its oxidised state (NADH transfer its charges to the enzyme and the enzyme simultaneously deprotonate Tet-CBQ with the 2^{e^-} transfer to it), and products of both substrates (NAD^+ and Tet-CHQ) are released simultaneously (Scheme 4).

The atypical or the Ping-Pong mechanism is more plausible, if the enzyme follows the classical ternary mechanism, reverses flux would occur between Tet-CBQ and product (Tet-CHQ) as previously reported in *S. chlorophenolicum* PCP degradation pathway [16,18,79]. The Ping-Pong mechanism is more plausible, because in this mechanism, the substrate (Tet-CBQ) and the cofactor (NADH) never bind together, preventing the non-enzymatic reduction of Tet-CBQ to Tet-CHQ in the presence of excess NADH (which may arise in the classical pathway) in the PheOHS, and the enzymatically catalysed reduction of Tet-CBQ to Tet-CHQ can be quantified stoichiometrically. The Ping-Pong mechanism may also prevent unproductive consumption of NADH and block the futile cycle, contrary to *S. chlorophenolicum* PCP degradation pathway [16,18,79].

4. Conclusion

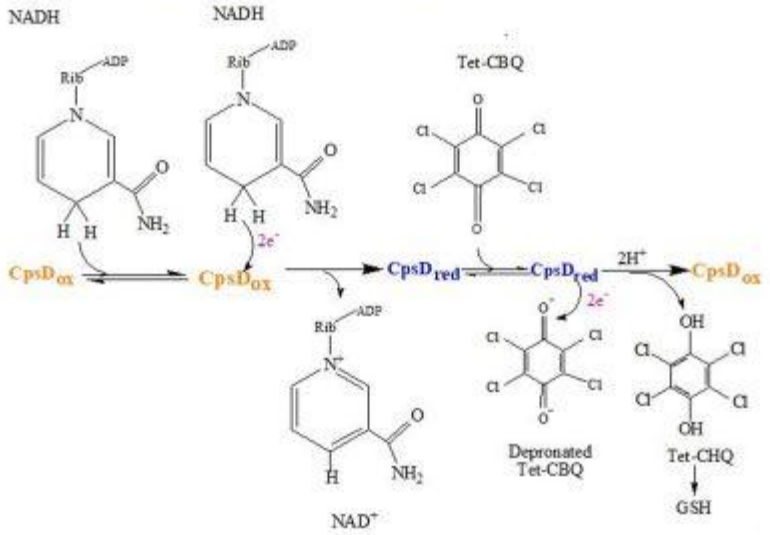
In conclusion, this study reports that CpsD may have a bifunctional catalytic and regulatory role in *Bc*AOA PheOHS. The theoretical 3D structure showed that the protein might occur in a dimeric form but in vivo expression of the enzyme in chemically competent cell showed that the purified protein can exist as a monomer and multi-tetramer. Based on the results and other reports on the multifunctional role of PCD/DCoH in mammalian, fungi and other bacterial species, this study proposed that CpsD act as a benzoquinone reductase and catalysed the reduction of Tet-CBQ to Tet-CHQ using a possible Ping-Pong or atypical ternary mechanism. It is also proposed that CpsD regulates the expression of Phe4MO and its hydroxylated product (Tet-CBQ) while coordinating the release of Tet-CHQ from the PheOHS (to prevent possible reverses flux in the PheOHS), using a possible Yin-Yang mechanism, to ensure that the amount of PCP hydroxylated is stoichiometrically equal to the amount of Tet-CBQ that it can process. However, to fully understand the concise mechanism of reduction of Tet-CBQ to Tet-CHQ by CpsD and the regulatory role of CpsD in *Bc*AOA PheOHS, further biological and crystallographic studies are required.

Acknowledgments

O.A. thanks Tertiary Education Fund (TETFund) Nigeria and Obafemi Awolowo University for the scholarship and study leave. Author also thanks Dr Mare Vlok of the central analytical facility, Stellenbosch, South Africa for the Liquid-Chromatography Mass-Spectrometry proteomics analysis of the protein.

Author contributions

O.A. and A.O. conceived and designed the project; O.A. and A.K. designed the experiments; O.A. performed the experiments; M.P. contributed reagents and materials; O.A., A.K., M.P. and A.O. wrote the manuscript; all the authors have read and approved the manuscript.



Scheme 2. The proposed Ping-Pong mechanisms of catalyses by CpsD in *Bacillus cereus* strain. The reaction occurred sequentially, the substrate (Tet-CBQ) and co-factor (NADH) were not bound together, NADH first bind, react and its oxidised product released, then Tet-CBQ bound, deprotonated and released.

Funding

National Research Foundation, South Africa (Grant No: 94036 and 92803).

Declaration of competing interest

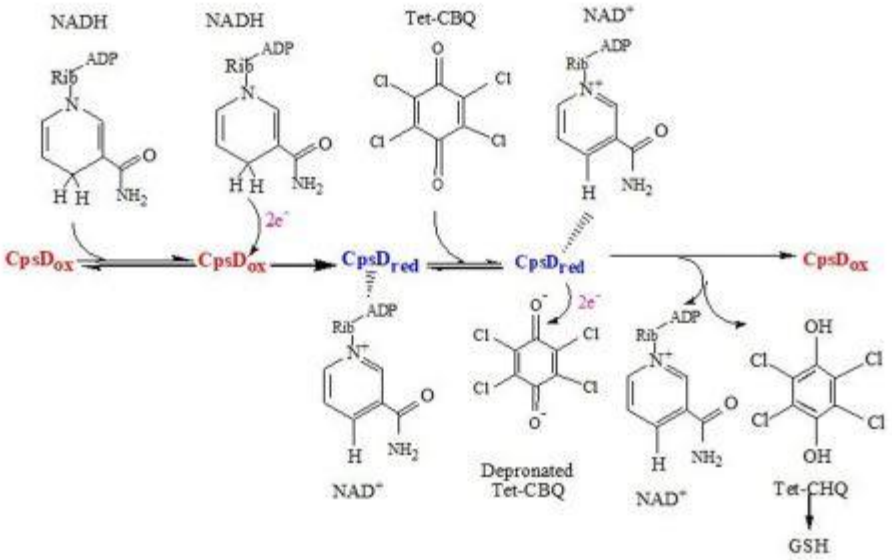
The authors declare that they have no conflicts of interest with the contents of this article.

Ethical statement

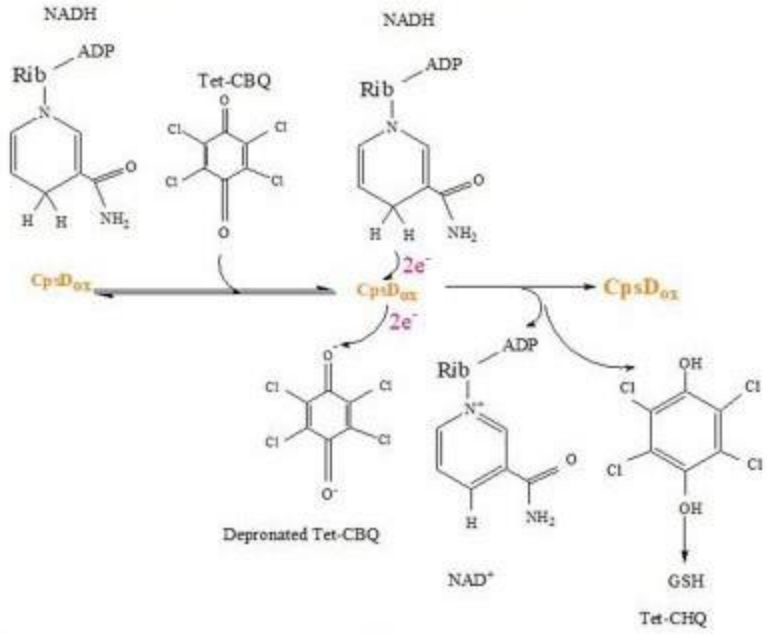
This article does not contain any studies with human participants or animals performed by any of the authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2020.06.083.



Scheme 3. The proposed atypical ternary mechanism of CpsD catalyses. The co-factor (NADH) bound, react and remain bounded, Tet-CBQ react with the complex in a substrate-complex ternary and both products released simultaneously.



Scheme 4. The proposed ternary mechanism of CpsD catalyses. The substrate (Tet-CBQ) and the co-factor (NADH) bound, NADH transfer 2^{e^-} to CpsD, the enzyme deprotonated Tet-CBQ with the 2^{e^-} and the products of both substrates were released simultaneously.

References

[1] R. Teufel, V. Mascaraque, W. Ismail, M. Voss, J. Perera, W. Eisenreich, W. Haehnel, G. Fuchs, Bacterial phenylalanine and phenylacetate catabolic pathway revealed, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 14390–14395, https://doi.org/10.1073/pnas.1005399107.

[2] G. Fuchs, M. Boll, J. Heider, Microbial degradation of aromatic compounds - from one strategy to four, Nat. Rev. Microbiol. 9 (2011) 803–816, https://doi.org/10.1038/nrmicro2652.

[3] G. Litwack, Metabolism of amino acids, Hum. Biochem. Elsevier 2018, pp. 359–394, https://doi.org/10.1016/b978-0-12-383864-3.00013-2.

[4] H. Wang, H. Chen, G. Hao, B. Yang, Y. Feng, Y. Wang, L. Feng, J. Zhao, Y. Song, H. Zhang, Y.Q. Chen, L. Wang, W. Chen, Role of the phenylalanine-hydroxylating system in aromatic substance degradation and lipid metabolism in the oleaginous fungus *Mortierella alpina*, Appl. Environ. Microbiol. 79 (2013) 3225–3233, https://doi.org/10.1128/AEM.00238-13.

[5] Z. Ding, C.O. Harding, A. Rebuffat, L. Elzouk, J.A. Wolff, B. Thöny, Correction of murine PKU following AAV-mediated intramuscular expression of a complete phenylalanine hydroxylating system, Mol. Ther. 16 (2008) 673–681, https://doi.org/10.1038/mt.2008.17.

[6] D. Wang, M.W. Coco, R.B. Rose, Interactions with the bifunctional interface of the transcriptional coactivator DCoH1 are kinetically regulated, J. Biol. Chem. 290 (2015) 4319–4329, https://doi.org/10.1074/jbc.M114.616870.

[7] H. Wang, B. Yang, G. Hao, Y. Feng, H. Chen, L. Feng, J. Zhao, H. Zhang, Y.Q. Chen, L. Wang, W. Chen, Biochemical characterization of the tetrahydrobiopterin synthesis pathway in the oleaginous fungus *Mortierella alpina*, Microbiology 157 (2011) 3059–3070, https://doi.org/10.1099/mic.0.051847-0.

[8] K. Zhang, D. Chen, K. Ma, X. Wu, H. Hao, S. Jiang, NAD(P)H:quinone oxidoreductase 1 (NQO1) as a therapeutic and diagnostic target in cancer, J. Med. Chem. 61 (2018) 6983–7003, https://doi.org/10.1021/acs.jmedchem.8b00124.

[9] C. Glorieux, P.B. Calderon, Cancer cell sensitivity to redox-cycling quinones is influenced by NAD(P)H:quinone oxidoreductase 1 polymorphism, Antioxidants 8 (2019) 1–10, https://doi.org/10.3390/antiox8090369.

[10] J.N. Blaza, H.R. Bridges, D. Aragão, E.A. Dunn, A. Heikal, G.M. Cook, Y. Nakatani, J. Hirst, The mechanism of catalysis by type-II NADH:quinone oxidoreductases, Nat. Publ. Gr. 7 (2016) 1–11, https://doi.org/10.1038/srep40165.

[11] A.K. Jaiswal, Characterization and partial purification of microsomal NAD(P)H:quinone oxidoreductases, Arch. Biochem. Biophys. 375 (1) (2000) 62–68, https://doi.org/10.1006/abbi.1999.1650.

[12] L. Chen, J. Yang, Biochemical characterization of the tetrachlorobenzoquinone reductase involved in the biodegradation of pentachlorophenol, Int. J. Mol. Sci. 9 (2008) 198–212, https://doi.org/10.3390/ijms9030198.

[13] D. Xu, L. Hu, X. Xia, J. Song, L. Li, E. Song, Y. Song, Tetrachlorobenzoquinone induces acute liver injury, up-regulates HO-1 and NQO1 expression in mice model: the protective role of chlorogenic acid, Environ. Toxicol. Pharmacol. 37 (2014) 1212–1220, https://doi.org/10.1016/j.etap.2014.04.022.

[14] C. Su, P. Zhang, X. Song, Q. Shi, J. Fu, X. Xia, H. Bai, L. Hu, D. Xu, E. Song, Y. Song, Tetrachlorobenzoquinone activates NRF2 signaling by keap1 cross-linking and ubiquitin translocation but not keap1-cullin3 complex dissociation, Chem. Res. Toxicol. 28 (2015) 765–774, https://doi.org/10.1021/tx500513v.

[15] B. Ling, B. Gao, J. Yang, Evaluating the effects of tetrachloro-1,4-benzoquinone, an active metabolite of pentachlorophenol, on the growth of human breast cancer cells, J. Toxicol. 2016 (2016), 8253726, https://doi.org/10.1155/2016/8253726.

[16] J. Rudolph, A.H. Erbse, L.S. Behlen, S.D. Copley, A radical intermediate in the conversion of pentachlorophenol to tetrachlorohydroquinone by *Sphingobium chlorophenolicum*, Biochemistry 53 (2014) 6539–6549, https://doi.org/10.1021/bi5010427.

[17] D.L. Fraser, Evaluation of the Neurotoxicity of Pentachlorophenol and Its Active Metabolites on SH-SY5Y Neuroblastoma Cells, University of Pretoria, 2017.

[18] K. Hlouchova, J. Rudolph, J.M. Pietari, L.S. Behlen, S.D. Copley, Pentachlorophenol hydroxylase, a poorly functioning enzyme required for degradation of pentachlorophenol by *Sphingobium chlorophenolicum*, Biochemistry 51 (2012) 3848–3860, https://doi.org/10.1021/bi300261p.

[19] Yunyou Su, Lifeng Chen, Brian Bandy, Jian Yang, The catalytic product of pentachlorophenol 4-monooxygenase is tetra-chlorohydroquinone rather than tetrachlorobenzoquinone, Open Microbiol. J. 2 (2008) 100, https://doi.org/10.2174/1874285800802010100.

[20] S.C. Mitchell, G.R. Steventon, Phenylalanine 4-monooxygenase: the "sulfoxidation polymorphism", Xenobiotica 50 (1) (2019) 51–63, https://doi.org/10.1080/00498254.2019.1636419.

[21] G. Anja, S. Jeanette E, Voss De, J. James J, C.J. Max, Unrivalled diversity: the many roles and reactions of bacterial cytochromes P450 in secondary metabolism, Nat. Prod. Rep. 35 (2018) 757–791, https://doi.org/10.1039/c7np00063d.

[22] G.R. Steventon, S.C. Mitchell, Phenylalanine 4-monooxygenase and the role of endobiotic metabolism enzymes in xenobiotic biotransformation, Expert Opin. Drug Metab. Toxicol. 5 (2009) 1213–1221, https://doi.org/10.1517/17425250903179318.

[23] G.R. Steventon, S. Khan, S.C. Mitchell, Comparison of the sulfur-oxygenation of cysteine and S-carboxymethyl-L-cysteine in human hepatic cytosol and the role of cysteine dioxygenase, J. Pharm. Pharmacol. 70 (2018) 1069–1077, https://doi.org/10.1111/jphp.12944.

[24] A.Y. Chang, V.W.Y. Chau, J.A. Landas, Y. Pang, Preparation of calcium competent *Escherichia coli* and heat-shock transformation, JEMI Methods 1 (2017) 22–25.

[25] J.R. Marchesi, T. Sato, A.J. Weightman, T.A. Martin, J.C. Fry, S.J. Hiorn, W.G. Wade, Design and evaluation of useful bacterium-specific pcr primers that amplify genes coding for bacterial 16S rRNA, Appl. Environ. Microbiol. 64 (1998) 795–799.

[26] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T.L. Madden, BLAST+: architecture and applications, BMC Bioinformatics 10 (2009) 421, https://doi.org/10.1186/1471-2105-10-421.

[27] J. Sambrook, D. Russell, Molecular Cloning: A Laboratory Manual, 3rd, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001.

[28] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Biochem. 72 (1976) 248–254, https://doi.org/10.1016/0003-2697(76)90527-3.

[29] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature 227 (1970) 680–685.

[30] S. Liu, T. Su, C. Zhang, W. Zhang, D. Zhu, J. Su, T. Wei, K. Wang, Y. Huang, L. Guo, S. Xu, N. Zhou, L. Gu, Crystal structure of PnpCD, a two-subunit hydroquinone 1, 2-dioxygenase, reveals a novel structural class of Fe^{2+}-dependent dioxygenases, J. Biol. Chem. 290 (2015) 24547–24560, https://doi.org/10.1074/jbc.M115.673558.

[31] B. Setlhare, A. Kumar, M.P. Moloena, B. Pilley, A.O. Olaniran, Phenol hydroxylase from *Pseudomonas* sp. KZNSA: purification, characterization and prediction of three-dimensional structure, Int. J. Biol. Macromol. 146 (2019) 1000–1008, https://doi.org/10.1016/j.ijbiomac.2019.09.224.

[32] A. Waterhouse, S. Bienert, T.A.P. de Beer, G. Tauriello, G. Studer, L. Bordoli, T. Schwede, SWISS-MODEL: homology modelling of protein structures and complexes, Nucleic Acids Res. 46 (W1) (2018) W296–W303.

[33] S. Bienert, A. Waterhouse, T.A.P. de Beer, G. Tauriello, G. Studer, L. Bordoli, T. Schwede, The SWISS-MODEL Repository—new features and functionality, Nucleic Acids Res. 45 (2017) D313–D319, https://doi.org/10.1093/nar/gkw1132.

[34] R.A. Laskowski, J. Jabłońska, L. Pravda, R.S. Vařeková, J.M. Thornton, PDBsum: structural summaries of PDB entries, Prot. Sci. 27 (2018) 129–134.

[35] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425.

[36] J. Felsenstein, Confidence limits on phylogenies: an approach using the bootstrap, Evolution (N. Y.) 39 (1985) 783–791.

[37] E. Zuckerkandl, L. Pauling, Evolutionary divergence and convergence in proteins, in: V. Bryson, H.J. Vogel (Eds.), Evolving Genes and Proteins, Academic Press, New York 1965, pp. 97–166.

[38] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms, Mol. Biol. Evol. 35 (2018) 1547–1549, https://doi.org/10.1093/molbev/msy096.

[39] V. Naponelli, A. Noiriel, M.J. Ziemak, S.M. Beverley, L.F. Lye, A.M. Plume, J.R. Botella, K. Loizeau, S. Ravanel, F. Rébeillé, V. De Crécy-Lagard, A.D. Hanson, Phylogenomic and functional analysis of pterin-4α-carbinolamine dehydratase family (COG2154) proteins in plants and microorganisms, Plant Physiol. 146 (2008) 1515–1527, https://doi.org/10.1104/pp.107.114090.

[40] E. Fredrik, X-ray Characterization of PaPheOH, a Bacterial Phenylalanine Hydroxylase, Ph.D. thesis Umea University, 2003.

[41] D. Kowlessur, S. Kaufman, Cloning and expression of recombinant human pineal tryptophan hydroxylase in *Escherichia coli*: purification and characterization of the cloned enzyme, Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol. 1434 (1999) 317–330, https://doi.org/10.1016/S0167-4838(99)00184-3.

[42] X.-D. Lei, S. Kaufman, Characterization of expression of the gene for human pterin carbinolamine dehydratase/dimerization cofactor of HNF1, DNA Cell Biol. 18 (1999) 243–252, https://doi.org/10.1089/104454999315466.

[43] B.A. Citron, M.D. Davis, S. Milstien, J. Gutierrez, D.B. Mendel, G.R. Crabtree, S. Kaufman, Identity of 4α-carbinolamine dehydratase, a component of the phenylalanine hydroxylation system, and DCoH, a transactivator of homeodomain proteins, Proc. Natl. Acad. Sci. U. S. A. 89 (1992) 11891–11894, https://doi.org/10.1073/pnas.89.24.11891.

[44] C.R. Hauer, I. Rebrin, B. Thony, F. Neuheiser, H.C. Curtius, P. Hunziker, N. Blau, S. Ghisla, C.W. Heizmann, Phenylalanine hydroxylase-stimulating protein/pterin-4α-carbinolamine dehydratase from rat and human liver. Purification, characterization, and complete amino acid sequence, J. Biol. Chem. 268 (1993) 4828–4831.

[45] B.A. Citron, S. Kaufman, S. Milstien, E.W. Naylor, C.L. Greene, M.D. Davis, Mutation in the 4α-carbinolamine dehydratase gene leads to mild hyperphenylalaninemia with defective cofactor metabolism, Am. J. Hum. Genet. 53 (1993) 768–774.

[46] G. Zhao, T. Xia, J. Song, R.A. Jensen, *Pseudomonas aeruginosa* possesses homologues of mammalian phenylalanine hydroxylase and 4α-carbinolamine dehydratase/DCoH as part of a three-component gene cluster, Proc. Natl. Acad. Sci. U. S. A. 91 (4) (1994) 1366–1370.

[47] B. Thöny, F. Neuheiser, N. Blau, C.W. Heizmann, Characterization of the human PCBD gene encoding the bifunctional protein pterin-4α-carbinolamine dehydratase/dimerization cofactor for the transcription factor HNF-1α, Biochem. Biophys. Res. Commun. 210 (1995) 966–973, https://doi.org/10.1006/bbrc.1995.1751.

[48] R. Ficner, U.H. Sauer, G. Stier, D. Suck, Three-dimensional structure of the bifunctional protein PCD/DCoH, a cytoplasmic enzyme interacting with transcription factor HNF1, EMBO J. 14 (1995) 2034–2042, https://doi.org/10.1002/j.1460-2075.1995.tb07195.x.

[49] I. Rebrin, S.W. Bailey, S.R. Beerth, M.D. Ardell, J.E. Ayling, Catalytic characterization of 4α-hydroxytetrahydropterin dehydratase, Biochemistry 34 (1995) 5801–5810, https://doi.org/10.1021/bi00017a011.

[50] S. Cameron, S.A. Fyffe, S. Goldie, W.N. Hunter, Crystal structures of *Toxoplasma gondii* pterin-4α-carbinolamine dehydratase and comparisons with mammalian and parasite orthologues, Mol. Biochem. Parasitol. 158 (2008) 131–138, https://doi.org/10.1016/j.molbiopara.2007.12.002.

[51] I. Yadid, J. Rudolph, K. Hlouchova, S.D. Copley, Sequestration of a highly reactive intermediate in an evolving pathway for degradation of pentachlorophenol, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) E2182–E2190, https://doi.org/10.1073/pnas.1214052110.

[52] I. Rebrin, B. Thöny, S.W. Bailey, J.E. Ayling, Stereospecificity and catalytic function of histidine residues in 4a-hydroxytetrahydropterin dehydratase/DCoH, Biochemistry 37 (32) (1998) 11246–11254, https://doi.org/10.1021/bi980663h.

[53] K. Kim, Characterization of 1,4-benzoquinone reductase from bovine liver, Biotechnol Bioproc E 7 (2002) 216, https://doi.org/10.1007/BF02932973.

[54] R. Hara, K. Kino, Enhanced synthesis of 5-Hydroxy-L-tryptophan through tetrahydropterin regeneration, AMB Express 3 (2013) 1–7, https://doi.org/10.1186/2191-0855-3-70.

[55] N. Pedrini, A. Ortiz-Urquiza, C. Huarte-Bonnet, Y. Fan, M.P. Juárez, N.O. Keyhani, Tenebrionid secretions and a fungal benzoquinone oxidoreductase form competing components of an arms race between a host and pathogen, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) E3651–E3660, https://doi.org/10.1073/pnas.1504552112.

[56] C.A. Bellin, G.A. O'Connor, Sorption and degradation of pentachlorophenol in sludge-amended soils, J. Env. Qual. 19 (1990) 603–608.

[57] P. Van Aken, N. Lambert, R. Van Den Broeck, J. Degrève, R. Dewil, Advances in ozonation and biodegradation processes to enhance chlorophenol abatement in multisubstrate wastewaters: a review, Environ. Sci. Water Res. Technol. 5 (2019) 444, https://doi.org/10.1039/c8ew00562a.

[58] U.W. Sauer, G. Stier, D. Suck, Human pterin-4α-carbinolamine dehydratase/dimerization cofactor of hepatocyte nuclear factor-1α: characterization and kinetic analysis of wild-type and mutant enzymes, Eur. J. Biochem. 231 (2) (1995) 414–423.

[59] F.J. van Spronsen, A.M. van Wegberg, K. Ahring, A. Bélanger-Quintana, N. Blau, A.M. Bosch, A. Burlina, J. Campistol, F. Feillet, M. Gizewska, S.C. Huijbregts, S. Kearney, V. Leuzzi, F. Maillot, A.C. Muntau, F.X. Trefz, M. van Rijn, J.H. Walter, A. MacDonald, Key European guidelines for the diagnosis and management of patients with phenylketonuria, Lancet Diabetes Endocrinol. 5 (2017) 743–756, https://doi.org/10.1016/S2213-8587(16)30320-5.

[60] J.C. Lewis, P.S. Coelho, F.H. Arnold, Enzymatic functionalization of carbon-hydrogen bonds, Chem. Soc. Rev. 40 (2011) 2003–2021, https://doi.org/10.1039/c0cs00067a.

[61] D. Buongiorno, G.D. Straganz, Structure and function of atypically coordinated enzymatic mononuclear non-heme-Fe(II) centers, Coord. Chem. Rev. 257 (2013) 541–563, https://doi.org/10.1016/j.ccr.2012.04.028.

[62] C.M. Seibert, F.M. Raushel, Structural and catalytic diversity within the amidohydrolase superfamily, Biochemistry 44 (2005) 6383–6391, https://doi.org/10.1021/bi047326v.

[63] G. Johnen, S. Kaufman, Studies on the enzymatic and transcriptional activity of the dimerization cofactor for hepatocyte nuclear factor 1, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 13469–13474, https://doi.org/10.1073/pnas.94.25.13469.

[64] S. Koster, G. Stier, N. Kubasch, H.C. Curtius, S. Ghisla, Pterin-4α-carbinolamine dehydratase from *Pseudomonas aeruginosa*: characterization, catalytic mechanism and comparison to the human enzyme, Biol. Chem. 379 (1998) 1427–1432, https://doi.org/10.1515/bchm.1998.379.12.1427.

[65] H. Rho, C.N. Jones, R.B. Rose, Kinetic stability may determine the interaction dynamics of the bifunctional protein DCoH1, the dimerization cofactor of the transcription factor HNF-1α, Biochemistry 49 (2010) 10187–10197, https://doi.org/10.1021/bi1015056.

[66] A. Marchler-Bauer, Y. Bo, L. Han, J. He, C.J. Lanczycki, S. Lu, F. Chitsaz, M.K. Derbyshire, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, F. Lu, G.H. Marchler, J.S. Song, N. Thanki, Z. Wang, R.A. Yamashita, D. Zhang, C. Zheng, L.Y. Geer, S.H. Bryant, CDD/SPARCLE: functional classification of proteins via subfamily domain architectures, Nucleic Acids Res. 45 (2017) D200–D203, https://doi.org/10.1093/nar/gkw1129.

[67] S. Lu, J. Wang, F. Chitsaz, M.K. Derbyshire, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, G.H. Marchler, J.S. Song, N. Thanki, R.A. Yamashita, M. Yang, D. Zhang, C. Zheng, C.J. Lanczycki, A. Marchler-Bauer, CDD/SPARCLE: the conserved domain database in 2020, Nucleic Acids Res. 48 (D1) (2020) 265–268, https://doi.org/10.1093/nar/gkz1221.

[68] A. Marchler-Bauer, S.H. Bryant, CD-Search: protein domain annotations on the fly, Nucleic Acids Res. 32 (2004)https://doi.org/10.1093/nar/gkh454.

[69] J.D. Cronk, J.A. Endrizzi, T. Alber, High-resolution structures of the bifunctional enzyme and transcriptional coactivator DCoH and its complex with a product analogue, Protein Sci. 5 (1996) 1963–1972, https://doi.org/10.1002/pro.5560051002.

[70] H. Erlandsen, J.Y. Kim, M.G. Patch, A. Han, A. Volner, M.M. Abu-Omar, R.C. Stevens, Structural comparison of bacterial and human iron-dependent phenylalanine hydroxylases: similar fold, different stability and reaction rates, J. Mol. Biol. 320 (2002) 645–661, https://doi.org/10.1016/s0022-2836(02)00496-5.

[71] L. Holm, Benchmarking fold detection by DaliLite v.5, Bioinformatics 35 (24) (2019) 5326–5327, https://doi.org/10.1093/bioinformatics/btz536.

[72] H. Suzuki, K. ichi Yoshid, T. Ohshima, Polysaccharide-degrading thermophiles generated by heterologous gene expression in *Geobacillus kaustophilus* HTA426, Appl. Environ. Microbiol. 79 (2013) 5151–5158, https://doi.org/10.1128/AEM.01506-13.

[73] S. Köster, G. Stier, R. Ficner, M. Hölzer, H.C. Curtius, D. Suck, S. Ghisla, Location of the active site and proposed catalytic mechanism of pterin-4α-carbinolamine dehydratase, Eur. J. Biochem. 241 (1996) 858–864, https://doi.org/10.1111/j.1432-1033.1996.00858.x.

[74] S. Köster, G. Stier, R. Ficner, M. Hölzer, H.C. Curtius, D. Suck, S. Ghisla, Three histidines in pterin-4α-carbinolalkine dehydratase are the important residues for substrate binding and catalysis, Pteridines 7 (1996) 98–100.

[75] J. Song, T. Xia, R.A. Jensen, PhhR, a *Pseudomonas aeruginosa* homolog of mammalian pterin 4α-carbinolamine dehydratase/dcoh, does not regulate expression of phenylalanine hydroxylase at the transcriptional level, J. Bacteriol. 1 (9) (1999) 2789–2796.

[76] J.A. Ronau, L.N. Paul, J.E. Fuchs, J.R. Corn, K.T. Wagner, K.R. Liedl, M.M. Abu-Omar, C. Das, An additional substrate binding site in a bacterial phenylalanine hydroxylase, Eur. Biophys. J. 42 (2013) 691–708, https://doi.org/10.1007/s00249-013-0919-8.

[77] D. Suck, R. Ficner, Structure Minireview and function of PCD/DCoH, an enzyme with regulatory properties, FEBS Lett. 389 (1) (1996) 35–39.

[78] D. Das, P. Kuzmic, B. Imperiali, Analysis of a dual domain phosphoglycosyl transferase reveals a ping-pong mechanism with a covalent enzyme intermediate, Proc. Natl. Acad. Sci. U. S. A. 114 (2017) 7019–7024, https://doi.org/10.1073/pnas.1703397114.

[79] S.D. Copley, J. Rokicki, P. Turner, H. Daligault, M. Nolan, M. Land, The whole genome sequence of *Sphingobium chlorophenolicum* L-1: insights into the evolution of the pentachlorophenol degradation pathway, Genome Biol. Evol. 4 (2012) 184–198, https://doi.org/10.1093/gbe/evr137.

SUPPORTING INFORMATION

Role of tetrachloro-1,4-benzoquinone reductase in phenylalanine hydroxylation system and pentachlorophenol degradation in *Bacillus cereus* AOA-CPS1

Oladipupo A. Aregbesola, Ajit Kumar, Mduduzi P. Mokoena, Ademola O. Olaniran*

Discipline of Microbiology, School of Life Sciences, College of Agriculture, Engineering and Science, University of KwaZulu-Natal (Westville Campus), Private Bag X54001, Durban 4000, Republic of South Africa.

***Corresponding author:** Ademola O. Olaniran
E-mail: olanirana@ukzn.ac.za
Phone: +27 31260 7400
Fax: +27 31260 7809

Running title: Benzoquinone reductase from *Bacillus cereus* AOA-CPS1

Supplementary material 1

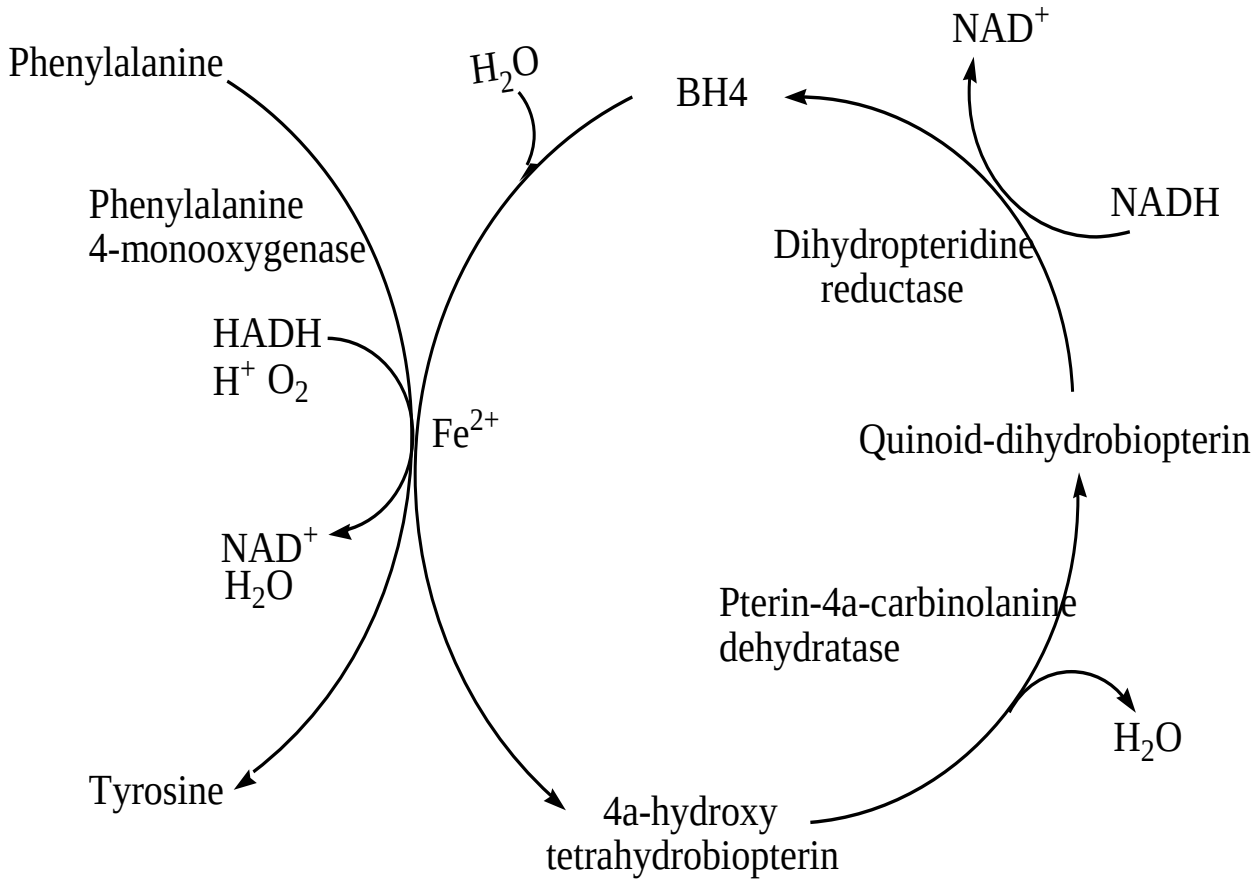


Figure S1: A typical mammalian phenylalanine hydroxylating system showing the enzymes and the substrates used by the system and the reactions catalysed by each of the enzyme. The mammalian phenylalanine hydroxylating system is made up of phenylalanine 4-monooxygenase, pteri-4α-carbinolamine dehydratase and dihydropteridine reductase (van Spronsen et al., 2017).

Supplementary material 2

Supplementary material 2a

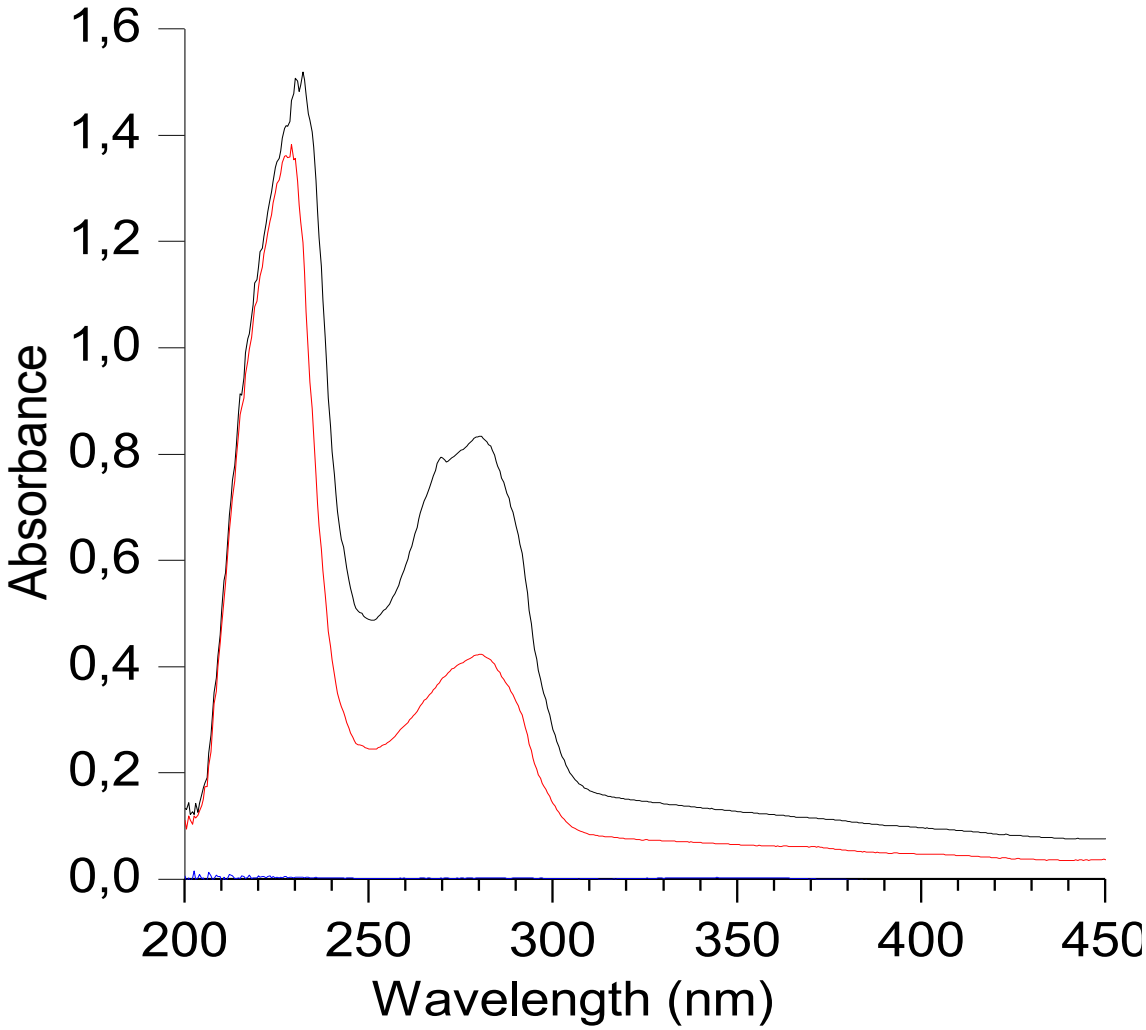


Figure S2(a): Absorption spectrum of purified CpsD. 50 (—) and 100 (—) µg of CpsD diluted with 50 mM sodium phosphate buffer pH 7.0.

Supplementary material 2b

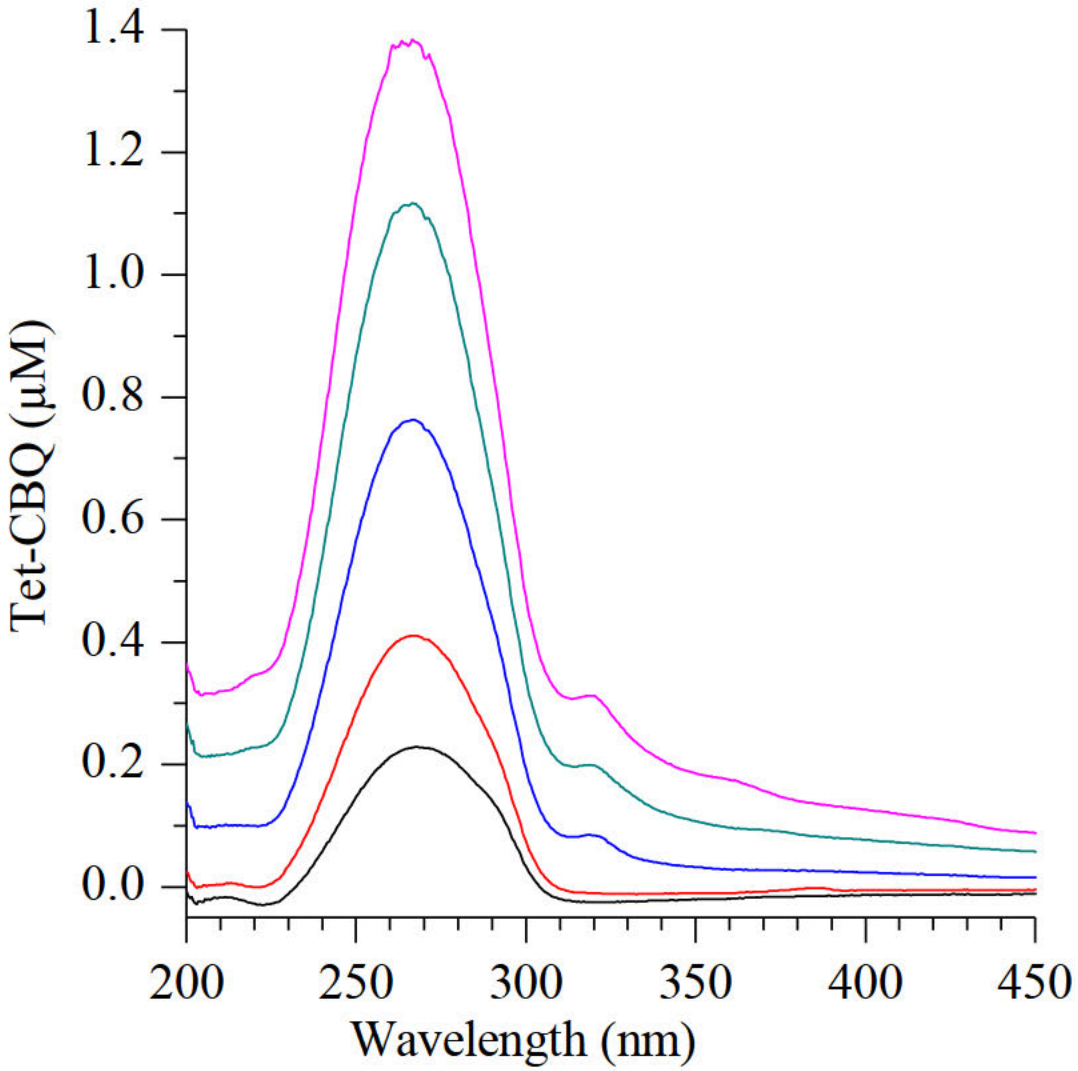


Figure S2b: Absorption spectrum of pure Tet-CBQ (at different concentration) dissolved in acetone and diluted with 50 mM sodium phosphate buffer (pH 7.0). 20 (—), 40 (—), 60 (—), 80 (—) and 100 (—) µM of Tet-CBQ diluted with sodium phosphate buffer pH 7.0.

Supplementary material 2c

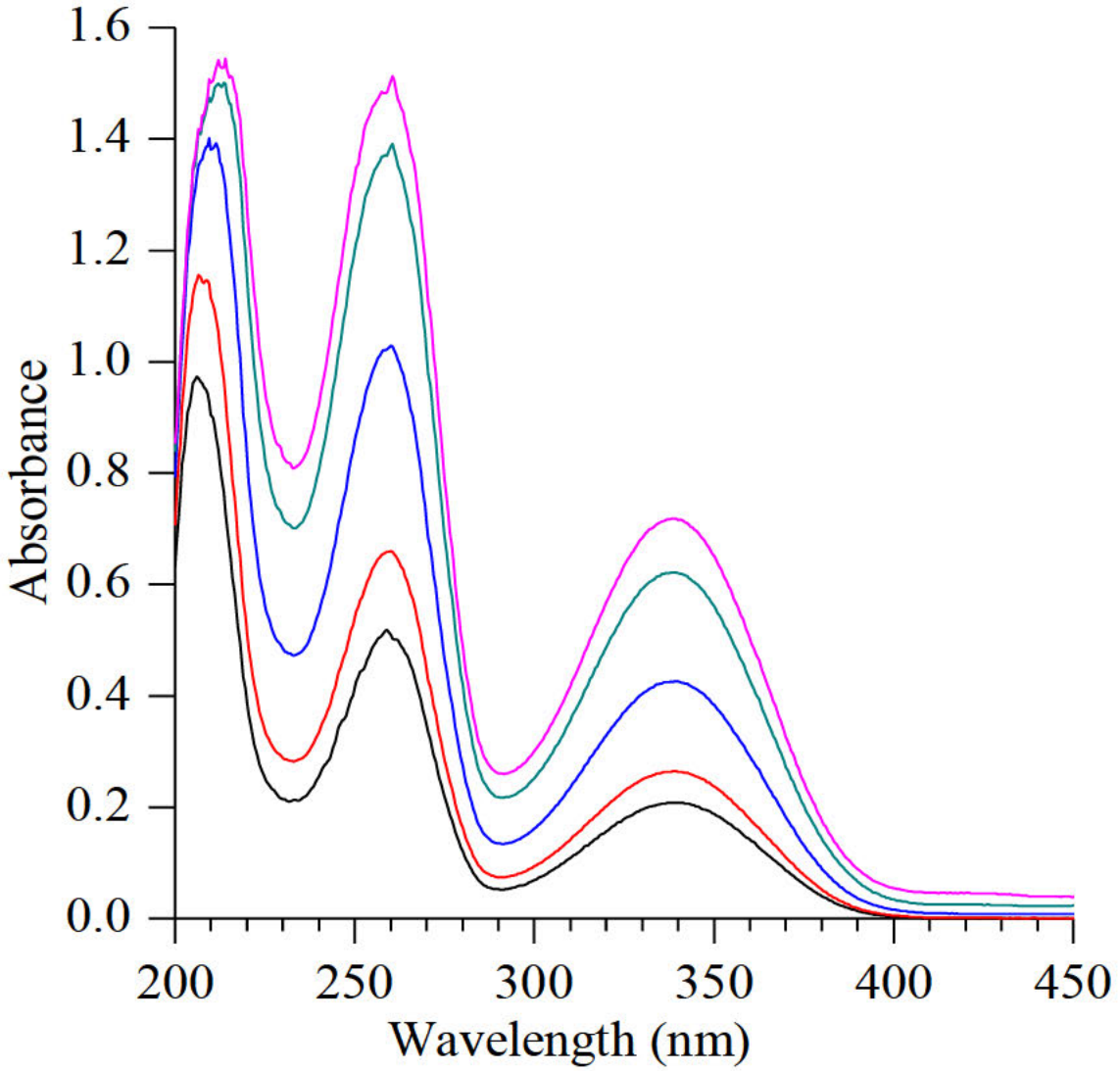


Figure S2(c): Absorption spectrum of pure NADH (at different concentration) dissolved in 50 mM sodium phosphate buffer (pH 7.0). 40 (—), 80 (—), 120 (—), 160 (—) and 200 (—) µM of NADH dissolved in 50 mM sodium phosphate buffer pH 7.0.

Supplementary material 2d

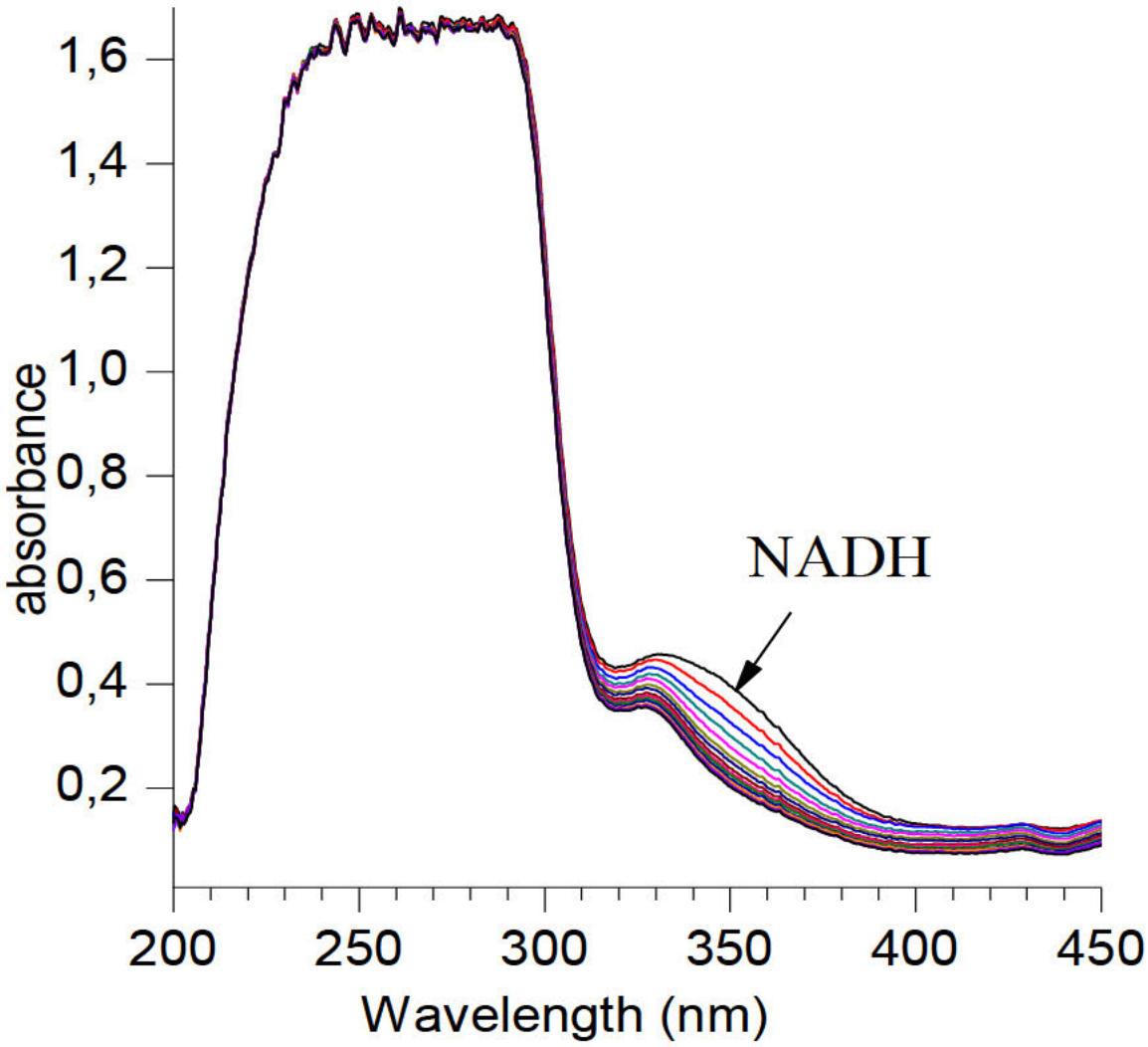


Figure S2(d): Continuous spectrophotometry reaction of CpsD with Tet-CBQ in the presence of NADH without ascorbate. The reaction mixture contained 30 µg of CpsD, 150 µM Tet-CBQ, 160 µM NADH and 50 mM $NaPO_4$ buffer (pH 7.0).

Supplementary material 2e

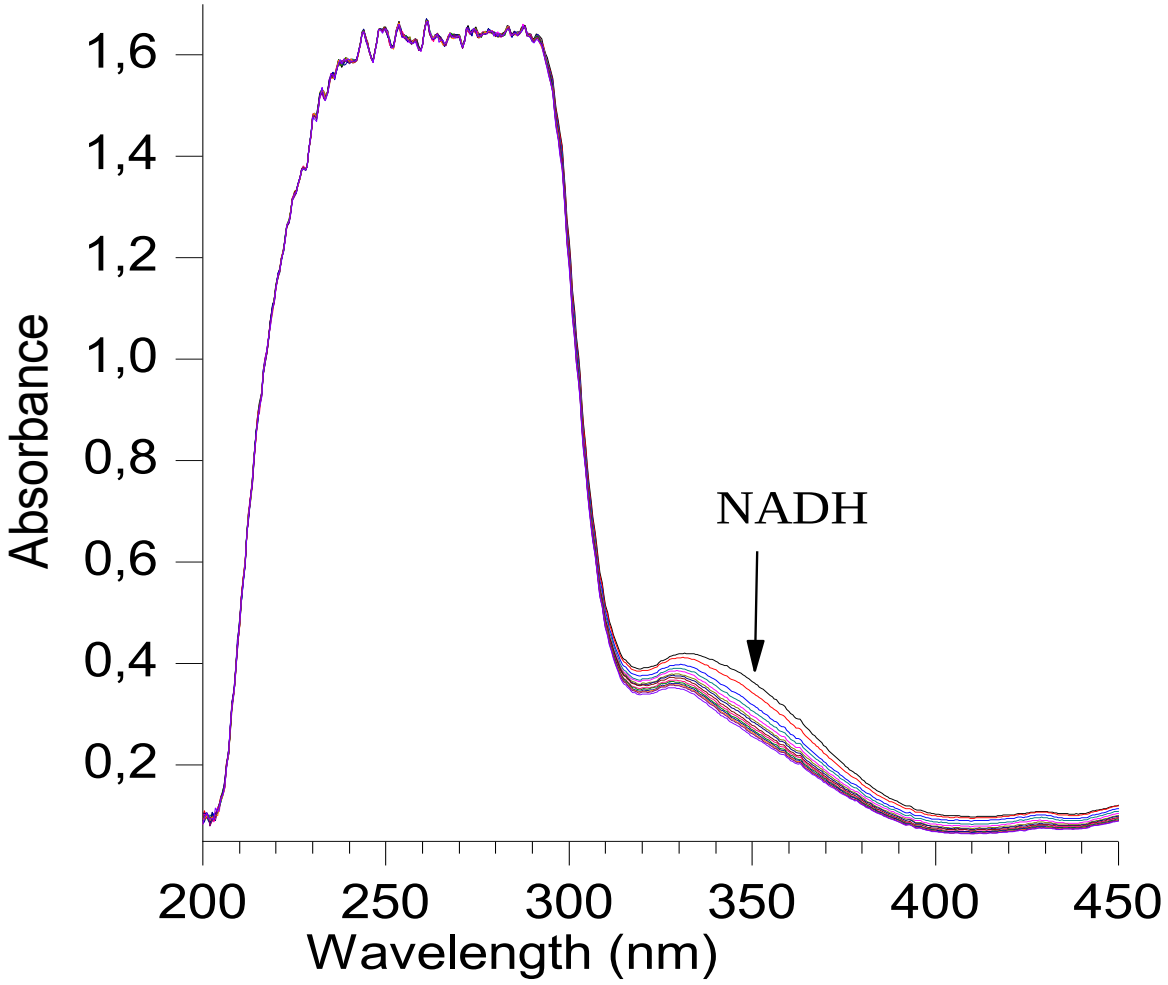


Figure S2(e): Non-enzymatic consumption of NADH in the presence of Tet-CBQ. The 1.0 ml non-enzymatic reaction 150 µM Tet-CBQ, 160 µM NADH and 50 mM $NaPO_4$ buffer (pH 7.0) only.

Supplementary material 2f

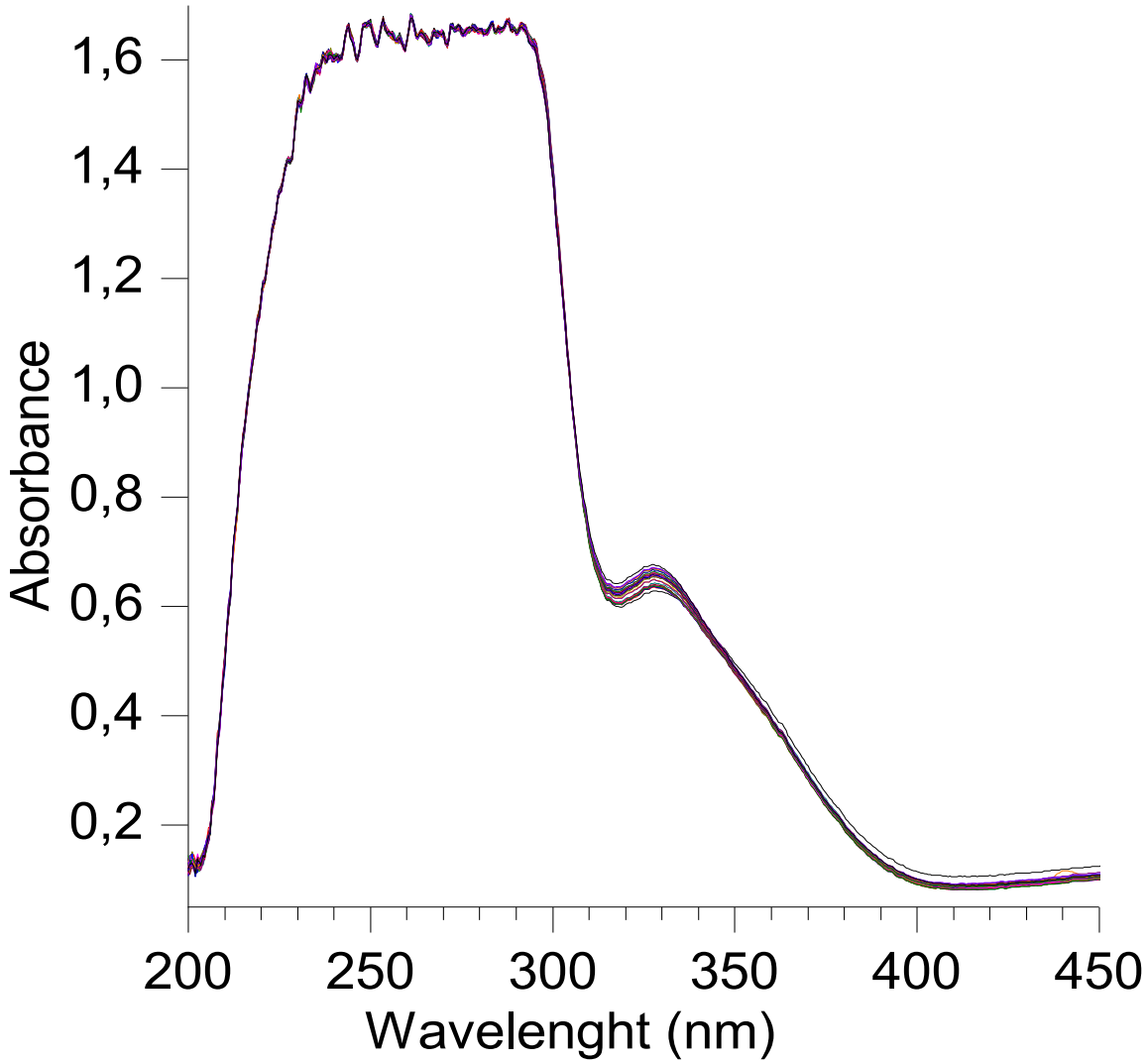


Figure S2(f): Continuous spectrophotometry reaction of CpsD with Tet-CBQ in the presence of excess ascorbate. The reaction mixture contained 350 µM ascorbate, 30 µg of CpsD, 150 µM Tet-CBQ, 160 µM NADH and 50 mM $NaPO_4$ buffer (pH 7.0). Ascorbate was added to the reaction to trap Tet-CBQ product as hydroquinone. Increase in absorbance at 327 nm represent formation and accumulation of hydroquinone.

Supplementary material 3

Supplementary material 3a

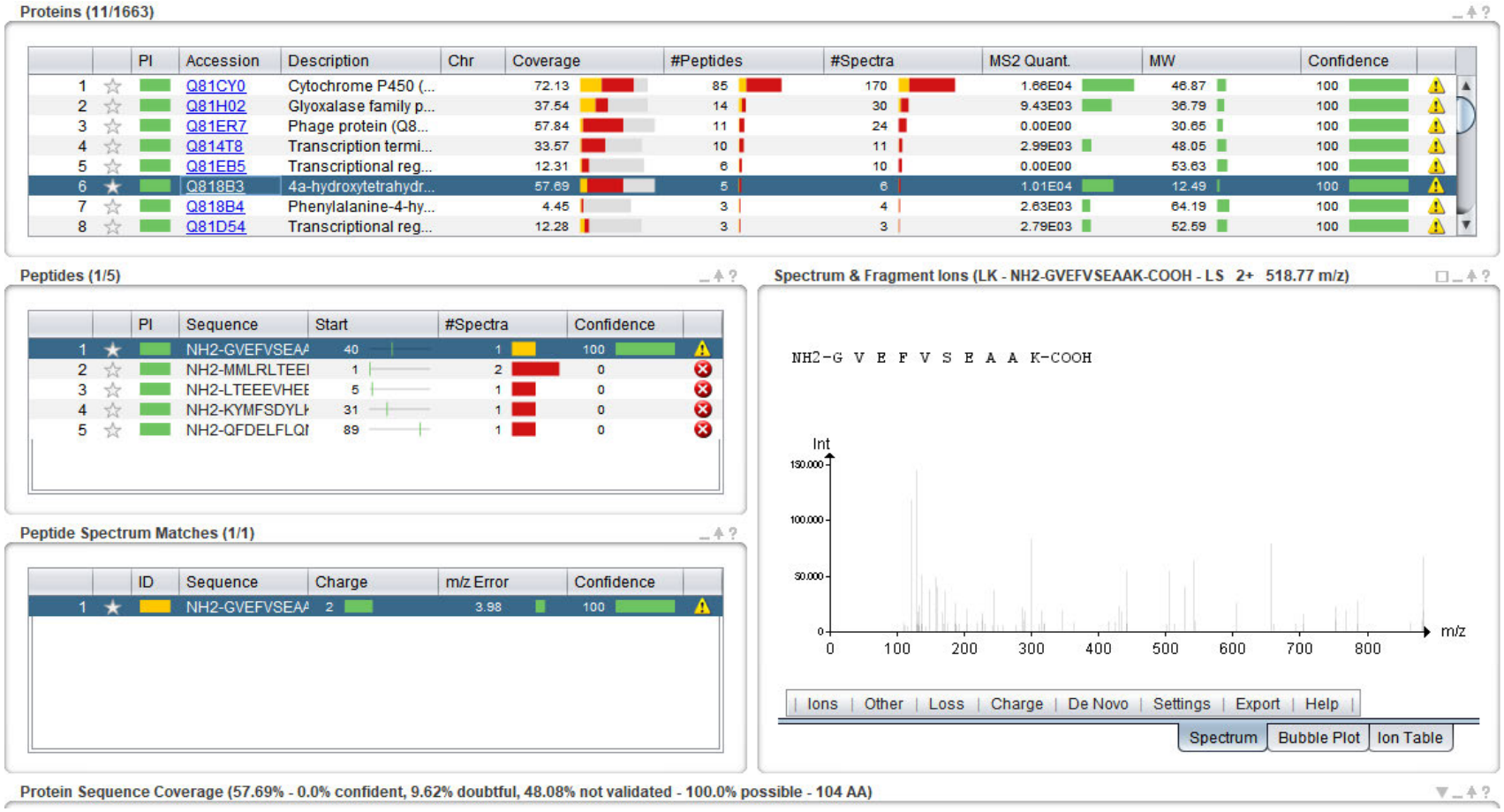


Figure S3(a): The spectrum overview

Supplementary material 3b

Peptide Spectrum Matches

	ID	Sequence	Protein(s)	Confidence	
1		NH2-GVEFVSEAAK-COOH	Q818B3	100	

Spectrum Identification Results

Validated

	SE	Rnk	Sequence	Charge	Confidence
1		1	NH2-GVEFVSEAAK-COOH	2	90
2		1	NH2-GVEFVSEAAK-COOH	2	87
3		7	NH2-AAESVFEVGK-COOH	2	0
4		8	NH2-AFVVMMNPK-COOH	2	0
5		4	NH2-ALITYDANR-COOH	2	0
6		5	NH2-GDLSFIQEK-COOH	2	0
7		1	NH2-GVEFVSEAAK-COOH	2	0
8		2	NH2-GVFYTFFR-COOH	2	0
9		9	NH2-IYDLEDIR-COOH	2	0

OMSSA X!Tandem Andromeda MS Amanda MS-GF+ DirecTag Comet MyriMatch Tide Novor PeptideShaker

Figure S3(b): Spectrum maches

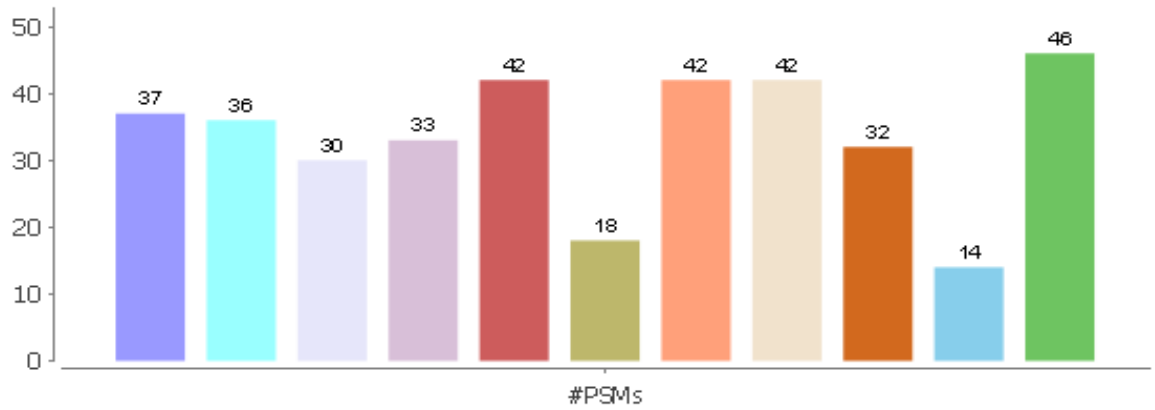


Figure S3(c): The PSMs of the peptides

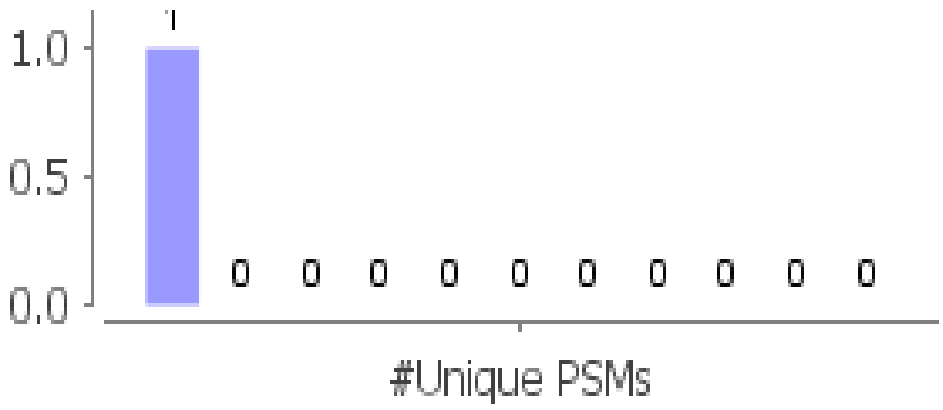


Figure S3(d): The PSMs of the peptides

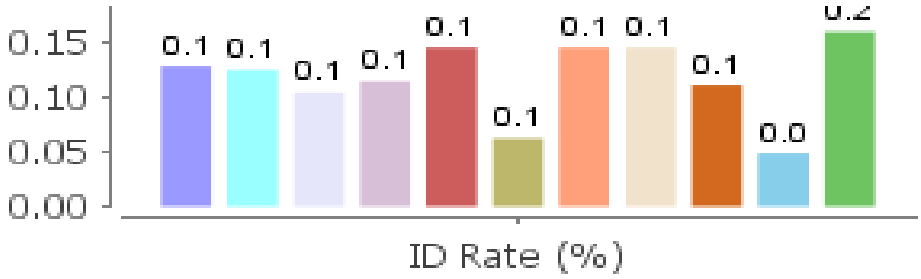


Figure S3(e): The identity (ID) rate of the peptides (%)

Figure S3(f): Protein validation plots

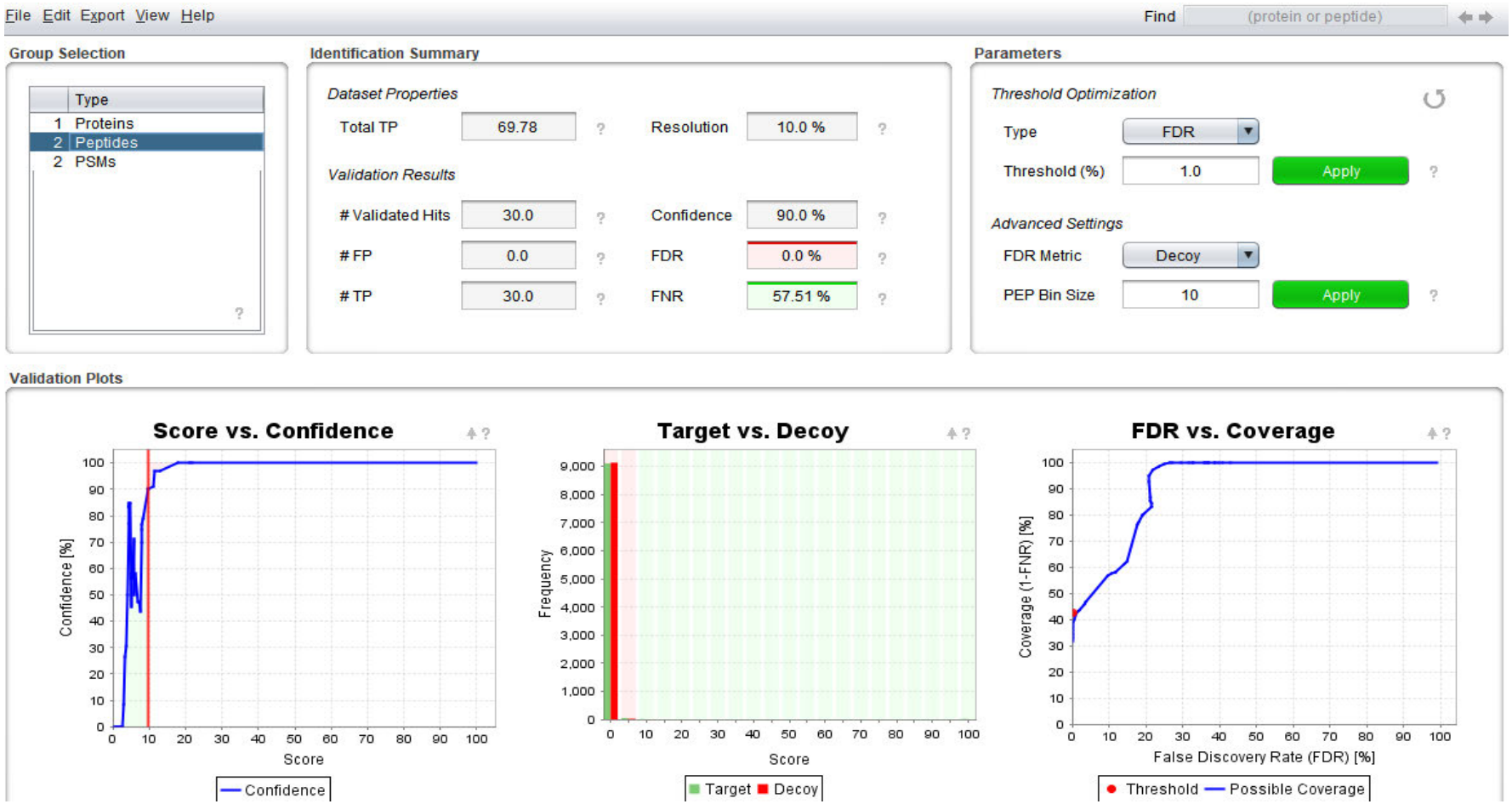


Figure S3(g): Peptide validation plots

Figure S3(h): PSMs validation

Supplementary material 4

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Manage Columns Show 100

select all *0 sequences selected* GenPept Graphics Distance tree of results Multiple alignment

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
MULTISPECIES: 4a-hydroxytetrahydrobiopterin dehydratase [Bacillus]	210	210	100%	1e-68	100.00%	WP_000979552.1
4a-hydroxytetrahydrobiopterin dehydratase [Bacillus toyonensis]	211	211	100%	1e-68	100.00%	WP_098203324.1
4a-hydroxytetrahydrobiopterin dehydratase [Bacillus toyonensis]	209	209	100%	2e-68	99.04%	WP_098854175.1
4a-hydroxytetrahydrobiopterin dehydratase [Bacillus toyonensis]	209	209	100%	2e-68	99.04%	WP_098652718.1
4a-hydroxytetrahydrobiopterin dehydratase [Bacillus toyonensis]	209	209	100%	4e-68	99.04%	WP_098710852.1
4a-hydroxytetrahydrobiopterin dehydratase [Bacillus toyonensis]	208	208	100%	5e-68	99.04%	WP_098137354.1
4a-hydroxytetrahydrobiopterin dehydratase [Bacillus sp. RM9]	208	208	99%	6e-68	100.00%	WP_002093255.1
4a-hydroxytetrahydrobiopterin dehydratase [Bacillus thuringiensis]	207	207	100%	1e-67	99.04%	WP_109737504.1
MULTISPECIES: 4a-hydroxytetrahydrobiopterin dehydratase [Bacillus]	207	207	100%	1e-67	99.04%	WP_000979538.1
MULTISPECIES: 4a-hydroxytetrahydrobiopterin dehydratase [Bacillus]	207	207	100%	2e-67	98.08%	WP_000979536.1
4a-hydroxytetrahydrobiopterin dehydratase [Bacillus cereus]	206	206	100%	3e-67	98.08%	WP_151153374.1
4a-hydroxytetrahydrobiopterin dehydratase [Bacillus thuringiensis]	206	206	100%	3e-67	98.08%	WP_109139927.1
4a-hydroxytetrahydrobiopterin dehydratase [Bacillus cereus]	206	206	100%	4e-67	98.08%	WP_098993731.1

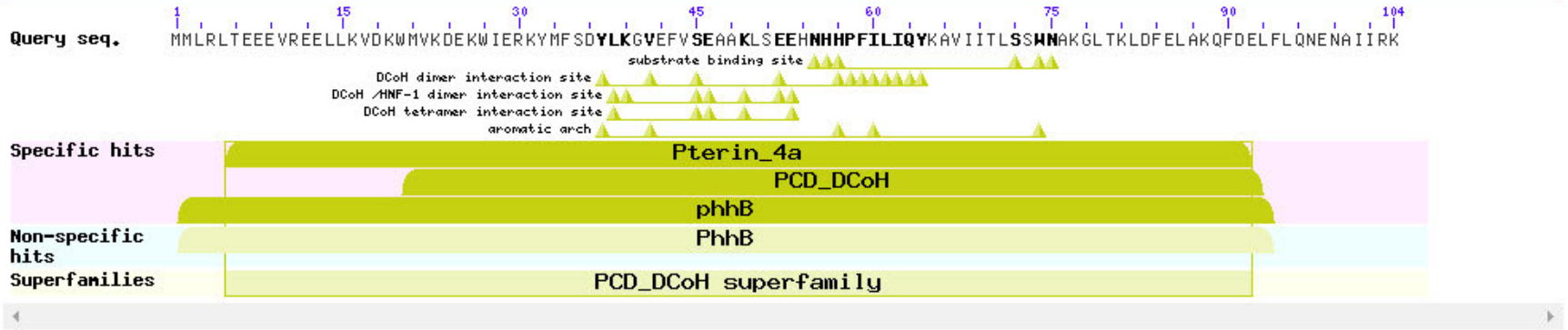
Figure S4(a): Sequences producing significant alignments

Local query sequence

Protein Classification

4a-hydroxytetrahydrobiopterin dehydratase (domain architecture ID 10473857)
4a-hydroxytetrahydrobiopterin dehydratase catalyzes the conversion from (6R)-6-(L-erythro-1,2-dihydroxypropyl)-5,6,7,8-tetrahydro-4a-hydroxypterin to (6R)-6-(L-erythro-1,2-dihydroxypropyl)-7,8-dihydro-6H-pterin in tetrahydrobiopterin biosynthesis

Graphical summary Zoom to residue level show extra options »



Search for similar domain architectures Refine search

List of domain hits

	Name	Accession	Description	Interval	E-value
[+]	Pterin_4a	pfam01329	Pterin 4 alpha carbinolamine dehydratase; Pterin 4 alpha carbinolamine dehydratase is also ...	5-92	1.14e-28
[+]	PCD_DCoH	cd00488	PCD_DCoH: The bifunctional protein pterin-4alpha-carbinolamine dehydratase (PCD), also known ...	20-93	4.86e-27
[+]	phhB	PRK00823	pterin-4-alpha-carbinolamine dehydratase; Validated	1-94	1.14e-26
[+]	PhhB	COG2154	Pterin-4a-carbinolamine dehydratase [Coenzyme transport and metabolism];	1-94	1.06e-23

Blast search parameters

Data Source: Live blast search RID = ZK7UVBTM01N
User Options: Database: CDSEARCH/cdd Low complexity filter: no Composition Based Adjustment: yes E-value threshold: 0.01 Maximum number of hits: 500

Figure S4(b): Conserved domains detected in CpsD structural model on the NCBI database.

Figure S4(c): The Dali search summary of the CpsD comparison to protein structures in 3D

S/N	Chain	*Z*-score	rmsd	lali	nres	% ID	Description
1	2ebb-A	21.8	0.1	96	96	59	molecule: pterin-4-alpha-carbinolamine dehydratase
2	3jst-a	17.2	1.1	93	97	32	molecule: pterin-4-alpha-carbinolamine dehydratase
3	1f93-a	16.8	1.5	96	103	27	molecule: pterin-4-alpha-carbinolamine dehydratase
4	2v6u-a	16.4	1.7	96	103	33	molecule: pterin-4-alpha-carbinolamine dehydratase
5	1uso-b	13.8	1.4	77	78	25	molecule: pterin-4-alpha-carbinolamine dehydratase
6	4low-a	12.3	1.7	76	84	26	molecule: acraf
7	4kyz-a	9.0	3.2	77	167	12	molecule: designed protein or327
8	4ney-b	8.5	1.9	68	173	12	molecule: engineered protein or277
9	5fxd-a	8.4	2.6	80	525	8	molecule: probable vanillyl-alcohol oxidase
10	3pm9-a	8.3	2.9	76	465	11	molecule: putative oxidoreductase
11	6c0d-a	8.1	3.6	73	404	8	molecule: amidase, hydantoinase/carbamoylase family
12	5i4m-a	7.7	2.2	71	414	7	molecule: amidase, hydantoinase/carbamoylase family
13	1dii-a	7.7	2.5	80	515	6	molecule: p-cresol methylhydroxylase
14	6gwj-b	7.5	2.3	68	84	9	molecule: ekc/keops complex subunit lage3
15	1ahu-a	7.4	2.6	80	555	6	molecule: vanillyl-alcohol oxidase
16	3n5f-a	7.3	1.7	69	406	6	molecule: n-carbamoyl-l-amino acid hydrolase
17	4pxc-a	7.3	2.1	71	424	13	molecule: ureidoglycolate hydrolase
18	2uuu-a	7.2	1.6	63	550	11	molecule: alkyldihydroxyacetonephosphate synthase
19	6o7v-a	7.2	1.4	57	625	11	molecule: v-type proton atpase subunit d
20	1yz6-a	7.2	2.2	69	261	10	molecule: probable translation initiation factor 2 alpha

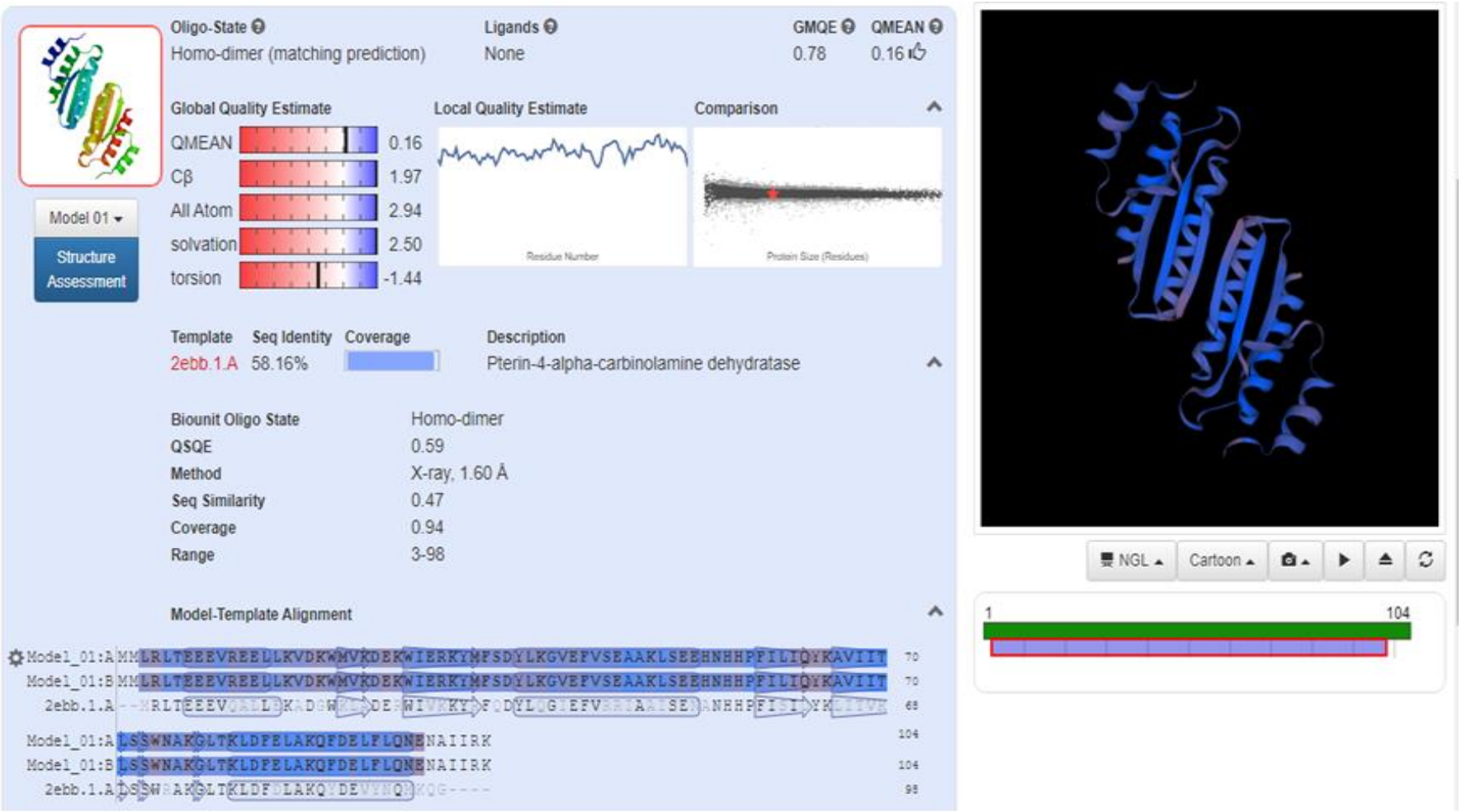


Figure S4(d): CpsD structural model built using PDB entry 2ebb as a template

Supplementary material 5

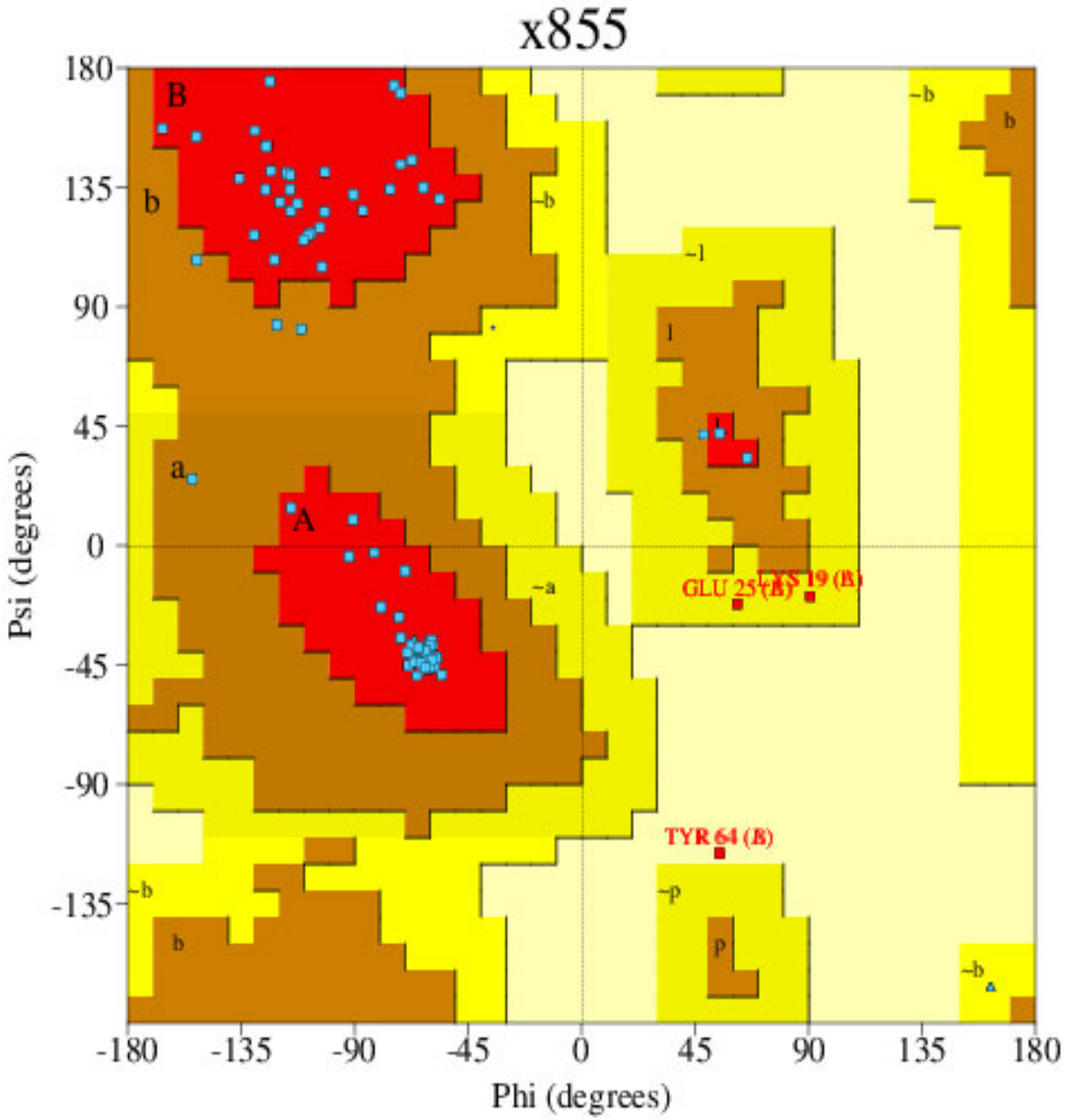


Figure S5a: Ramachandran Plot statistics for CpsD structural protein.

Figure S5b: Beta sheets found in CpsD secondary structure

Sheet	View	No. of strands	Type	Barrel	topology
A		8	Antiparallel	No	1 2X – 1 5 – 1 -2X 1

Figure S5c: List of beta hairpins found in CpsD secondary structure

	Strand 1			Strand 2			Hairpin
S/N	**Start**	**End**	**Length**	**Start**	**End**	**Length**	**class**
1	Met21	Lys23	3	Trp27	Met23	7	3:3
2	Phe59	Gin63	5	Ala66	Leu71	6	2:2 llP

Note: the table gives the beginning and the end residues, and lengths, of the two strands involved in the hairpin together with the hairpin class

Figure S5d: Beta bulges found in CpsD secondary structure

No.	Bulge type	Res X	Res 1	Res 2	Res 3	Res 4
1	Antiparallel class	Lys23A	Lys26A	Trp27A		

The table gives the bulge type, and residues X, 1, 2, 3 and 4. Most of the bulges will have just residues X, 1 and 2; the special bulges may have one or more of residues 3 and 4, and the bent bulges will just have residues 1 and 2. A beta bulge is a region of irregularity in a beta sheet, where the normal pattern of hydrogen bonding is disrupted e.g. by the insertion of an extra residue

Figure S5(e): Lists of beta strands found in CpsD secondary structure

No.	Start	End	Sheet	Number of residues	Edge	Sequence
1	Met21	Lys23	A	3	Yes	MVK
2	Trp27	Met33	A	7	No	WIERKYM
3	Phe59	Gin63	A	5	No	FILIQ
4	Ala66	Leu71	A	6	No	AVIITL

Note: for each beta strand in the protein chain the table gives the strand number (assigned sequentially from the N-terminus of the protein), the start and end residues, the letter corresponding to the sheet in which the strand is involved, the number of residues in the strand, whether the strand is at the edge of a beta sheet or not and its amino acid sequence.

Figure S5(f): Helix-helix interactions found in CpsD secondary structure

No.	View	Helices		Helix types		distance (Å)	Angle (◦)	Interaction type		No. of interacting residues		
										Total	Helix 1	Helix 2
1		**A1**	A3	**H**	H	11.4	-11.1	**N**	N	4	3	2
2		**A2**	A3	**H**	H	10.6	148.3	**1**	1	13	6	5
3		**A2**	B2	**H**	H	10.0	178.0	**1**	1	15	7	7

Figure S5(g): Helices found in CpsD secondary structure

No.	Start	End	Type	No. of Residues	Length	Unit Rise	Residues Per turn	Pitch	Deviation from ideal	Sequence
1	Glu7	Leu14	H	8	12.36	1.48	3.71	5.49	12.5	EEEVREEL
2	Tyr37	Glu53	H	17	26.51	1.53	3.59	5.50	8.0	YLKGVEFVSEAAKLSEE
3	Lys81	Asn97	H	17	26.36	1.52	3.59	5.45	5.9	KLDFELAKQFDELFLON

Note: Table for each helix includes the helix number (assigned sequentially starting with 1 at the N-terminus of the protein), the residue numbers corresponding to the start and end of the helices, the helix type (**H** (alpha helix) or **G** (3, 10) helix. This is followed by the number of residues in the helix and information about the geometry of the helix as follows: length and unit rise (both in Angstroms), the number of residues per turn (ideally 3.6) for alpha helices), the helix pitch in Angstroms and perfect helix. The geometrical parameters are not calculated for helices with fewer than four residues. The final column in the table gives the helix's amino acid sequence.

Figure S5(h): Turns found in CpsD secondary structure

			Turn	Residue i+1			Residue i+2			I to i+3	
No.	**Turn**	**Sequence***	**Type**	**Phi**	**Psi**	**Chi1**	**Phi**	**Psi**	**Chi1**	**CA-dist**	**H-bond**
*1	Val17-Try20	VDKW	II	-56.2	130.7	-69.2	90.7	-19.1	-58.5	5.8	No
*2	Val22-Glu25	VKDE	IV	-152.7	154.0	-168.6	48.3	41.8	-49.7	6.9	Yes
*3	Lys23-Lys26	KDEK	I	48.3	41.8	-49.7	61.7	-22.0	-56.7	6.4	No
*4	Asp24-Trp27	DEKW	IV	61.7	-22.0	-56.7	-154.2	25.1	-63.6	5.9	Yes
*5	Phe34-Trp37	FSDY	VIII	-79.4	-23.3	51.8	-123.2	141.2	176.5	6.7	Yes
6	Gln63-Ala66	QYKA	II	54.7	-116.0	-169.2	-90.8	9.9	-65.9	5.1	No
7	Ser73-Lys76	SWNA	IV	-60.7	-41.2	68.2	-59.0	-37.3	-72.0	5.4	Yes
8	Trp74-Lys77	WNAK	I	-59.0	-37.3	-72.0	-92.2	-4.2	-	4.8	no

Note: A beta is defined for 4 consecutive residues (denoted by i, i+1, i+2 and i+3), if the distance between the Capha atom of residue I and the Capha atom of residue i+3 is less than 7Å and if the central two residues are not helical. The residue numbers of residues i and i+3 in the turn, the one-letter amino acid code of residues i, i+1, i+2 and i+3 in the turn, and the turn type. For each of the central two residues (i+1 and i+2) phi, psi and chi1 are recorded. The final column shows the distance between the C alpha atoms of residues I and i+3 and whether or not a hydrogen bond exists between these two residues. The phi and psi angles are allowed to vary by +/- 30 degrees from these ideal values with the added flexibility of one angle being allowed to deviate by as much as 40 degrees. Types VIa1, VIa2 and VIb turns are subject to the additional condition that residue i must be a cis-proline. Turns which do not fit any of the above criteria are classified as type IV.

Supplementary material 6

Figure S6: Cleft found in CpsD structure

Clefts

	Volume	R1 ratio	Accessible vertices		Buried vertices		Average depth		Residue type							Ligands
1	1964.25	1.06	63.33	5	9.42	2	11.35	1	10	5	8	7	5	1	0	
2	1859.20	0.00	62.03	7	9.72	1	10.18	4	8	5	8	6	4	1	0	
3	1201.08	0.00	60.59	10	8.43	4	10.61	2	6	4	1	8	1	1	0	
4	1193.06	0.00	60.71	9	8.98	3	10.30	3	5	4	1	8	1	0	0	
5	1292.20	0.00	64.78	3	6.55	5	8.87	5	7	5	2	6	2	0	0	
6	1290.52	0.00	64.74	4	6.14	6	6.92	7	7	5	2	8	4	0	0	
7	399.94	0.00	65.57	2	5.15	7	8.06	6	3	2	2	4	0	0	0	
8	183.09	0.00	62.83	6	0.38	9	4.98	8	2	2	1	1	2	0	0	
9	177.61	0.00	61.28	8	0.19	10	0.00	9	2	2	1	1	2	0	0	
10	151.88	0.00	75.23	1	4.95	8	0.00	10	3	2	0	1	1	0	0	

Protein structure

Residue-type colouring

Positive	Negative	Neutral	Aliphatic	Aromatic	Pro & Gly	Cysteine
H,K,R	D,E	S,T,N,Q	A,V,L,I,M	F,Y,W	P,G	C

Supplementary material 7

Figure S7(a): List of hydrogen bonds interactions across protein-protein interface

S/N	ATOM 1						ATOM 2					Distance
	Atom No.	Atom name	Res name	Res No.	Chain		Atom No.	Atom name	Res name	Res No.	Chain	
1	496	N	ILE	60	A	<->	1333	N	ILE	60	B	3.20
2	499	O	ILE	60	A	<->	1330	O	ILE	60	B	2.73
3	512	N	ILE	62	A	<->	1317	N	ILE	62	B	2.73
4	515	O	ILE	62	A	<->	1314	O	ILE	62	B	3.20

Figure S7(b): List of non-bonded contacts across protein-protein interface

S/N	ATOM 1						ATOM 2					Distance
	Atom No.	Atom name	Res name	Res No.	Chain		Atom No.	Atom name	Res name	Res No.	Chain	
1	311	CB	TYR	37	A	<->	1247	OE1	GLU	52	B	3.28
2	312	CG	TYR	37	A	<->	1247	OE1	GLU	52	B	3.62
3	312	CG	TYR	37	A	<->	1293	CD2	HIS	57	B	3.77
4	314	CD2	TYR	37	A	<->	1294	OE1	GLU	52	B	3.33
5	314	CD2	TYR	37	A	<->	1293	CB2	HIS	57	B	3.54
6	316	CE2	TYR	37	A	<->	1215	O	ALA	48	B	3.55
7	316	CE2	TYR	37	A	<->	1216	CB	ALA	48	B	3.80
8	316	CE2	TYR	37	A	<->	1293	CD2	HIS	57	B	3.62
9	318	OH	TYR	37	A	<->	1216	CB	ALA	48	B	3.73
10	326	CD2	LEU	38	A	<->	1215	O	ALA	48	B	3.79
11	345	CG1	VAL	41	A	<->	1193	CA	SER	45	B	3.90
12	346	CG2	VAL	41	A	<->	1216	CB	ALA	48	B	3.58
13	372	CG1	VAL	44	A	<->	1337	CD1	ILE	62	B	3.63
14	375	CA	SER	45	A	<->	1163	CG1	VAL	41	B	3.90
15	379	OG	SER	45	A	<->	1197	OG	SER	45	B	3.19
16	397	O	ALA	48	A	<->	1134	CE2	TYR	37	B	3.55
17	397	O	ALA	48	A	<->	1144	CD2	LEU	38	B	3.79

Figure S7(b): List of non-bonded contacts across protein-protein interface (Cont'd)

	ATOM 1						ATOM 2					
S/N	**Atom No.**	**Atom name**	**Res name**	**Res No.**	**Chain**		**Atom No.**	**Atom name**	**Res name**	**Res No.**	**Chain**	**Distance**
18	398	CB	ALA	48	A	<->	1134	CE2	TYR	37	B	3.80
19	398	CB	ALA	48	A	<->	1136	OH	TYR	37	B	3.73
20	398	CB	ALA	48	A	<->	1164	CG2	VAL	41	B	3.58
21	429	OE1	GLU	52	A	<->	1129	CB	TYR	37	B	3.28
22	429	OE1	GLU	52	A	<->	1130	CG	TYR	37	B	3.62
23	429	OE1	GLU	52	A	<->	1132	CD2	TYR	37	B	3.33
24	473	CG	HIS	57	A	<->	1353	CD1	TYR	64	B	3.64
25	473	CG	HIS	57	A	<->	1355	CE1	TYR	64	B	3.74
26	474	ND1	HIS	57	A	<->	1153	CD1	TYR	64	B	3.82
27	474	ND1	HIS	57	A	<->	1355	CE1	TYR	64	B	3.54
28	475	CD2	HIS	57	A	<->	1130	CG	TYR	37	B	3.77
29	475	CD2	HIS	57	A	<->	1132	CD2	TYR	37	B	3.54
30	475	CD2	HIS	57	A	<->	1134	CE2	TYR	37	B	3.62
31	475	CD2	HIS	57	A	<->	1353	CD1	TYR	64	B	3.88
32	475	CD2	HIS	57	A	<->	1355	CE1	TYR	64	B	3.81
33	476	CE1	HIS	57	A	<->	1355	CE1	TYR	64	B	3.45
34	477	NE2	HIS	57	A	<->	1355	1	TYR	64	B	3.62
35	486	CA	PHE	59	A	<->	1333	O	ILE	62	B	3.77
36	494	CE2	PHE	59	A	<->	1328	CD1	LEU	61	B	3.55
37	495	CZ	PHE	59	A	<->	1328	CD1	LEU	61	B	3.83
38	496	N	ILE	60	A	<->	1333	O	ILE	62	B	3.20
39	499	O	ILE	60	A	<->	1323	CA	LEU	61	B	3.40
40	499	O	IE	60	A	<->	1324	C	LEU	61	B	3.53
41	499	O	ILE	60	A	<->	1326	CB	LEU	61	B	3.71
42	499	O	ILE	60	A	<->	1330	N	ILE	62	B	2.73.

Figure S7b: List of non-bonded contacts across protein-protein interface (Cont'd)

	ATOM 1						ATOM 2					
S/N	**Atom No.**	**Atom name**	**Res name**	**Res No.**	**Chain**		**Atom No.**	**Atom name**	**Res name**	**Res No.**	**Chain**	**Distance**
43	499	O	ILE	60	A	<->	1331	CA	ILE	62	B	3.65
44	499	O	ILE	60	A	<->	1332	C	ILE	62	B	3.90
45	499	O	ILE	60	A	<->	1333	O	ILE	62	B	3.49
46	499	O	ILE	60	A	<->	1334	CB	ILE	62	B	3.84
47	505	CA	LEU	61	A	<->	1317	O	ILE	60	B	3.40
48	506	C	LEU	61	A	<->	1317	O	ILE	60	B	3.53
49	508	CE	LEU	61	A	<->	1317	O	ILE	60	B	3.71
50	510	CD1	LEU	61	A	<->	1312	CE2	PHE	59	B	3.55
51	510	CD1	LEU	61	A	<->	1313	CZ	PHE	59	B	3.85
52	512	N	ILE	62	A	<->	1317	O	ILE	60	B	2.73
53	513	CA	ILE	62	A	<->	1317	O	ILE	60	B	3.65
54	514	C	ILE	62	A	<->	1317	O	ILE	60	B	3.90
55	515	O	ILE	62	A	<->	1304	CA	PHE	59	B	3.77
56	515	O	ILE	62	A	<->	1314	N	ILE	60	B	3.20
57	515	O	ILE	62	A	<->	1317	O	ILE	60	B	3.49
58	516	CB	ILE	62	A	<->	1317	O	ILE	60	B	3.84
59	519	CD1	ILE	62	A	<->	1190	CG1	ILE	44	B	3.63
60	535	CD1	TYR	64	A	<->	1291	CG	VAL	57	B	3.64
61	535	CD1	TYR	64	A	<->	1292	ND1	HIS	57	B	3.82
62	535	CD1	TYR	64	A	<->	1293	CD2	HIS	57	B	3.88
63	537	CE1	TYR	64	A	<->	1291	CG	HIS	57	B	3.74
64	537	CE1	TYR	64	A	<->	1292	ND1	HIS	57	B	3.54
65	537	CE1	TYR	64	A	<->	1293	CD2	HIS	57	B	3.81
66	537	CE1	TYR	64	A	<->	1294	CE1	HIS	57	B	3.45
67	537	CE1	TYR	64	A	<->	1295	NE2	HIS	57	B	3.62

Supplementary material 8

```
CpsD          ------------------------------------------------------------   0
ZP00628055    ------------------------------------------------------------   0
ZP00400454    --------MVHLVLRAGLEDGQE-----IRRQPRLQRVRAERARRHRHRRRDRAEREEEP   47
CAE78833      ------------------------------------------------------------   0
NP840178      ------------------------------------------------------------   0
AAU28068      ------------------------------------------------------------   0
Q9PAB4        ------------------------------------------------------------   0
ABB41227      ------------------------------------------------------------   0
BAD72380      MTRGVAMAHARLLLARYYAMAAPSWPTVSKNLPLLGHG------RS-HHPM-YASQDEI-   51
NP174274      --------MSRLLLPKLFSISRTQVPAASLFNNLYRRH-----KRFVHWTS-KMSTDSVR   46
XP641894      -----MKVLTSLKLKNN-------CKILSAYLPNLRNL-----KSNNKTRL-FV------   36
YP436101      ------------------------------------------------------------   0
Q8KFI4        ------------------------------------------------------------   0
BAC88867      ------------------------------------------------------------   0
AAZ21144      ------------------------------------------------------------   0
ZP00798856    ------------------------------------------------------------   0
ZP01012805    ------------------------------------------------------------   0
ZP00864663    ------------------------------------------------------------   0
AAU93190      ------------------------------------------------------------   0
CAI06204      ------------------------------------------------------------   0
AAZ59799      ------------------------------------------------------------   0
ABB53415      ------------------------------------------------------------   0
BAE59828      ------------------------------------------------------------   0
AAN51894      -------------------------------MPIL-------------RRLIPLLLNYDP   16
AAK47768      ------------------------------------------------------------   0
NP982992      ------------------------------------------------------------   0
CAG98922      --------MALK----------------VFVQSTF-------------DKMVLLVFCF--   21
AAK42360      ------------------------------------------------------------   0
ZP00520697    ------------------------------------------------------------   0
ZP00774037    ------------------------------------------------------------   0
AAZ59312      ------------------------------------------------------------   0
AAP99535      ------------------------------------------------------------   0
AAP06387      ------------------------------------------------------------   0
Q8XU38        ------------------------------------------------------------   0
AAN29038      ------------------------------------------------------------   0
YP428662      ------------------------------------------------------------   0
ZP00548450    ------------------------------------------------------------   0
CAG90387      ------------------------------------------------------------   0
EAN85354      ------MVVCCL----------------FFFFTLFT------------RQPFFLIFNFN-   25
ZP00412175    ------------------------------------------------------------   0
O86722        ------------------------------------------------------------   0
AAZ18005      ------------------------------------------------------------   0
ZP00950887    ------------------------------------------------------------   0
ABC46318      -------------------------------MNRSRRS--------SRLRPDAPILESG-   20
AAY42130      ------------------------------------------------------------   0
ZP00396203    ------------------------------------------------------------   0
AAG19997      ------------------------------------------------------------   0
ZP00658229    ------------------------------------------------------------   0
BAB97857      ------------------------------------------------------------   0
ZP00283397    ------------------------------------------------------------   0
ZP00571129    ------------------------------------------------------------   0
AAK45453      ------------------------------------------------------------   0
ZP00993810    ------------------------------------------------------------   0
BAD59613      ------------------------------------------------------------   0
```

Figure S8a: Multiple sequences alignments of CpsD with PCD/DCoH from other members of PCD/DCoH superfamily. Conserved residues were shaded in green**.**

```
CpsD          ----------------------------------------------------MMLRLTE   7
ZP00628055    ------------------------------------------------------------   0
ZP00400454    VHGRGVS---PIARPGGRPA---LALRRRGPARTRSAMDLDFAKRKCVPCEG-GIPALAP   100
CAE78833      -----------------------------------MMSQTELLRKKSHPV----DQALTP   21
NP840178      -----------------------------------MTNVCDLTDRKCKPCEG-GVPPLEM   24
AAU28068      ------------------------------------MTSDLSSKHCESCEG-IGAALNS   22
Q9PAB4        -----------------------------------MNDLITLAQAHCQPREK-KEHKLGQ   24
ABB41227      -----------------------------------MTTEIALKDQSCETIEK-GSKAMII   24
BAD72380      ----------KM-SSRRWCH---GS-----------PDNQELAKKICVPCNSKDIHAMPE   86
NP174274      SSTTGGS---AS-GARTFCS------------------LADLSTKKCVPCNAKDLRAMTE   84
XP641894      VQENNNN---IM-NINAFCS---SSISNNTNKETNIDEKVDLSKKHCQPCEG-GIPPLDF   88
YP436101      --------------------------------------MSDTIPSCEACRP-DAEKVDP   20
Q8KFI4        -------------------------------------MTQLENKHCVPCEG-TAAPMAS   21
BAC88867      --------------------------------------MNLTEQRCTACRP-DAPRVGA   20
AAZ21144      -------------------------------------MSDLLNKKCVPCEG-GILPFDI   21
ZP00798856    --------------------------------------MNNLAEKKCIPCSL-GTPPLSS   21
ZP01012805    ----------------------------------------MTEKSCTDLPK-GTPALKA   18
ZP00864663    -------------------------------------MSLTDKVCTACQG-GVEPMDE   20
AAU93190      -----------------------------------MDACSLTAKQCTPCQG-GIPPLTA   23
CAI06204      -----------------------------------MSDELQSRTCTPCRG-DVPPMTK   22
AAZ59799      -----------------------------------MSEKLESQTCTPCRG-GIPPLER   22
ABB53415      ------------------------------------------------------------   0
BAE59828      --------------------------------------MAAAPKFARG----MNKGQLQ   17
AAN51894      SRILFIDCILVENKTTTLCG--RS----------------------------FMSDLNN   45
AAK47768      ----------MAHGNVSRCE--ESSLHD-----------------VCCG----RLSALTD   27
NP982992      ----------------------------------------MYNKIAKQ----LPILLRP   15
CAG98922      FSIFLVNKDFLHLGILSTTH--REMLEKVINATLQALNSIDMYNKIVKQ----VPVLLPP   75
AAK42360      ----------------------------------MILFKKLFLNSSFFLF----LMSGISS   23
ZP00520697    ----------------------------------------------------MTTKLSE   7
ZP00774037    ----------------------------------------MNRKKEDY----MVDKLSE   15
AAZ59312      ---------------------------------------------MKYEKL----MVSLIEK   13
AAP99535      ------------------------------------------------M----ENKLLSS   8
AAP06387      -----------------------------------------------MP----LLTDLER   9
Q8XU38        ---------------------------------------------MMP----TLNDEQR   10
AAN29038      ------------------------------------------------M----ARNRLTE   8
YP428662      ----------------------------------------------------MTDRLDD   7
ZP00548450    ----------------------------------------------MG----LRDPLPA   9
CAG90387      ------------------------M----------------SSKTI--R----KAAVIAG   14
EAN85354      FLFF------FSETIFSGCVFLEEM----------------RRATILCA----IWMPFSQ   59
ZP00412175    ---------------------------------------------MTG----TEDILTQ   10
O86722        ------------------------------------------------M----PAEPLSP   8
AAZ18005      -------------------------------------------------------MSSLSN   6
ZP00950887    ---------------------------------------------------------MKLTE   5
ABC46318      FSNF-----------------LSE------------------HQPEHMP----SREPLSD   41
AAY42130      ------------------------------------------------MS----RVEPLSR   9
ZP00396203    --------------MTAMC---PEL----------------YDPRMGYD----PSRKLTD   23
AAG19997      ----------------------------------------------------MSDRLDD   7
ZP00658229    ---------------------------------------------MTD----PKQVLRF   10
BAB97857      ----------------------------------------------------MCAILRQ   7
ZP00283397    ---------------------------------------------MAN----AERVYSD   10
ZP00571129    ----------------------------------------------------MPTRLDN   7
AAK45453      ----------------------------------------MTVYRR----GMAVLTD   13
ZP00993810    ----------------------------------------------------MSRQLTD   7
BAD59613      ------------------------------------------------M----TTELLSD   8
```

Figure S8a: Multiple sequences alignments of CpsD with PCD/DCoH from other members of PCD/DCoH superfamily. Conserved residues were shaded in green.

```
CpsD          EEV-------REELLKV-D------KWMVKD-----E----K---WIERKYMFSDYLKGV   41
ZP00628055    -------------MTDD-F------QGWERR-------DKPP---TLFRRFAFAQYAQTR   30
ZP00400454    DAV-------D-AGLRGL-------DGWDAQ-------QGKT---RLHKHLRFDDFVAAM   135
CAE78833      EEI-------Q-QYLTVL-------DGWSLQ---------GL---HIAKSFEFKNYYQTI   54
NP840178      EEA-------E-KLLKQLE------QGWQLA-------DN-----KISRTFSFKNYYQTM   58
AAU28068      EQI-------K-NLLPQLN------TKWEVT-------EDNR---IIKRAFSFKNFYETM   58
Q9PAB4        ARL-------A-ELLPQV-------PGWELS-------NNGH---ALTRTFRFDNYYRTL   59
ABB41227      PRI-------E-SYLSQM-------PGWDVP-------LDYQ---TLTKTFSFKNYHQTV   59
BAD72380      DSA-------K-KMLEQ-V------GGWELA-------TEGD-ILKLHRAWKVKNFVKGL   123
NP174274      QSA-------Q-DLLQK-V------AGWDLA-------NDND-TLKLHRSWRVKSFTKGL   121
XP641894      NSK-------I-SLLKNID------KDWKLS-------DDSK---KIFRNWKI-PFSKSV   123
YP436101      AKL-------E-KYLSQ-V------PEWRLE-------ERNG-VQMISRDYKFKNFALAL   57
Q8KFI4        EEL-------Q-RQLSS-L------PEWTLV-------DDSG-TSKLVRVFTFKDFQSAL   58
BAC88867      AEI-------A-ELHPQ-I------PAWRIV-------EIEG-TPRLERQFRLRDFREAI   57
AAZ21144      SEI-------H-KYQKK-V------DGWDVK-------KNTKEIYFLEKNFIFKNFVNSQ   59
ZP00798856    DEI-------K-RYISQLH------EEWKVI-------ND-H---HLEREFKFKNFKEAL   56
ZP01012805    DEL-------E-ALKAQID------PAWQVD---------ET---VLTRRYETKNFAKAL   52
ZP00864663    AAA-------Q-RYLAE-V------PGWELT-------HAGT---RIERHFKTGDFATAL   55
AAU93190      EEA-------E-KLLVH-V------PRWELK-------DAAT---KLKRTFRFENFMEAL   58
CAI06204      AEA-------K-RQLAQ-T------PAWTLS-------DDGR---CIERSFTFDDFKDAM   57
AAZ59799      AEA-------E-ALLVE-T------PGWTLA-------DDAG---RLERSFTFRNFAQAL   57
ABB53415      ----MKLFDKT-KINSL-I------PSWNYKVKP----DCYN---YIQRKIKFPTFNEAC   41
BAE59828      PEL----------NSLL-E------QGWAL--D-----EDGM---GVKKTYYFKTYFKAV   50
AAN51894      A---------D-LELFK-Q------KGWEIKFR-----TEIP---LLSKTFLFPTYLSGL   80
AAK47768      REL-------S-ERLTA-L------PGWEL---------VDG---KLRHTFGFGSFDQSM   60
NP982992      PEL-------Q-AGLAA-L------PHWRL---------VGA---SLQRELRLADFEATW   48
CAG98922      KLL-------T-QELRK-L------PKWRL---------MEN---ELVRDYKFRDFEETW   108
AAK42360      KEL----------EELK-I------NGWIV--L-----ENGK---KIKKEFRFKDFKQSV   56
ZP00520697    SEI-------Q-SALRE-R------SAWTV---------VNG---KLHREYKFADFIHAF   40
ZP00774037    SETQTKLIALN-QGLPA-E------QHWHI---------NDG---KLAKNFKFKSFIRAF   55
AAZ59312      NQL-------D-SFIEK-N------PSWII---------DNK---TIKKEFKFENFIEAF   46
AAP99535      QEI-------Q-ELRKS-L------PTWEE---------AEG---KLVKRFKFTNFIEAF   41
AAP06387      EQL-------L-SPLLN-T------HHWELCRN-----NSRD---AIQRSFVFKDFDAAF   46
Q8XU38        KAL-------F-AE----L------PGWSL--Q-----NDRD---AIHKRFTFTDFNAAF   42
AAN29038      SEM-------N-EALRA-L------DGWQK--V-----DGRE---AITRSFKFKDFSTAF   43
YP428662      PAL-------D-RQLAD-H------PDWTI--T-----SART---ALTRSFGFKDFSEAF   42
ZP00548450    EEV-------T-RRLAA-L------DGWVKAT-------DRE---EIHKTFPVE-YYAAI   43
CAG90387      KQL-------E-EQISR-LNCGSHKNYWKIRSSLAADGSVEN---RLEVDYKFKSFAKTW   62
EAN85354      PAI-------T-QALQE-L------QGWRME-----GNMNKG---AICRDFEFRDFKQAM   96
ZP00412175    TQI-------D-AALTE-L------SDWLYQ-----P----G---ALVTVFKAPTAAAAL   43
O86722        QEV-------E-ERLAT-L------PGWSLD-----A----G---RLTRSYRLGSHFAAA   41
AAZ18005      QQV-------D-LQLEE-L------PGWQRD-----G----N---AIVKTYHFSDFVEAM   39
ZP00950887    QQI-------E-NALKE-L------KGWEFG-----D----D---AIHTSFEFNNFKEAF   38
ABC46318      DAI-------D-DALAD-L------DGWSHA-----D----D---KLHKTYEFSDFREAI   74
AAY42130      EAI-------E-AALAE-L------PGWTYA-----D----D---RLQKTYTFGSFREAV   42
ZP00396203    GDV-------LDRM--P-E------GWW-GD-----A----G---KIGRDFGFDTYQAGV   54
AAG19997      DTI-------SDRL--P-D------DWI-HD-----G----D---AITRTYTFEEYLDGV   38
ZP00658229    P----------QVQAEG-L------DDWRFF-----L----M---KLHARFETGSFTKGL   41
BAB97857      DGA-------MNSLFDV-S------PHWSSA-----N----A---KLTAHFNTGKFSTGM   41
ZP00283397    EEI-------QQRLVGP-L------QHWYLE-----E----G---WLRRKYRTEGWKGTL   44
ZP00571129    SAV-------S-TALEA-L------PGWAGD-----A----D---RIRLEILVDG-DESR   39
AAK45453      EQV-------D-AALHD-L------NGWQRA-----G----G---VLRRSIKFPTFMAGI   46
ZP00993810    EEI-------E-RQLAD-L------PLWTRE-----G----D---TIVATIEAPDFPAAI   40
BAD59613      EQI-------A-TALQD-L------PDWTRS-----G----D---EISRTVQAESFPAAI   41
                                                                   :
```

Figure S8a: Multiple sequences alignments of CpsD with PCD/DCoH from other members of PCD DCoH superfamily. Conserved residues were shaded in green.

```
CpsD          EFVSEAAKLSEEHNHHPFILI-QY-KA-VIITLSSWN----AKG-LTKLDFELAKQFDEL    93
ZP00628055    AFLDALSLLSEEMRHHPQNINFGT-TY-VNVTLGAAD----DAT-LSAEDELFARRIAAL    83
ZP00400454    KFVNAMADLAEAEGHHPDFCV-LY-AT-VDVTLWTHA----VGG-LSENDFILAAKLDRL    187
CAE78833      AFVNAIAFIVHTEDHHPELEV-GY-NR-CVVKFYTHSVNEGLGG-ISENDFICAAKIDAL    110
NP840178      AFVNAIAWVSHREDHHPDMMV-GY-DW-CRVEYMTHA----IGG-LSENDFICAAKVDML    110
AAU28068      AFVNAIAWIANIENHHPDLEV-GY-NY-CRVHFMTHA----LNG-LTHNDFICAAKIDKL    110
Q9PAB4        AFVNALAFIAHCEDHHPDMSV-HY-GR-AVVCFSTHK----IGG-ISENDFICAAKTSAL    111
ABB41227      AFVNAITWVAHKEDHHPEICF-GY-NE-CKVILTTHS----IKG-ISQNDFIMAAKIDAL    111
BAD72380      EFLQLVAAVAEEEGHHPDLHLVGW-NN-VKIDVWTHS----VRG-LTDNDFILAAKINNL    176
NP174274      DFFQRVADIAESEGHHPDLHLVGW-NN-VKIEIWTHA----IGG-LTENDFILAAKINEL    174
XP641894      EYLNDISKIADEEGHHPDISI-ESFWN-FKITIYTHF----IND-LTENDYILASKIDNL    176
YP436101      EFTNKVGAIAEEINHHPELVT-EW-GK-VRVTWWSHT----IKG-LHELDFAMAKRCEAV    109
Q8KFI4        DFTNRVGQLAEAEGHHPALLT-EW-GK-VTVSWWTHA----IGG-IHLNDVIMATKTEKL    110
BAC88867      AFTVRVGEEAEAEGHHPALLT-EW-GS-VKVSWWTHA----IAG-LHRNDFVMAAKTDAI    109
AAZ21144      NFINKVGEISENENHHPDISF-GW-GY-AKIIITTHA----IEG-LSENDFILAAKIDQI    111
ZP00798856    SYTNVIGQLAEKEGHHPDMLL-SW-GK-VKITLFTHK----IDG-LSESDFVFAAKVDKQ    108
ZP01012805    MLVNGVGYLAEAQGHHPDIKL-GW-GY-AEVSFTTHS----VGG-LSENDFICAAKLDAM    104
ZP00864663    GFTQTVGELAEQEDHHPQITL-GW-GF-VQVELYTHK----IGG-LHENDFILAAKINQA    107
AAU93190      DFARKVGELCEAEGHHPDIGI-GW-GY-CRVEFQTHK----ING-LHENDFIMAAKVDEL    110
CAI06204      SFVAKLGELAETEGHHPDICF-GW-GW-ARVTWQTKK----ING-LHDNDFIMAAKTDGL    109
AAZ59799      EFVSGVGRLAEEQGHHPEISF-GW-GH-ATVSWRTKK----IKG-LHRNDFVMAAKTSEL    109
ABB53415      TFLNKLFEENKKLDHHCKYIS-DY-NK-IKIKIYTHT----SKD-VTEKDIQLAQIIDDI    93
BAE59828      SFVNVVASQSAAKKHHPTITV-RI-GS-VDIHWTTHQ----PRG-LTDKDLTMAQHCDEA    102
AAN51894      EFVNSLAHIAERLDHHPDLFL-SY-RK-VTVEIFTHS----KNT-ITDLDLRFAEETETI    132
AAK47768      KFVAKIAAIADKFNHHPDICV-HN-KRSVRLTCWTRQ----MHC-LTRVDFDLAEAFSAV    113
NP982992      GLLTQVAMRAHLWGHHPTITT-TH-TR-ATIALTTHD----AGG-VTDIDLRMARRIERL    100
CAG98922      SFLNKVAMRSHLWGHHPTITT-TY-NR-VQFRLTTHD----VSG-ISDADIMMASRIEKY    160
AAK42360      DFLKDIQPSADALDHHPDVCV-YY-NR-VVVELTTHD----VGG-LTDLDYKLAIKLDEL    108
ZP00520697    GFMTCAALSAEAMNHHPEWFN-VY-NR-LTIDLTTHD----AGG-ITAKDFQLAAKLDAL    92
ZP00774037    GWMSQIAIWAEKLNHHPEWFN-VY-NK-VEVKLTTHD----VGG-LSELDFKLASKMDLL    107
AAZ59312      GFMSKVALLSEKIDHHPDWQN-IY-NK-VKINLTTHD----KGG-ITTNDIKLAEAIDKL    98
AAP99535      FMTKIAIISESLSHHPEWTN-IY-SE-VIIRLSTHD----MGG-ITMLDYKLAKAIDAI     93
AAP06387      DFMTKIASKSKVMNHHPEWSN-VY-NK-VDILLTSHD----VGG-ISKRDVDLANFINDA    98
Q8XU38        GFMTRVALKAEQVNHHPEWFN-VW-NR-VDITLSTHD----ANG-LTHRDADLARFIEQA    94
AAN29038      GFMAQAALYAEKLDHHPEWFN-AY-NR-VDVTLATHS----ENG-VTELDIKMARKMNAI    95
YP428662      GFMARVALEAQAQDHHPDWSN-SY-NR-VDITLSTHD----SGG-LSAKDFALAKAIDRI    94
ZP00548450    EALGTVAEAAKELEHHPDVEL-HW-GE-LTFSLTTYS----AGQRITDLDFRLVDRIEAV    96
CAG90387      SFLNLIAYHADNVKHHPSIDT-TY-NK-VNIKLTTHD----AGNRVTYNDLKFAQFVRDE    115
EAN85354      AFMNAVAVDCERMGHHPSWTN-TY-NR-LQVQLTTHG----SGNRVTQKDIDLARRMNDV    149
ZP00412175    ELIAAVGRLAEEQNHHPDLDW-RY-NR-VFIRFSSHD----AGTRVTGRDIAAAAAVSRA    96
O86722        AMVVHVAQVQEELDHHSDLTL-GY-HT-VALAVHTHS----AGGAVTEKDVELARRVEDL    94
AAZ18005      SFMNQAAFHAEALEHHPEWSN-AY-NV-VEVRLTTGD----TGG-ITSHDVRLAKRMEHI    91
ZP00950887    SVMTRIAFEAEAQQHHPEWTN-VY-NE-LSITLSTHD----AGG-VTEKDIEMAKTIEDI    90
ABC46318      SFVVRLSFYAEEMMHHPELEN-VY-NT-VSIALTTHD----AGGKVTEMDVELASQIEEL    127
AAY42130      SFIVRIAFEAEQLNHHPELHN-VY-NR-VTLALTTHA----AGNRVTARDVELARAIERI    95
ZP00396203    DFVVRVAALAEARGHHPDIHL-YY-RR-VRLTFFTYE----AGG-VTQLDLDAARAVNAL    106
AAG19997      AFASEVGDLADEAFHHPEITI-RY-DE-VEVRFTDHE----AGG-VTSQDIELARRTDDR    90
ZP00658229    ELVTRITEAAEAANHHPDVVL-TY-PQ-VDVDLQSHD----VHG-VTSRDVDLARRISEI    93
BAB97857      KFVNLIADSAEEANHHPDILL-TY-GF-VEITLTSHD----VGE-ITDRDVALAKVIDAH    93
ZP00283397    MVVNAVGHLAEAAWHHPDLTV-SY-AF-VTVKLKTHS----AKG-ITDKDFALASKIESF    96
ZP00571129    AVVDEVMREADAMDHHPVVEQ-GP-GT-TTFTVWTHS----AGG-VTELDIELARKISAI    91
AAK45453      DAVRRVAERAEEVNHHPDIDI-RW-RT-VTFALVTHA----VGG-ITENDIAMAHDIDAM    98
ZP00993810    RIVAEAAEVAEEMNHHPDIDI-RW-RT-TSWLLTTHD----AGG-LTQLDIEQAHRINEI    92
BAD59613      ALVDRVAEAAERAGHHPDIDI-RW-RT-VTFTLSTHS----AGG-LTGRDIDLARQIDEL    93
                                :*                            :   *   .
```

Figure S8a: Multiple sequences alignments of CpsD with PCD/DCoH from other members of PCD DCoH superfamily. Conserved residues were shaded in green.

```
CpsD         FLQNENAIIRK-------------------------------------------------   104
ZP00628055   AERGE-----------------------------------------------------   88
ZP00400454   REAAAARR--------------------------------------------------   195
CAE78833     AGNQFAPMSH-------------------------------------------------   120
NP840178     FKS--------------------------------------------------------   113
AAU28068     LVD--------------------------------------------------------   113
Q9PAB4       YEQGI------------------------------------------------------   116
ABB41227     LD---------------------------------------------------------   113
BAD72380     NLEGLLSKKATVQK---------------------------------------------   190
NP174274     QVEDLLRKKKVAK----------------------------------------------   187
XP641894     NIKSIEPRRKKITK---------------------------------------------   190
YP436101     FNA--------------------------------------------------------   112
Q8KFI4       V----------------------------------------------------------   111
BAC88867     AAQVGAV----------------------------------------------------   116
AAZ21144     FNV--------------------------------------------------------   114
ZP00798856   QSE----------------------------------------------------------   111
ZP01012805   GLAG---------------------------------------------------------   108
ZP00864663   WDQARR-------------------------------------------------------   113
AAU93190     AAEA-------------------------------------------------------   114
CAI06204     APT--------------------------------------------------------   112
AAZ59799     AAGMTQG----------------------------------------------------   116
ABB53415     LKCHNHQIIEKNQK---------------------------------------------   107
BAE59828     AELMGAVEKDQGKKCGPS----------------------------SPTPSSPGL-----   129
AAN51894     IRQG-------------------------------------------------------   136
AAK47768     HDEQCSQ---QVAR---------------------------------------------   124
NP982992     LARSECI---S------------------------------------------------   108
CAG98922     IKQIDPK---KGILIGDK-----------------------------------------   175
AAK42360     YKMKTS-----------------------------------------------------   114
ZP00520697   AAG------------------------------------------------------   95
ZP00774037   QV-------------------------------------------------------   109
AAZ59312     INS--------------------------------------------------------   101
AAP99535     KYE--------------------------------------------------------   96
AAP06387     AFEYQAK----------------------------------------------------   105
Q8XU38       AQLTGAK----------------------------------------------------   101
AAN29038     AG---------------------------------------------------------   97
YP428662     VG---------------------------------------------------------   96
ZP00548450   LARYVQAEA----------AG-----------------------------------   107
CAG90387     FERERPQNVLKMEELIKDARGQFSFSQASQIIDDLVASDKSSNLEKASKHSDPSKNSNGK   175
EAN85354     FREISVSQ---------------------------------------------------   157
ZP00412175   AAAVSASA----------EPGKYPPPGASRS-----------------------------   117
O86722       AAGHGAH----------------------------------------------------   101
AAZ18005     VQPKCL-----------------------------------------------------   97
ZP00950887   IDAD-------------------------------------------------------   94
ABC46318     A----------------------------------------------------------   128
AAY42130     AWVK-------------------------------------------------------   99
ZP00396203   WDELSSTGGTGA----------------------------------------------   118
AAG19997     R----------------------------------------------------------   91
ZP00658229   AAELGVESAPRDVSTLELALD---VPDAGAVKPFWRAVLGYQDNQDWPEVMDPGGRNNTL   150
BAB97857     AKTLAISAEA-------------------------------------------------   103
ZP00283397   IQWQPATEGGPLEGTPTD----------DQ---RFR-YIK------YDKPKDPGSD----   132
ZP00571129   LRSAGIAF-------------------------------------------------   99
AAK45453     FGA--------------------------------------------------------   101
ZP00993810   ARRHGVSFGG--------------------------------------------------   102
BAD59613     AR---------------------------------------------------------   95
```

Figure S8a: Multiple sequences alignments of CpsD with PCD/DCoH from other members of PCD/DCoH superfamily. Conserved residues were shaded in green.

```
CpsD         ------------------------------------------------------------   104
ZP00628055   ------------------------------------------------------------   88
ZP00400454   ------------------------------------------------------------   195
CAE78833     ------------------------------------------------------------   120
NP840178     ------------------------------------------------------------   113
AAU28068     ------------------------------------------------------------   113
Q9PAB4       ------------------------------------------------------------   116
ABB41227     ------------------------------------------------------------   113
BAD72380     ------------------------------------------------------------   190
NP174274     ------------------------------------------------------------   187
XP641894     ------------------------------------------------------------   190
YP436101     ------------------------------------------------------------   112
Q8KFI4       ------------------------------------------------------------   111
BAC88867     ------------------------------------------------------------   116
AAZ21144     ------------------------------------------------------------   114
ZP00798856   ------------------------------------------------------------   111
ZP01012805   ------------------------------------------------------------   108
ZP00864663   ------------------------------------------------------------   113
AAU93190     ------------------------------------------------------------   114
CAI06204     ------------------------------------------------------------   112
AAZ59799     ------------------------------------------------------------   116
ABB53415     ------------------------------------------------------------   107
BAE59828     ------------------------------------------------------------   129
AAN51894     ------------------------------------------------------------   136
AAK47768     ------------------------------------------------------------   124
NP982992     ------------------------------------------------------------   108
CAG98922     ------------------------------------------------------------   175
AAK42360     ------------------------------------------------------------   114
ZP00520697   ------------------------------------------------------------   95
ZP00774037   ------------------------------------------------------------   109
AAZ59312     ------------------------------------------------------------   101
AAP99535     ------------------------------------------------------------   96
AAP06387     ------------------------------------------------------------   105
Q8XU38       ------------------------------------------------------------   101
AAN29038     ------------------------------------------------------------   97
YP428662     ------------------------------------------------------------   96
ZP00548450   ------------------------------------------------------------   107
CAG90387     ------------------------------------------------------------   175
EAN85354     ------------------------------------------------------------   157
ZP00412175   ------------------------------------------------------------   117
O86722       ------------------------------------------------------------   101
AAZ18005     ------------------------------------------------------------   97
ZP00950887   ------------------------------------------------------------   94
ABC46318     ------------------------------------------------------------   128
AAY42130     ------------------------------------------------------------   99
ZP00396203   ------------------------------------------------------------   118
AAG19997     ------------------------------------------------------------   91
ZP00658229   WFQEAPDATGEVQQRFHLDIVVPREVAEERVAAAVAAGGRLVSEDAVPAFWVLADAHGNK   210
BAB97857     ------------------------------------------------------------   103
ZP00283397   ------------------------------------------------------------   132
ZP00571129   ------------------------------------------------------------   99
AAK45453     ------------------------------------------------------------   101
ZP00993810   ------------------------------------------------------------   102
BAD59613     ------------------------------------------------------------   95
```

Figure S8a: Multiple sequences alignments of CpsD with PCD/DCoH from other members of PCD/DCoH superfamily. Conserved residues were shaded in green.

```
CpsD            --------------    104
ZP00628055      --------------    88
ZP00400454      --------------    195
CAE78833        --------------    120
NP840178        --------------    113
AAU28068        --------------    113
Q9PAB4          --------------    116
ABB41227        --------------    113
BAD72380        --------------    190
NP174274        --------------    187
XP641894        --------------    190
YP436101        --------------    112
Q8KFI4          --------------    111
BAC88867        --------------    116
AAZ21144        --------------    114
ZP00798856      --------------    111
ZP01012805      --------------    108
ZP00864663      --------------    113
AAU93190        --------------    114
CAI06204        --------------    112
AAZ59799        --------------    116
ABB53415        --------------    107
BAE59828        --------------    129
AAN51894        --------------    136
AAK47768        --------------    124
NP982992        --------------    108
CAG98922        --------------    175
AAK42360        --------------    114
ZP00520697      --------------    95
ZP00774037      --------------    109
AAZ59312        --------------    101
AAP99535        --------------    96
AAP06387        --------------    105
Q8XU38          --------------    101
AAN29038        --------------    97
YP428662        --------------    96
ZP00548450      --------------    107
CAG90387        --------------    175
EAN85354        --------------    157
ZP00412175      --------------    117
O86722          --------------    101
AAZ18005        --------------    97
ZP00950887      --------------    94
ABC46318        --------------    128
AAY42130        --------------    99
ZP00396203      --------------    118
AAG19997        --------------    91
ZP00658229      VCVCTADGRETAGE    224
BAB97857        --------------    103
ZP00283397      --------------    132
ZP00571129      --------------    99
AAK45453        --------------    101
ZP00993810      --------------    102
BAD59613        --------------    95
```

Figure S8a: Multiple sequences alignments of CpsD with PCD/DCoH from other members of PCD/DCoH superfamily. Conserved residues were shaded in green. PCD/DCoH from other members of PCD/DCoH superfamily were from *Schistosoma japonicum* (AAP06387), *Peptoclostridium acidaminophilum* (AAN86540), PHS_CHLTE (Q8KFI4), PHS_RALSO (Q8XU38), PHS_XYLFA (Q9PAB4), *Nitrosomonas europaea*

ATCC 19718 (NP840178), *Arthrobacter* sp. FB24 (ZP00412175), *Janibacter* sp. HTCC2649 (ZP00993810), *Nocardia farcinica* IFM 10152 (BAD59613), *Frankia* sp. EAN1pec (ZP00571129), *Frankia* sp. CcI3 (ZP00548450), *Nocardioides* sp. JS614 (ZP00658229), *Corynebacterium glutamicum* ATCC 13032 (BAB97857), PHS_STRCO (O86722), *Leptospira interrogans* serovar *Lai* str. 56601 (AAN51894), *Aspergillus oryzae* RIB40 (BAE59828), *Mycobacterium tuberculosis* CDC1551 (AAK47768), *Nitrobacter hamburgensis* X14 (ZP00628055), *Arabidopsis thaliana* (AAM66965), *Dictyostelium discoideum* AX4 (XP641894), *Oryza sativa Japonica* Group (BAD72380), *Arabidopsis thaliana* (NP174274), *Brucella suis* 1330 (AAN29038), *Candidatus Pelagibacter ubique* HTCC1062 (AAZ21144), *Hahella chejuensis* KCTC 2396 (YP436101), *Alkaliphilus metalliredigenes* QYMF (ZP00798856), *Hydrogenovibrio crunogenus* XCL-2 (ABB41227), *Anaeromyxobacter dehalogenans* 2CP-C (ZP00400454), *Alkalilimnicola ehrlichei* MLHE-1 (ZP00864663), *Maritimibacter alkaliphilus* HTCC2654 (ZP01012805), *Legionella pneumophila* subsp. *pneumophila* str. Philadelphia 1 (AAU28068), *Gloeobacter violaceus* PCC 7421 (BAC88867), *Aromatoleum aromaticum* EbN1 (CAI06204), *Methylococcus capsulatus* str. *Bath* (AAU93190), *Debaryomyces hansenii* CBS767 (CAG90387), *Cupriavidus pinatubonensis* JMP134 (AAZ59799), *Salinibacter ruber* DSM 13855 (ABC46318), *Trypanosoma cruzi* (EAN85354), *Solibacter usitatus* Ellin6076 (ZP00520697), *Prochlorococcus marinus* str. NATL2A (AAZ59312), *Eremothecium gossypii* ATCC 10895 (NP982992), *Kluyveromyces lactis* (CAG98922),*Pseudoalteromonas atlantica T6c* (ZP00774037),Rhodothermus marinus (AAY42130), *Prochlorococcus marinus* subsp. *marinus* str. CCMP1375 (AAP99535), Psychrobacter arcticus 273-4 (AAZ18005), *Croceibacter atlanticus* HTCC2559 (ZP00950887), *Bdellovibrio bacteriovorus* HD100 (CAE78833), *Plasmodium falciparum* (ABB53415), *Deinococcus geothermalis* DSM 11300 (ZP00396203), *Burkholderia fungorum* LB400 (ZP00283397), *Halobacterium salinarum* NRC-1 AAG19997), *Mycobacterium tuberculosis* CDC1551 (AAK45453), *Rhodospirillum rubrum* ATCC 11170 (YP428662), *Saccharolobus solfataricus* P2 (AAK42360)

```
CpsD          ------------------------------------------------------------  0
XP018653366   ------------------------------------------------------------  0
XP004996804   ------------------------------------------------------------  0
XP002431808   ------------------------------------------------------------  0
EPB90052      ------------------------------------------------------------  0
JAI53288      ------------------------------------------------------------  0
XP012997458   -------------------------------------------------------MAAALG-  6
XP010592093   ------------------------------------------------------------G-  1
ETN63961      -MVFVCTPCPCFRMVSINNSFLAGLRLS----------------------------RLSA  31
XP003378337   ------------------------------------------------------------  0
XP015375611   ------------------------------------------------------------  0
XP015791586   MNISFALRVK----SSKSVTFRLLSDLDHHF--------HPRSIQQISSVSKLVKPV---  45
XP799142      ------------------------------------------------------------  0
XP004365728   ------------------------------------------------------------  0
XP005995953   ------------------------------MWEERRSQVMS-----------WLMFTT---  17
ERG82286      ------------------------------------------------------------  0
XP003091027   ----------------------------------MRHLFISPFFQR---------------  12
XP017846937   ------------------------------------------------------------  0
XP011061072   -------------------------------------------------------MIAVL--  5
XP002194428   MALLYHMPTPCRSTSCAPGEERGAKRIAQGFPSKRPCLATPSPHRQVASAGKALLAVGGP  60
XP002399841   ------------------------------------------------------------  0
XP002596035   ------------------------------------------------------------  0
XP008428253   ------------------------------------------------------------  0
XP003963732   ------------------------------------------------------------  0
A5GMR6        ------------------------------------------------------------  0
WP014881094   ------------------------------------------------------------  0
WP014746679   ------------------------------------------------------------  0
Q0K5I5        ------------------------------------------------------------  0
WP023598290   ------------------------------------------------------------  0
WP014751602   ------------------------------------------------------------  0
WP013299844   ------------------------------------------------------------  0
CBN79409      ------------------------------MSSGRRCCY----------VGNAQAAVVA-  19
Q1GS35        ------------------------------MSSR--------------------------  4
WP013433887   ------------------------------------------------------------  0
Q13SM2        ------------------------------------------------------------  0
Q63PF1        ------------------------------------------------------------  0
WP025294361   ------------------------------------------------------------  0
WP015436031   ------------------------------------------------------------  0
A5V667        ------------------------------------------------------------  0
A5FWK8        ------------------------------------------------------------  0
WP015165790   ------------------------------------------------------------  0
Q54RY8        ------------------------------------------------------------  0
XP004359573   ------------------------------------------------------------  0
EFA77375      ------------------------------------------------------------  0
WP014102349   ------------------------------------------------------------  0
Q89WZ6        ------------------------------------------------------------  0
WP013651003   ------------------------------------------------------------  0
WP012568145   ------------------------------------------------------------  0
Q11D07        ------------------------------------------------------------  0
A6WV21        ------------------------------------------------------------  0
```

Figure S8b: Multiple sequences alignments of CpsD with PCD/DCoH from other members of the Pterin_4a superfamily.

```
CpsD          ------------------------------------------------------------  0
XP018653366   ------------------------------------------------------------  0
XP004996804   -------------------------------------------------------MSDD  4
XP002431808   -----------------------------------MPFTLSLTGKDILGYR-FQSTPVK  23
EPB90052      ------------------------------------------------------------  0
JAI53288      ------------------RLTSLVNLRGVG------------V-VGEYCSYR-F-ASSVK  27
XP012997458   -------------------------------------ARRTRVRSLAALRGG-S-RSLAT  27
XP010592093   -------------------------------------PRRP-FPGAAAPGAR-T-LSLAA  21
ETN63961      GRSLAVP-----------ALSRTRW-TNPNLICADSRSSQTHFPRTSAHRTF-L-TTTVI  77
XP003378337   ------------------------M-RLLKLGAK-------NLFCARALPFS-S-SSTVC  26
XP015375611   -------------------MTAVFF-NNTLR--------------PVLLCFA-K-YSTKR  24
XP015791586   -------STI---KTSTSSLSSLVT-RNL---AT-------SSTSKQAKDLS-E-NSVKS  82
XP799142      ------------------------------------------------MRFIR-T-FARMA  11
XP004365728   ------------------------------------------------------------  0
XP005995953   ------------RAKVAAVLQSVLW-KRTSV------------------SVL-H-KNTAT  44
ERG82286      ----------------------------ML------------SASFRILPLS-R-RVVHT  18
XP003091027   ------------------LRTSVSF-LCVSPISS-------SLPPTSHRLFS-S-TIAVS  44
XP017846937   ------------------------------------------------------------  0
XP011061072   --------------TRRASGGDHIR-KCISR-----------ILDR----RD-L-SAAVS  33
XP002194428   GGLATPPGQSGEAPAPAGSGTESIP-RHGEH-----------PPARRASPAR-R-ARRAA  106
XP002399841   -------------------------------------------------------------L  1
XP002596035   ------------------------------------------------------------  0
XP008428253   --------------------MSVFW-AEMST-----------SPF-----CP-S-CPATL  21
XP003963732   ------------------------------------------------------------  0
A5GMR6        ------------------------------------------------------------  0
WP014881094   ------------------------------------------------------------  0
WP014746679   ------------------------------------------------------------  0
Q0K5I5        ------------------------------------------MICGLRFRY-------DQ  11
WP023598290   ------------------------------------------------------------  0
WP014751602   ------------------------------------------------------------  0
WP013299844   ------------------------------------------------------------  0
CBN79409      ---------L---STKKNS-S-----SCPI------------AVCKRE--------LRTG  41
Q1GS35        ----------------------------SR------------LVAHRALSYGDRTALTSG  24
WP013433887   ------------------------------------------------------------  0
Q13SM2        ------------------------------------------------------------  0
Q63PF1        ------------------------------------------------------------  0
WP025294361   ------------------------------------------------------------  0
WP015436031   ------------------------------------------------------------  0
A5V667        ------------------------------------------------------------  0
A5FWK8        ------------------------------------------------------------  0
WP015165790   ------------------------------------------------------------  0
Q54RY8        ------------------------------------------------------------  0
XP004359573   ---------M---TFSKASAI-----ILVL------------IICLSS-------FITAS  24
EFA77375      ------------------------------------------------------------  0
WP014102349   ------------------------------------------------------------  0
Q89WZ6        ------------------------------------------------------------  0
WP013651003   ------------------------------------------------------------  0
WP012568145   ------------------------------------------------M-------ATTQT  6
Q11D07        ------------------------------------------------------------  0
A6WV21        ------------------------------------------------------------  0
```

Figure S8b: Multiple sequences alignments of CpsD with PCD/DCoH from other members of the Pterin_4a superfamily.

```
CpsD          ---MMLRLTEEEVREELLKV---DKWMVK---DEKWIERKYMFSDYLKGVEFVSEAAKLS   51
XP018653366   ----MPLLTGPEREQMLPPLLNIHHWELCQHSSREAIRRSFLFKDFDVAFDFMKKIAEKS   56
XP004996804   PATKKAKLSHEERLDRLTKAQ-ERGWKLQ--KDRDAIEKSFNFKDFSEAWAFMSRVALKA   61
XP002431808   KGKMAGKLTQEQRDTLLKPLL-SSGWSMV--KDRDAIYREFVFKDFNEAFGFMTRIALKA   80
EPB90052      --MSVTTLTQAQRDQLLTPLL-SSGWTMV--ENRDAIIKKYSFQDFNEAFGFMTRVALKA   55
JAI53288      KRKMPSKLSDEERISLIKPLQ-EVGWVVL--DNRDAIRKEFVFKNFVKAFDFMKAVADEA   84
XP012997458   QPSDSQLLIAEERNQILLDLK-NVGCSEL--SERNAIYKEFSFKNFNQAFGFMSRVALQA   84
XP010592093   MSADAHRLTAEERNQVILDLK-AAGWSEL--SERNAIHKEFSFKNFNQAFGFMSRVALQA   78
ETN63961      YRKMLAKLTDAERAEKLKPLLDNAGWKLV--ENRDAIYKEYLFADFNQAWAFMSAVALKA   135
XP003378337   RKMPLCLLNDQEREQHLNELI-GQGWKLQ--EKRDAIQKLFTFGDFNEAFGFMTQIALKA   83
XP015375611   TRKMPILLTKEQREELLQPLF-SNQWSLV--KDRDAIYKEFLFSNFIEAFGFMTQVALKS   81
XP015791586   RKMAPKGLTSDERDQLLKPLL-ESGWKMD-DSGRDAIKKELVFKDFIQAFGFMSSVALKA   140
XP799142      SEAKRVKLSGDDRSQKLSALT-GAGWKEV--EGRDAIQKEFTFGNFNQAFGFMTRVALKA   68
XP004365728   ---MLQKLTPEARTAKLAQL---PGWNEV--AGRDAIRKDFVFADFNQAFGFMTRIALKA   52
XP005995953   MSSDSQWLTADERGQLLLELK-AAGWVEL--EDRDAIYKEFLFKSFNQAFGFMTRVALQA   101
ERG82286      KPKLLMSLDTEERKQLLEPLF-ANGWTMV--DGRDAIRKNFQFKNFNEAFAFMVRVALQA   75
XP003091027   ARKKMPLLTESERNDQLSGLK-SAGWKLV--EGRDAIQKEFQFKDFNEAFGFMTRVGLKA   101
XP017846937   ---MVAKLTEEERKEKLQPLL-DAGWKMV--EGRDAIYKEFLLKDFNQAFSFMTGVALLA   54
XP011061072   KKTKMSKLSPEEREQNLSSLL-STGWTVQ--ANRDAIYKEFVFKNFNEAFGFMTRVALQA   90
XP002194428   GHGKAHRLSTEEREQLLPNLR-AVGWNEV--EGRDAIFKEFHFKDFNRAFGFMTRVALQA   163
XP002399841   QSSKRAKLTEEERKTKLAPLL-SAGWTTV--KDRDAIYKEFLFKNFNQSFGFMTRIAMQA   58
XP002596035   -------LTDSERESQLEPLM-GKGWTMV--DGRDAIYKEFLFKDFNQAFGFMSRVALRA   50
XP008428253   QAGKIQTLSDEERAHLLPLLR-NAQWVEA--VGRDAIYKEFIFKDFNQAFGFMSRVALQA   78
XP003963732   MAGKIQSLTEEERAHLLPLLH-NAQWVEV--VGRDAIYKEFIFKDFNQAFGFMSRVALQA   57
A5GMR6        ---MASLLPQSERETLSNTL---PHWQV--EA--GRLKRNWQFKDFSEAFAFMTRVALLA   50
WP014881094   ---MTEKLSDATRGPLLDPLF-SAGWQV--VNGRDAITKTYKFDSFVDAFGWMTRAAIWA   54
WP014746679   ---MVEKLDPAARAELLASL---SGWTE--VEGRDAIEKTFRFTDFKAAFGFMTRVAIAA   52
Q0K5I5        IEPPMTPLSPQARATLLADL---PGWTT--VPDRDAIFKRFTFHDFNAAFGFMTRVAIQA   66
WP023598290   ---MTTRLSSEARASLAEVL---PEWHQ--VVGRDAIQRSFTFHDFREAFGFMVQVALSA   52
WP014751602   ----MTPLNDADIGIALASL---DGWGL--DEDGKALRRHFKFDSFVQAFGFMTKVALLA   51
WP013299844   ----MDKLSAEDRKRFLADH---PNWTL--VDGRDAMSASFRFENFIDAFGFMTQIALRA   51
CBN79409      KHV---FTDARERETAIATL---SGWAEA-GEGRDAISKSFCFSDFNQAFAFMTRAALVA   94
Q1GS35        RKTMVQKLDDAQRSALLARF---PQWTH--DPVRDAITRQFRFDDFAQAFGFMASVAIVA   79
WP013433887   ---MIQKLSTDERAKLGSTL---PAWQP--VSGRDAIRRQLQFADFNAAFGFMTRVAIKA   52
Q13SM2        ---MIHKLTSEERATQIAAL---HGWQA--ATGRDAIQRQFKFADFNEAFGFMTRVAIKA   52
Q63PF1        ---MIHKLTSEERKTQLESL---HHWTA--VPGRDAIQRSLRFADFNEAFGFMTRVAIKA   52
WP025294361   ---MVEQLDEDARARALDDL---DEWDY--DVSRDAITRSLVFDDFVTAFGFMTQVALLA   52
WP015436031   ---MISKLTASERQAALAAL---PLWAE--IPDRDAIHRSLRFKDFNAAFAFMTRVALMA   52
A5V667        --MSIEPLTDEERADALDAL---PDWDY--DDGRDAISRSFTFPDFSAAFAFMTRVALYA   53
A5FWK8        ---MIPRLTDAERAALVDLL---PEWSL--AKDRDAIERRFAFADFSEAFAFMTRVALLA   52
WP015165790   --MAVEKLSPQAINDGLSTL---TGWEIK-N---EKLNKSFKFKDFNEAFGFMTRIALVA   51
Q54RY8        --MAPTLLTEEQRKELIP-----KDWEMV-V-GRDAIKKTFTFKDFNQAFSFMTRVALVA   51
XP004359573   EMAVPVKLTAEERPALLATI---PQWAMV-EGGRDAIKRTYLFADFNQAFSFMTRVALVA   80
EFA77375      --MAPIKLTEEERATLLPTV---PSWSMV-E-GRDAIKRTYIFKDFITAFSFMTRTALVA   53
WP014102349   --MVQTKMTRDAVSTALQEL---EGWALA-T-DKDALHKSFRFADFNEAWGFMARCALLA   53
Q89WZ6        ---MAERLTTEARKQALGGI---PDWTEV-S-GRDAIGKTFVFKDFNEAFGFMTRAALVA   52
WP013651003   ---MSARLTDAERLEALAGL---DAWTLL-A-GREAIRRTYAFADFNAAFGFMTRVALLA   52
WP012568145   GTLTGAALTGAARAEALAGL---AAWREV-E-GRDAIRRSFVFPDFNAAFGFMTRVALLA   61
Q11D07        --MARQKLDRKEIENGLSEL---DEWELA-E-GGNAIHRRFVFSNFNEAFGFMTRAALAA   53
A6WV21        --MARNRLTENELNEALTEL---DGWQKV-D-GREAIAKSFKFKDFNAAFGFMTRAALHA   53
                                                :     : .:  .  ::   .  :
```

Figure S8b: Multiple sequences alignments of CpsD with PCD/DCoH from other members of the Pterin_4a superfamily.

```
CpsD            EEHNHHPFILIQYKAVIITLSSWNAKGLTKLDFELAKQFDELFLQNENAIIRK 104
XP018653366     KVMNHHPEWFNVYNKVDILLTSHDAGGLSRRDIDLANFINDAAFEYQAK---- 105
XP004996804     EELCHHPEWFNVYNRLDVTLSTHDVSGLSAKDFTMAEFMDSLTGK-------- 106
XP002431808     DKMDHHPEWFNVYNKVQITLASHDVAGLSQRDITLANFVNQASKLGAS----- 128
EPB90052        DKLDHHPEWFNVYNRVEITLSTHDCQGLSTRDIELATFCDKAQK--------- 99
JAI53288        EKMNHHPEWFNVYNKVDVTLSSHDVNGLSKRDVKLAKLIDERASVYKE----- 132
XP012997458     EKINHHPEWFNVYK-VQLTLTSHDCGGLTKRDVQLAKFIEKASIFV------- 129
XP010592093     EKMNHHPEWFNVYNKVQITLTSHDCGGLTKRDVKLAQFIEKAAASM------- 124
ETN63961        EQMNHHPEWFNVYNKVQVTLATHDCGGLSERDVRLASFMDSKVPSK------- 181
XP003378337     EKMNHHPEWFNVYNKVDITLSTHDCNGLSMKDITLGKFIEGVAAVRKK----- 131
XP015375611     EKMDHHPEWFNVYNKVQVTLSTHDVGGLSSNDVNLATFLDKTEKKIKSSS--- 131
XP015791586     EKMNHHPEWFNCYNKLNILLSSHDVNGLSERDIKLANFIDKAWSQINK----- 188
XP799142        ETMNHHPEWFNVYNRVDITLSTHDVGGLSDNDVNMANFIEKAATSAK------ 115
XP004365728     DKMDHHPEWFNVYNQVHITLSTHDCGGLSERDITLATFIESVAAESRAK---- 101
XP005995953     EKMNHHPEWFNVYHKVQITLSTHSCAGLTKKDVKLAQFIEKAAASML------ 148
ERG82286        EKMDHHPEWFNVYNKVDVTLSSHDVNGLSQRDVKLAKFMDKAST--------- 119
XP003091027     EQMNHHPEWFNVYNKVDITLSTHDCGGLSQNDVKLATFIETVAKK-------- 146
XP017846937     EKMNHHPEWFNVYNKVQVTLSTHDVGGLSSQDIRMGNYFEAQAKLLNNC---- 103
XP011061072     EKMDHHPEWFNVYNKVNITLSSHDVNGLSQRDIKLATFIDKVAAPTGN----- 138
XP002194428     EKLDHHPEWFNVYNKVHITLSTHDCGGLSERDINLASFIEQVAASLS------ 210
XP002399841     EKMDHHPEWFNVYNKVQVTLSTHDVGGLSENDVKLANFIETAAKSMSWLQFLP 111
XP002596035     EKMDHHPEWFNVYNKVQITLSTHDVSGLSGRDITLATFIEKAASSIQ------ 97
XP008428253     EKMDHHPEWFNVYNKVQITLSTHDCGGLSQRDITLATFIDQASLM-------- 123
XP003963732     EKMDHHPEWFNVYNKVQITLSTHDCGGLSQRDITMATFMDQASLM-------- 102
A5GMR6          EAMQHHPNWSNVYNRVSIELTTHDLGGLSDLDVQLARSIDALC---------- 93
WP014881094     EKWDHHPEWSNVYNTVTVTLTTHDVGGISALDAKLARKMESLF---------- 97
WP014746679     DGQDHHPEWSNVYNRVTILLTTHDADGLSARDAKLARVIEEAAAGSGLST--- 102
Q0K5I5          EKADHHPEWFNVYNRVDITLSTHDANGLTQRDIDLAHFIERAADAVTD----- 114
WP023598290     EKMNHHPEWFNVYNRVDITLSTHDADGLTQRDIDLAREIDQFAGDRASS---- 101
WP014751602     EQANHHPEWSNVYNKVDIRLTTHDADGLTERDVKLAQAIDAL----------- 93
WP013299844     EKMDHHPEWSNVYNKVEITLTTHDVDGLSPRDTELAAFIDKLAP--------- 95
CBN79409        EKMDHHPEWFNVYNRVDVTLSTHDLGGLSRKDIRLAQRLDAFAKDAPEALP-- 145
Q1GS35          EKLDHHPEWSNVYNRVDVLLTTHDANGLSERDAKLAEAIEALL---------- 122
WP013433887     QEMNHHPEWSNVYNRVEITLATHEAGGITPRDIELARFIDSIADASCC----- 100
Q13SM2          QEMDHHPEWFNVYDKVEITLSTHEANGVTERDIRLAAFIDSITA--------- 96
Q63PF1          QEMNHHPEWFNVYNRVDVTLSTHDANGLTERDIKLAHFIDEVGKHAKAA---- 101
WP025294361     EKADHHPEWFNVWNRVDILLTTHDAGGLSQRDIDLAEKIDALYVPFEKD---- 101
WP015436031     EKMDHHPEWSNVYNRVDITLSTHDAGGLSERDLALAQFIDAAAPE-------- 97
A5V667          EKHDHHPEWSNVWNRVDILLTTHDAGGLSHRDVAMAEAIEALAE--------- 97
A5FWK8          EKHDHHPEWSNVYNRVTILLSTHDAGGLSARDIRLAEAIDGLA---------- 95
WP015165790     ESINHHPELFNVYNNVTIDLTTHDAGGISNLDLEFAKRVNSF----------- 93
Q54RY8          EQMNHHPEWFNVYNRVEITLATHDCSGLSVNDTKMADIMNQFFNQLHQ----- 99
XP004359573     EQMAHHPEWFNVYNRVEITLATHDCNGLSMNDINLAKTMDTFFGQYKL----- 128
EFA77375        EKMDHHPEWFNVYNKVDVTLATHDCSGLSKNDINLAKTMDEFFVQYTQ----- 101
WP014102349     EKMNHHPEWFNVWNRVDVTLNTHDAGGITALDIKMATAMNDYAKALNT----- 101
Q89WZ6          EKMDHHPEWRNVYKTVEVVLSTHDAGGVTALDIELARAMNAIAKLTPG----- 100
WP013651003     EKMDHHPEWFNVYRTVEVTLTTHDAGGLTLRDIDLARAMDRIAGQT------- 98
WP012568145     ETMNHHPEWSNVYNRVEVTLTTHDAGGVTALDVRMAGFMDGIAG--------- 105
Q11D07          EKLDHHPEWTNVYKTVDVTLTTHASGGLTELDFKLAKKMNAYAAK-------- 98
A6WV21          EKLDHHPEWFNVYNRVDVTLATHSENGITELDIKLARKMNAIA---------- 96
                .   ***     :  : : * :    *::  *  :.   :
```

Figure S8b: Multiple sequences alignments of CpsD with PCD/DCoH from other members of the Pterin_4a superfamily. The conserved residues were shaded in yellow. Sequences alignments was done using CLUSTAL O (1.2.4) multiple sequence alignment tool (Madeira et al., 2019) . PCD/DCoH from other members of the pterin_4a superfamily were

from PHS_SYNPW (A5GMR6), PHS_CHESB (Q11D07), *Cupriavidus necator* H16 (Q0K5I5), PHS_PARXL (Q13SM2), PHS_BURPS (Q63PF1), PHS_DICDI (Q54RY8), PHS_BRADU (Q89WZ6), PHS_OCHA4 (A6WV21), Q1GS35_SPHAL (Q1GS35), PHS_SPHWW (A5V667), PHS_ACICJ (A5FWK8), PHS_ACICJ (A5FWK8), *Caenorhabditis remanei* (XP003091027), *Ixodes scapularis* (XP002399841), *Ascaris suum* (ERG82286), *Taeniopygia guttata* (XP002194428), *Capsaspora owczarzaki* ATCC 30864 (XP004365728), *Anopheles darling* (ETN63961), *Polymorphum gilvum* (WP013651003), *Poecilia reticulate* (XP008428253), *Trichinella spiralis* (XP003378337), *Ectocarpus siliculosus* (CBN79409), *Branchiostoma floridae* (XP002596035), *Advenella kashmirensis* (WP014751602), *Strongylocentrotus purpuratus* (XP799142), *Pediculus humanus corporis* (XP002431808), *Diuraphis noxia* (XP015375611), *Micavibrio aeruginosavorus* (WP014102349), *Paraburkholderia rhizoxinica* (WP013433887), *Drosophila busckii* (XP017846937), *Latimeria chalumnae* (XP005995953), *Pseudanabaena* sp. PCC 7367 (WP015165790), Schistosoma mansoni (XP018653366), *Tetranychus urticae* (XP015791586), *Acromyrmex echinatior* (XP011061072), *Mucor circinelloides f. circinelloides* 1006PhL (EPB90052), *Rhodnius neglectus* (JAI53288), *Parvularcula bermudensis* (WP013299844), *Azoarcus* sp. KH32C (WP015436031), *Takifugu rubripes* (XP003963732), *Cavia porcellus* (XP012997458), *Loxodonta africana* (XP010592093), *Tistrella mobilis* (WP014746679), *Rhodospirillum centenum* (WP012568145), Phaeobacter inhibens (WP014881094),*Sphingomonas sanxanigenens* (WP025294361), *Cavenderia fasciculate* (XP004359573), *Heterostelium album* PN500 (EFA77375),*Salpingoeca rosetta* (XP004996804), Pandoraea (WP023598290).

CHAPTER EIGHT

The chapter has been ***published in International Journal of Biological Macromolecules***

(2020 Impact Factor 5.162)

International Journal of Biological Macromolecules 161 (2020) 247–257

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: http://www.elsevier.com/locate/ijbiomac

Cloning, overexpression, purification, characterization and structural modelling of a metabolically active Fe^{2+} dependent 2,6-dichloro-*p*-hydroquinone 1,2-dioxygenase (CpsA) from *Bacillus cereus* strain AOA-CPS_1

Oladipupo A. Aregbesola, Ajit Kumar, Mduduzi P. Mokoena, Ademola O. Olaniran *

Discipline of Microbiology, School of Life Sciences, College of Agriculture, Engineering and Science, University of KwaZulu-Natal (Westville Campus), Private Bag X54001, Durban 4000, Republic of South Africa

ARTICLE INFO

Article history:
Received 5 January 2020
Received in revised form 24 May 2020
Accepted 30 May 2020
Available online 05 June 2020

Keywords:
Pentachlorophenol
2,6-Dichloro-*p*-hydroquinone
2,6-Dichloro-*p*-hydroquinone 1,2-dioxygenase
Bacillus cereus

ABSTRACT

2,6-Dichloro-*p*-hydroquinone (DiCHQ) aromatic-ring cleavage by DiCHQ 1,2-dioxygenase (CpsA) is very crucial for complete transformation of pentachlorophenol (PCP) to 2-chloromaleylacetate in *Bacillus cereus* AOA-CPS_1 (*Bc*AOA). The 978 bp gene (*cpsA*) was detected and amplified in the genome of *Bc*AOA; cloned, overexpressed and purified to homogeneity. CpsA showed a single ≈36.9 kDa protein band on SDS-PAGE and exhibited optimum activity at 30 °C and pH 9.0. CpsA was stable between 20 °C and 40 °C, and also retained about 90% of its activity at 60 °C for 120 min. The enzyme retained about 90% activity between pH 9.0 and 11.5 and 60% activity at pH 13.0. CpsA was found to be Fe^{2+} dependent as about 90% increased activity was observed in the presence of $FeSO_4$. CpsA showed apparent v_{max}, K_m, k_{cat} and k_{cat}/K_m of 27.77 ± 0.9 μMs^{-1}, 0.990 ± 0.03 mM, 4.20 ± 0.04 s^{-1} and 4.24 ± 0.03 s^{-1} mM^{-1}, respectively at pH 9.0. Analysis of the reaction products via GC–MS confirmed 2-chloromaleylacetate as the ring-cleavage product. CpsA 3D structure revealed a conserved 2-His-1-carboxylate facial triad motif (His 9, His 244 and Thr 11), with Fe^{3+} at the centre. Findings from this study provide new insights into the involvement of this enzyme in PCP degradation and suggests alternate possible mechanism of ring-cleavage by dioxygenases.

1. Introduction

Pentachlorophenol (PCP) is a synthetic and persistent organochlorine pesticide [1] and has been used extensively as fungicides, pesticides, defoliant and wood preservative [2]. Although it's use was restricted to certified applicators due to its toxicological potentials [3], it is still widely used in industry as a wood preservative for railroad ties, wharf pilings and utility [1]. Apart from being listed as a persistent organic pollutant [4,5], PCP is also classified as Group 1 human carcinogen [6–8] based on sufficient epidemiological evidence [9]. In African and South African context, the presence of organochlorine pesticides and chlorophenol in waters, soil and bio life have been reported [10–17].

PCP is highly resistant to microbial degradation due to its recalcitrant nature, however, microorganisms such as *Sphingobium chlorophenolicum* L-1 [18], *Sphingomonas* sp. UG30 [19] and the *Bacillus cereus* strain AOA-CPS_1 recently isolated in our laboratory (Manuscript under review) have evolved pathways for its complete degradation. PCP degradation pathway and the catabolic enzymes involved have been studied extensively in *S. chlorophenolicum* L-1 [20], while PCP degradation has also been reported in other strains [21–26].

In most of the microorganisms, PCP degradation is initiated by hydroxylation of PCP to tetrachloro-1,4-benzoquinone (TCBQ) by pentachlorophenol-4-monooxygenase (PcpB) [27], which is reduced to tetrachloro-1,4-hydroquinone (TCHQ) by tetrachloro-1,4-benzoquinone reductase (PcpD) [28]. TCHQ reductase (PcpC) sequentially removes two atoms of chlorine from TCHQ to form 2,6-dichloro-*p*-hydroquinone (DiCHQ) [29]. A ring-cleaving dioxygenase: DiCHQ 1,2-dioxygenase (PcpA) cleaves DiCHQ aromatic ring to form 2-chloromaleylacetate (2-ClMA) [27], which is further dechlorinated to maleylacetate (MA) by 2-chloromaleylacetate reductase (PcpE) [30,31] which finally funnelled to the tricarboxylic acid (TCA) cycle. DiCHQ is a common metabolic intermediate of several chloroaromatic compounds, such as PCP, tetrachlorophenol, trichlorophenol (TCP) and other polychlorophenolic compounds [27]. The ring cleavage and dehalogenation steps catalysed by PcpA are considered very crucial in polychlorophenols degrading microorganisms [27,32,33]. Two pathways are known for DiCHQ transformation to 2-ClMA and MA. One is direct cleavage to 2-ClMA and second is hydroxyquinol (HQ) pathway in

* Corresponding author.
E-mail address: olaniranai@ukzn.ac.za (A.O. Olaniran).

https://doi.org/10.1016/j.ijbiomac.2020.05.268
0141-8130/© 2020 Elsevier B.V. All rights reserved.

which DiCHQ is converted to chloro-hydroxyquinol (CHQ) [34]. PcpA is an inducible enzyme and its mRNA expression level increases in the presence of PCP [27]. Moreover, PcpA is highly specific for DiCHQ as preferred substrate [33]. PcpA from *S. chlorophenolicum* strain ATCC 39723 is a member of the Fe^{2+}-dependent ring-cleaving hydroquinone dioxygenases (HQDOs) [27,32] of the vicinal oxygen chelate (VOC) superfamily proteins, alongside 2-chlorohydroquinone 1,2-dioxygenase (LinE) from *Sphingomonas paucimobilis* (UT26) [35] and hydroquinone 1,2-dioxygenase (MnpC) from *Cupriavidus necator* JMP 134 [36]. Moreover, unlike extradiol, intradiol, and gentisate dioxygenases that are inactivated by chlorinated substrates [37], PcpA catalyses the cleavage of aromatic rings between a hydroxyl and chlorine group [35].

To date, only PcpA from *S. chlorophenolicum* was reported to open the aromatic ring of DiCHQ, however the catalytic mechanism is not clear [32]. Furthermore, thorough studies are needed on enzymes with the catalytic properties like PcpA to better understand the mechanism of catalysis. Hence, this study reports on the detection, cloning, overexpression and characterization of a PcpA analogue enzyme: 2,6-Dichloro-*p*-hydroquinone 1,2-dioxygenase (CpsA) from *Bacillus cereus* strain AOA-CPS_1 indigenous to contaminated wastewater in South Africa. To the best of our knowledge, this is the first report on the biological function of a multispecies ring-cleaving dioxygenase (a multispecies of *Bacillus* family protein whose function remain putative to date), and its involvement in 2,6-dichloro-*p*-hydroquinone aromatic ring-cleavage. Additionally, structure analysis of the CpsA was elucidated using molecular modelling and Bioinformatics tools.

2. Materials and methods

2.1. Materials

2,6-Dichloro-1,4-benzoquinone (DiCBQ, 98%), isopropyl β-D-1-thiogalactopyranoside (IPTG) ≥ 99%, Luria Bertani agar and Broth (LB) (Vegitone), imidazole, *N*-methyl-*N*-trimethylsilyl-trifluoroacetamide (TMS), phenylmethylsulfonyl fluoride (PMSF), ampicillin sodium salt and 4-Morpholineethanesulfonic acid hemisodium salt (MES) were purchased from Merck (Merck & Company, Inc., NJ, USA). Dithiothreitol (DTT), ethylenediaminetetraacetic acid (EDTA), restriction endonucleases (*Nde*I and *Bam*HI), T4 DNA ligase, DNA ladders (100 bp plus, 1 kb and 1 kb plus), page ruler plus prestained protein ladder, *Taq* polymerase and PCR reaction mix were obtained from ThermoFisher Scientific (Waltham, MA, USA). *Escherichia coli* strains DH5α and BL21 (DE3) (Invitrogen, ThermoFisher Scientific, Waltham, USA) were used as cloning and protein overexpression hosts, respectively. Plasmid vector: pET15b DNA (Novagen, Merck, Germany), was used as a cloning and protein expression vector. Chemically competent cells of *E. coli* strains DH5α and BL21(DE3) were prepared as previously described [38]. All chemicals and reagents used in this study were of analytical grade standards.

2.2. Isolation and identification of Bacillus cereus strain AOA-CPS_1 (BcAOA)

*Bc*AOA was isolated from the effluent of a wastewater treatment plants in Durban, South Africa, enriched using a minimal salt medium (MSM) supplemented with 50 mg l^{-1} of PCP and 5 ml of sludge. After three successive transfers, the isolate was purified via successive subculturing on sterile nutrient agar (NA) plates until distinct colonies were obtained. Glycerol stocks of the pure isolates were prepared and stored in a bio-freezer at −80 °C. The 16S rDNA was amplified from genomic DNA of the isolate using 63F and 1387R universal primers pair [39] to identify the strain. The amplified 16S rDNA gene partial fragment was sequenced (Inqaba Biotechnical Company, Pretoria, South Africa) and submitted at NCBI BLASTn server [40].

2.3. Detection, amplification and cloning of cpsA

To detect and amplify *cps*A from *Bc*AOA, the PCR was performed using genomic DNA (purified using Quick-DNA Fungal/Bacteria Kit, Zymo Research Corporation, USA) as template and degenerate primer pair 1 (forward: 5'-TCTGTYMCRAWTTCAAAWA-3' and reverse: 5'-ATRADRVAGGWARNSCHGGWA-3'). To design the primer pair 1, the *Bacillus* spp. strains entries (from the NCBI BLASTn for 16S rDNA) with whole genome sequencing projects were selected and genome sequences searched for the presence of ring cleaving 1,2-dioxygenases. Subsequently, gene sequences of three strains; *B. thuringiensis* strain ATCC 10792, *B. thuringiensis* strain L-7601, and *Bacillus* sp. FDAARGOS_235 showing the presence of putative ring cleaving 1,2-dioxygenases were aligned using DNAMAN v9.0 software (Lynnon, Biosoft, USA) and the primer pair 1 were designed manually. The PCR resulted in the amplification of ≈687 bp (Fig. S1) fragment and sequenced as described above. The DNA sequence was again submitted at NCBI BLASTn search resulted in 100% homology with a 978 bp nucleotide sequence from many *Bacillus* spp. Primer pair 2, forward primer: 5'-GGGCATATGATGAACCAATTAAAAGGA-3' and reverse primer: 5'-ACGGGATCCTTACTCTTTAATAAATTCCTT-3'), inserting *Nde*I and *Bam*HI restriction sites (underlined) targeting expression vector pET15b for cloning purpose, was designed. The 10 μl PCR reaction mix for *cps*A amplification contained: 1 μl of 10× buffer, 1.5 mM of $MgCl_2$, 20 μM of dNTPs, 1.0 μM of each primer (primer pair 1), 1 μl of genomic DNA, 1.25 U of High-fidelity DNA polymerase and 5.15 μl of ddH_2O. The reaction mix was amplified in a thermal cycler (T100™ Thermal Cycler, Bio-Rad, USA), with the following amplification conditions: initial denaturation at 95 °C for 5 min, followed by denaturation at 95 °C for 1 min, annealing at 58 °C for 1 min, 25 cycles; extension at 72 °C for 1 min, and a final extension at 72 °C for 10 min. The amplicon was extracted from agarose gel using a gene gel purification kit. The purified DNA was sequenced and confirmed to be the expected product. The pET15b plasmid DNA and the gel-purified *cps*A gene fragment were double digested using *Nde*I and *Bam*HI FastDigest restriction enzymes. The digested pET15b plasmid DNA and *cps*A were gel purified and ligated using T4 DNA Ligase. The ligated product pET15b-*cps*A was transformed into chemical competent *E. coli* DH5α by heat shock technique [38] and selected on LB ampicillin (50 μg ml^{-1}) agar plates followed by colony PCR to select for positive transformants. The recombinant pET15b-*cps*A plasmid was purified using a plasmid miniprep kit and again transformed into *E. coli* BL21 (DE3) expression host followed by colony PCR to confirm successful transformation. The presence of *cps*A in pET15b-*cps*A plasmid was confirmed by restriction digestion using the corresponding restriction enzymes. The in-frame cloning of *cps*A and correct sequence was confirmed by sequencing using universal T7 promotor and T7 terminator primers.

2.4. Overexpression and purification of CpsA

The transformed *E. coli* BL21(DE3) containing pET15b-*cps*A was pre-grown overnight at 37 °C in 100 ml of LB broth containing ampicillin (50 μg ml^{-1}). One hundred-milliliter inoculum was used to inoculate 2 L of LB broth containing ampicillin (50 μg ml^{-1}) and incubated at 37 °C to an optical density (OD) of 0.6 at 600 nm. The culture broth was induced with 1 mM IPTG and incubated at 20 °C for 24 h, with shaking at 200 rpm. The cell pellets were harvested by centrifugation at 8000 rpm for 10 min and the supernatant was decanted. The pellets were re-suspended in 50 mM sodium phosphate ($NaPO_4$) buffer (pH 7.0) containing 1 mM each of DTT and PMSF, washed twice and re-suspended in 100 ml of the same buffer. The re-suspended cells were sonicated at Psi 40 and 40 V for 5 min with 10 s pulse (Omni Sonic Ruptor 400, Omni International, Kennesaw, Georgia). The cell lysate was centrifuged for 30 min at 15000 rpm and 4 °C and the supernatant containing CpsA was purified. The total protein concentration in the supernatant was determined by Bradford method [41].

To purify the recombinant 6xHis-tagged CpsA, 10 ml (10 mg total protein) of supernatant was loaded in 5 ml Pierce HisPur Cobalt Chromatography Cartridge (#90094, Thermo Scientific, IL, USA) connected to ÄKTA purifier system (GE Healthcare Life Sciences, IL, USA) and equilibrated with buffer 'A' (50 mM $NaPO_4$, 300 mM NaCl, 5 mM imidazole, pH 7.4) followed by elution with buffer 'B' (50 mM $NaPO_4$, 300 mM NaCl, 150 mM imidazole, pH 7.4). The elution fractions showing a single protein peak of interest at 280 nm were pulled together and concentrated using 10 kDa cut-off Amicon Ultra-15 centrifugal filter unit (Merck, NJ, USA). Expression of CpsA, its level in cell lysate, homogeneity of the purified protein and its molecular weight were determined by loading the samples in 12% SDS-PAGE [42]. The enzymatic activity of CpsA at each purification steps was determined as described below.

2.5. Enzyme activity assay

Enzyme activity for CpsA was determined in a 1 ml reaction mixture by continuous spectrophotometric method [43] with modifications [27,44]. Briefly, 12 μl of sodium borohydride (100 mM stock, freshly prepared in ddH_2O) was used to reduce 120 μl of DiCBQ (2 mM stock, dissolved in 5% ethanol) to DiCHQ. Ten microliters of 1 mM $FeSO_4.7H_2O$ and 10 μl of 750 mM of imidazole were mixed aerobically with 10 μl (100 μg) of purified CpsA and allowed to interact for about 10 min. Finally, DiCHQ (132 μl) and 30 μl of the CpsA, $FeSO_4.7H_2O$ and imidazole mixture (prepared as above) were mixed with 838 μl of 50 mM $NaPO_4$ buffer (pH 7.0). The consumption of DiCHQ and formation of 2-ClMA were monitored spectrophotometrically at 305 nm and 253 nm, respectively. The absorbance at 253 nm representing the formation of 2-ClMA was converted to the nanomole (nm) of the product by using extinction coefficient of 2-ClMA i.e. 9600 $\mu M^{-1}\ cm^{-1}$ [43]. One unit of CpsA activity was defined as the enzyme producing 1 nM of product per min.

2.6. Determination of optimum pH and pH stability

To determine the optimum pH for the CpsA activity, 250 μl (200 μg) of the purified enzyme was assayed as described above except that the reactions were performed using different pH buffers i.e. acetate (pH 4.0–5.0), $NaPO_4$ (pH 6.0–8.0), glycine-NaOH (pH 9.0–10.0), Na_2HPO_4–NaOH (pH 11–12) and KCl–NaOH (pH 13.0). To determine the pH stability of purified CpsA, the enzyme (200 μg) was pre-incubated in buffers ranging from pH 7 to 13 and the residual activity was determined as described above. The activity obtained at optimum pH 9.0 was considered as 100% residual activity.

Table 1
Purification of 2,6-dichloro-p-hydroquinone-1,2-dioxygenase from *Bacillus cereus* AOA-CPS_1.

Purification step	Vol (ml)	Total protein (mg ml^{-1})	Total activity (U)	Specific activity (U mg^{-1})	Purification (fold)	Yield (%)
Crude	400.00	2965.471	1980	0.667	1.0	100
HisTrap	122.00	2.364	241.42	102.124	153	12
Ultrafiltration	22.70	6.818	1341.23	196.719	285	67.74

2.7. Determination of optimal temperature and temperature stability

To determine the optimum temperature for the CpsA activity, 250 μl (200 μg) of the purified enzyme was assayed as described above except that the reactions were performed at different temperatures ranges from 20 °C to 90 °C at the optimal pH 9.0. The enzyme thermostability was investigated by incubating the enzyme (200 μg) at different temperatures (40 °C–90 °C) for 2 h (aliquots taken out after every 10 min) and assayed at optimum conditions as described above. The residual activity obtained at optimum conditions was considered as 100% residual activity.

2.8. Determination of kinetic parameters

The purified CpsA (200 μg) was incubated with varying substrate concentrations (50 to 500 μM of DiCHQ) and the activity was determined at optimum conditions as described above. K_m and v_{max} values

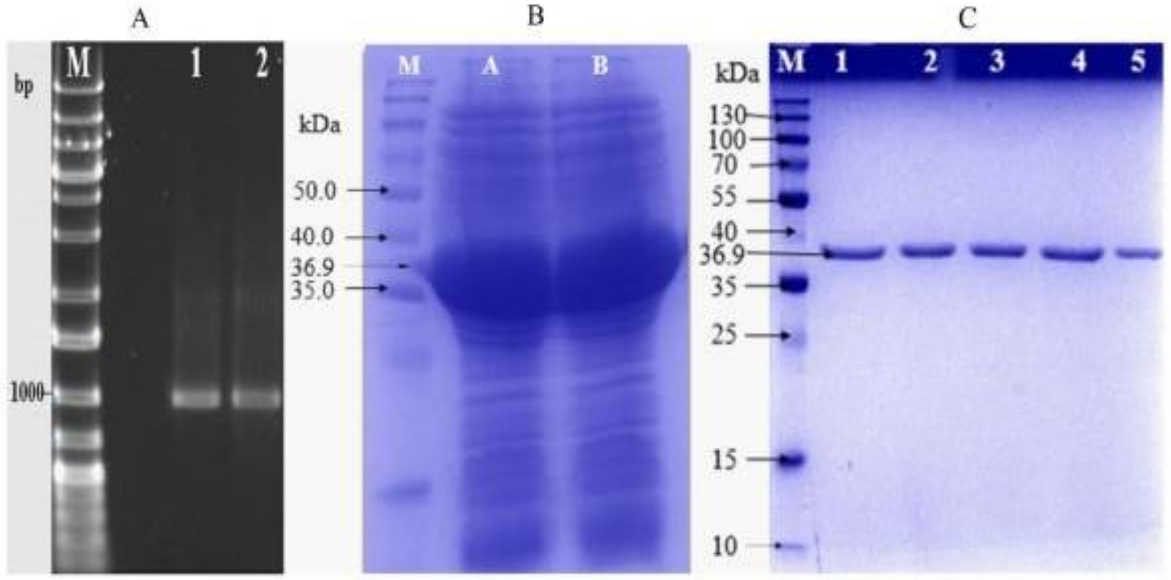


Fig. 1. PCR amplification, overexpression and purification of CpsA. (A) Amplification of *cpsA* gene from *BcAOA*; DNA marker (Lane M), amplified *pcpA* gene (Lane 1 and 2). (B) 12% SDS-PAGE showing overexpression of CpsA; protein marker (Lane M), uninduced (Lane A) and induced (Lane B) CpsA; (c) AKTA fractions of purified recombinant 6xHis-tagged CpsA (Lanes 1–5), protein marker (Lane M). *36.9 kDa molecular weight indicated is based on the biophysical properties calculated at ProtParam tool at ExPASy for CpsA amino acid sequence. Uninduced BL21(DE3) cells also showed the expression of CpsA as it is well established that T7 RNA polymerase is expressed at basal level in BL21(DE3) and may express the recombinant proteins in some cases [63,64].

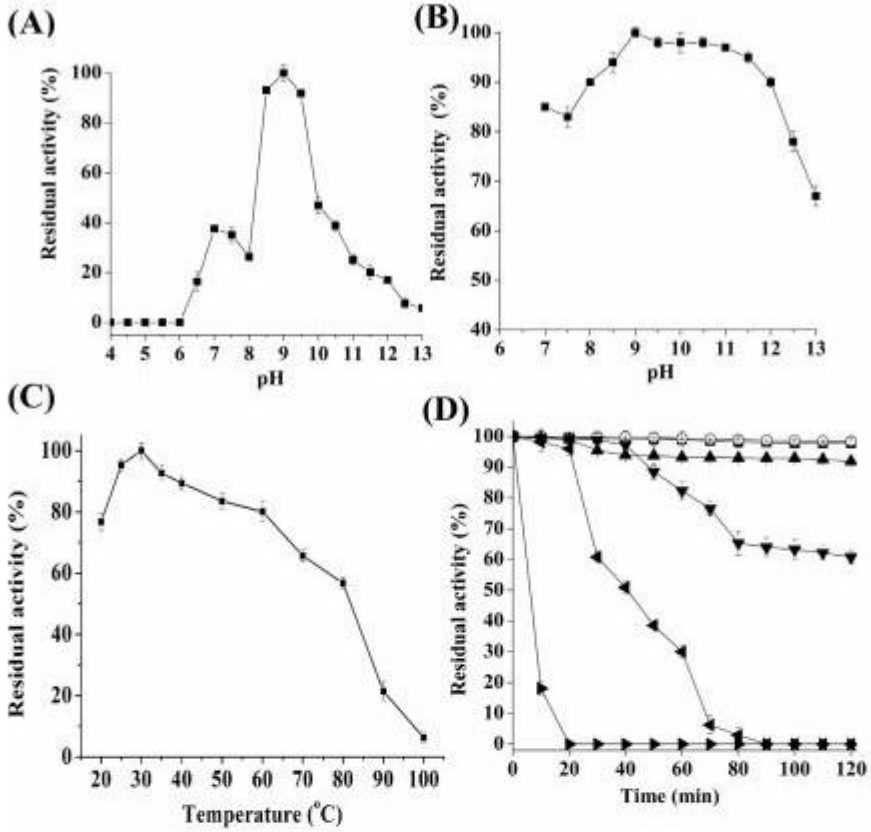


Fig. 2. Characterization of CpsA. (A) Optimum pH, (B) pH stability at different pH range for 120 min, (C) optimum temperature, (D) temperature stability at 40 °C (■), 50 °C (○); 60 °C (▲); 70 °C (▼); 80 °C (◀), 90 °C (▶) over time.

were obtained by applying Michaelis-Menten equation in Line-weaver Burk plots. The catalytic constant of the enzyme substrate reaction (k_{cat}), also referred to as the turnover number, represents the number of reactions catalysed per unit time by each active site; was determined using the equation, $k_{cat} = v_{max}/[E]t$, where v_{max} is the maximum velocity and $[E]t$, is the total enzyme concentration. Catalytic efficiency was calculated by the equation, k_{cat}/K_m.

2.9. Effects of metal ions on CpsA activity

To determine the effect of different concentrations of metal ions, 100 μg of purified CpsA was pre-incubated with the metal ions for 10 min and assays were performed as described above. The enzyme activity determined without addition of any metal ion was considered as 100% residual activity [45].

Table 2
Effect of metal ions on 2,6-dichloro-p-hydroquinone-1,2-dioxygenase from *Bacillus cereus* AOA-CPS_1.

Metal ions	Residual activity (%)
Zn^{2+}	6.07 ± 0.16
Mn^{2+}	6.18 ± 0.11
Fe^{3+}	13.02 ± 1.12
Fe^{2+}	100 ± 2.08
Ca^{2+}	9.07 ± 2.09
Cu^{2+}	11.13 ± 2.13
Mg^{2+}	14.87 ± 2.11
Pb^{2+}	4.08 ± 0.95
Ni^{2+}	5.92 ± 1.11
CpsA	18.34 ± 2.13[a]

[a] Assay with buffer only.

2.10. Detection of metabolites in the reaction mixture

The products or metabolites of reaction between DiCHQ and CpsA were detected by GC–MS analysis [35]. The metabolites were extracted with ethyl acetate at a ratio 1:1 after acidification by HCl from the reaction mixture obtained as described above. The extracted metabolites were derivatized using TMS at 65 °C for 15 min. The derivatized TMS samples were analyzed by loading 1 μl in the GC–MS column (HP-5MS, 5% Phenyl Methyl Silox 325 °C: 30 m × 250 μm × 0.25 μm) connected to 7890A GC system equipped with a 5975C MS detector (Agilent Technologies, CA, USA). The system was run at 80 °C raised at 5 °C min^{-1} to 160 °C for 3 min, then raised at 10 °C min^{-1} to 260 °C and held at 260 °C for another 3 min. The other parameters were set as; Mass ranges (m/z), 50–700, ionization energy, 70 eV. The identity of the compounds was determined from National Institute of Standard and Technology (NIST) library database. Pure 2,4-dichlorophenol was used as an internal standard.

2.11. In-gel trypsin digestion and identification of the purified CpsA in ES-MS

The pure protein (50 μg) was loaded onto 12% SDS-PAGE and stained with Coomassie blue R250. The protein band was excised carefully and

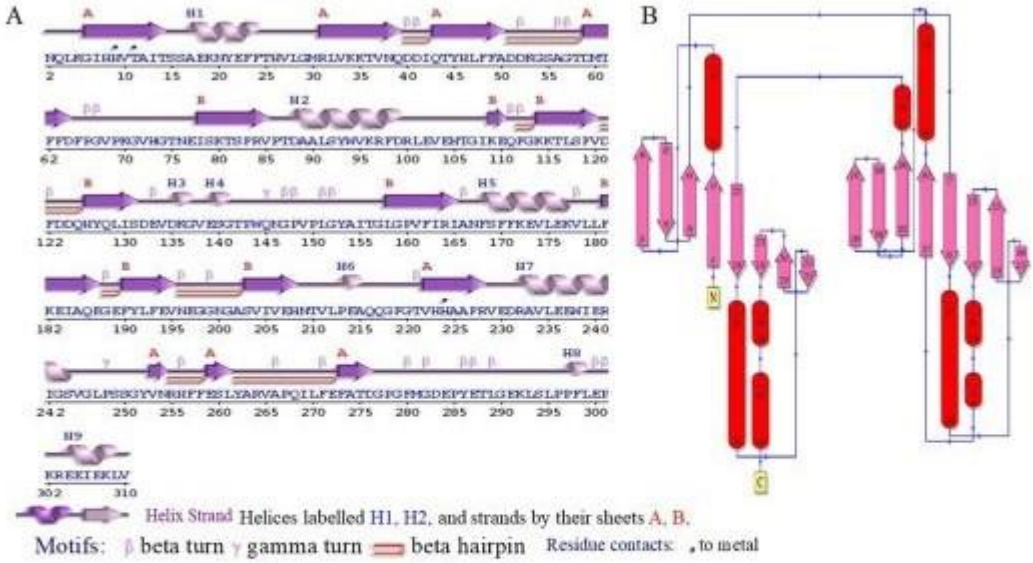


Fig. 3. The predicted 3D and secondary structural framework of CpsA. (A) Secondary structural elements of CpsA obtained using PDBsum indicating βαβββ motifs, 2 β-sheets, 8 β-hairpins, 1 ψ-loop, 10 β-bulges, 16 strands, 9 α-helixes, 28 β-turns and 3 γ-turns. (B) Topology of CpsA obtained using PDBsum showing the βαβββ motifs repeats.

digested with trypsin and fragments analyzed by electrospray mass spectrometry (ES-MS) (Central Anlalytical Facility, Stellenbosch University, Stellenbosch, South Africa). The data analysis and the protein identification were done as described previously [46].

2.12. Template-based structure prediction and homology modelling for CpsA

Three-dimensional structure and homology modelling of the protein were predicted by submitting the amino acid sequence at SWISS-MODEL tool at ExPASy bioinformatics resource portal workspace (https://swissmodel.expasy.org/interactive). The default parameters used for performing the automated SWISS-MODEL were as explained previously [30] and elaborated at (https://swissmodel.expasy.org/docs/help) webpage. The modelled PDB files were submitted to online tool PDBsum for determining structural summary [47].

2.13. Evolutionary relationships of CpsA with other dioxygenases

To study the evolutionary relatedness of CpsA with dioxygenases from vicinal oxygen chelate (VOC) family, intradiol, extradiol and other meta-ring cleavage dioxygenases, the phylogenetic based analysis was performed using Neighbor-Joining method [48]. The evolutionary history, bootstrap consensus tree [49], percentage of replicate trees [49] and the evolutionary distance tree [50] were performed using MEGA7 software [51].

3. Results and discussion

3.1. Cloning, overexpression and purification of 6xHis-tagged CpsA

Analysis of the 16S rDNA (submitted to NCBI GenBank under accession number MH504118.1.) sequence of the indigenous PCP degrading isolate used in this study identified it as *Bacillus cereus* strain AOA-CPS_1 (*Bc*AOA). Further, a 978 bp ring cleaving dioxygenase gene *cpsA* was amplified from the genome of *Bc*AOA using PCR (Fig. 1A) and cloned to overexpress and purify it to homogeneity. The *cpsA* gene sequence shared above 99% homology with previously reported multi species ring cleaving dioxygenase genes in many *Bacillus* spp. (unpublished data). The *cpsA* gene fragment was successfully cloned into an expression vector pET15b, and the recombinant C-terminal 6xHis-tagged CpsA was overexpressed in *E. coli* BL21(DE3) (Fig. 1B). The recombinant CpsA protein was purified to homogeneity and showed Mr. of ≈36.9 kDa (Fig. 1C). The molecular weight of CpsA from *Bc*AOA is similar to that of other ring cleaving dioxygenases reported previously [36,52,53]. A single step purification strategy purified the enzyme 285-fold with a total 67.74% yield (Table 1). The total enzyme activity and specific activity of CpsA at each purification step are also shown in Table 1.

3.2. Optimum pH and pH stability of CpsA

CpsA showed an optimum activity at pH 9.0. The activity at pH 8.5 and 9.5 was slightly lower than that at pH 9.0 but drastically decreases at pH 8.0 and 10.0 (Fig. 2A). The enzyme profile shows a peak at pH 7.0 as well, in consonance with PcpA enzyme from *S. chlorophenolicum* L-1 which exhibited two peaks at neutral and alkaline pH [27]. However, the activity at pH 9.0 was much higher than that at pH 7.0. CspA was stable at a wide range of alkaline pH. The enzyme retained >90% of its activity between pH 9.0 and pH 11.5 and demonstrated over 60% residual activity at pH 13.0 relative to the activity at pH 9.0 (Fig. 2B). Similar results were reported for PcpA from *S. chlorophenolicum* L-1 [27]. The alkaline stability of these enzyme may be attributed to the fact that >80% of PCP was extensively used as wood preservative in combination with sodium hydroxide which may initially turn most PCP polluted sites into alkaline environments [54]. Moreover, it is now established that some microbes have evolved efficient pathways to degrade PCP under alkaline conditions, which buttress the fact that most bacterial PCP transformation occurred more rapidly under alkaline than acidic conditions.

3.3. Optimum temperature and temperature stability of CpsA

CpsA showed an optimum temperature at 30 °C and then decreased to about 60% at 80 °C and only retained <6% of its activity at 100 °C

(Fig. 2C). Moreover, CpsA showed good stability at temperature between 20 °C and 40 °C. CpsA retained about 90% of its activity at 60 °C for 50 min, however, decreased sharply to about 60% and 20% at 70 °C and 80 °C within 30 min and 10 min, respectively, with no activity at 90 °C (Fig. 2D). PcpA from *S. chlorophenolicum* L-1 also exhibited similar results for its temperature stability [27].

3.4. Effect of metal ions on CpsA activity

CpsA exhibited maximum activity in the presence of Fe^{2+} ion as compared to activity in buffer only. The other metal ions like Zn^{2+}, Mn^{2+}, Fe^{3+}, Ca^{2+}, Cu^{2+}, Mg^{2+}, Pb^{2+} and Ni^{2+} could not enhance CpsA activity (Table 2). The observed 5.5-fold increase in CpsA activity in this study in the presence of Fe^{2+} is slightly higher than the 4.2-fold increase in PcpA activity reported previously [53]. Fe^{3+} ions could not enhance the enzymatic activity of CpsA (this study) nor in PcpA, a very closely related enzyme to CpsA from *S. chlorophenolicum* L-1 [27,53].

3.5. Steady-state kinetic parameters of CpsA

CpsA exhibited activity peaks at two different pH (Fig. 2A), therefore pH 7.0 and pH 9.0 were used to determine the kinetic parameters of CpsA. The apparent v_{max}, K_m, k_{cat} and k_{cat}/K_m for CpsA at pH 7.0 were found to be 8.55 ± 0.96 μMs^{-1}; 0.656 ± 0.04 mM; 1.26 ± 0.06 s^{-1} and 1.92 ± 0.01 s^{-1} mM^{-1}, respectively, while 27.77 ± 0.9 μMs^{-1}; 0.990 ± 0.03 mM; 4.20 ± 0.04 s^{-1} and 4.24 ± 0.03 s^{-1} mM^{-1}, respectively, were obtained at pH 9.0. The k_{cat} for CpsA at pH 9.0 is comparable to that of PcpA from *S. chlorophenolicum* L-1 [33], and in agreement with the earlier findings of Sun et al. [27]. The k_{cat} for CpsA calculated at pH 7.0 is also similar to previous reports [32]. The K_m of CpsA is similar to reported for a ring cleaving dioxygenase enzyme MnpC from *Cupriavidus nacator* JMP134 [36] and PcpA from *S. chlorophenolicum* L-1 [27].

3.6. Ring-cleavage product of CpsA and DiCHQ reaction

The GC–MS analysis of the derivatized TMS catalytic products of DiCHQ by CpsA yielded a keto form of 2-CIMA (dimethyl methyl maleate, Tris(trimethylsilyl) malic acid and 2-Keto tri-TMS) (Fig. S2).

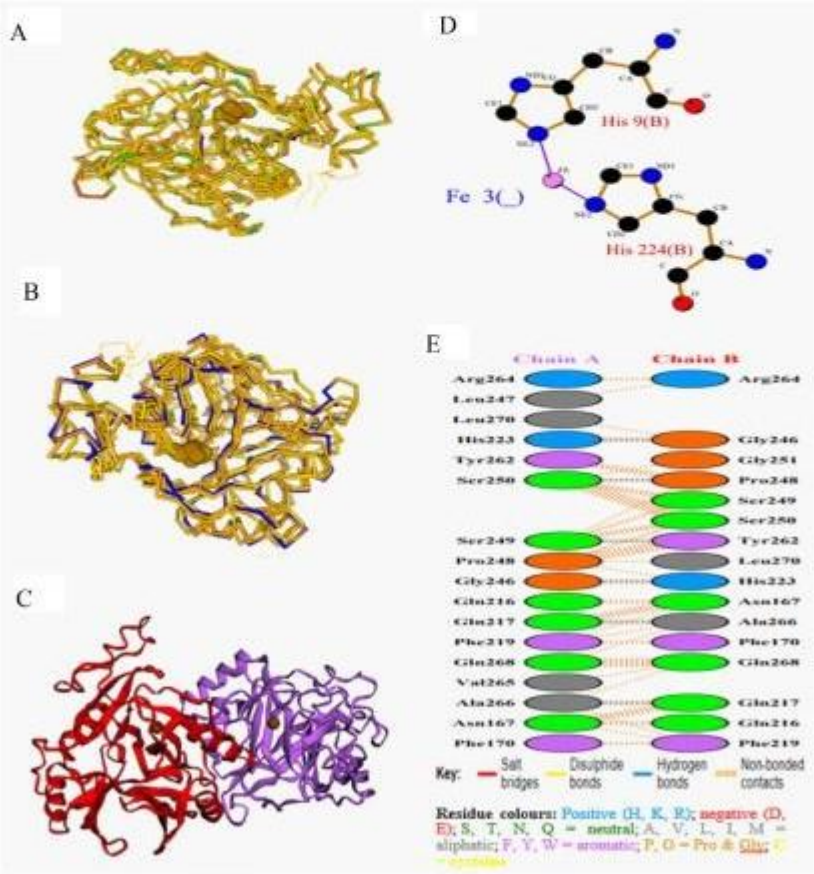


Fig. 4. Graphics of the 3D superimposition of 4HUZ, MHQO, 30AJ and CpsA showing: (A) the conserved and (B) conserved structures; (C) 3D structure of CpsA; (D) ligand-metal interactions (LIGPLOT v.4.5.3); (E) schematic diagram of interactions between protein chains. Interacting chains are joined by coloured lines, each representing a different type of interaction, as per the key above. The area of each circle is proportional to the surface area of the corresponding protein chain. The extent of the interface region on each chain is represented by the black wedge whose size signifies the interface surface area. The number of H-bond lines between any two residues indicates the number of potential hydrogen bonds between them. For nonbonded contacts, which can be plentiful, the width of the striped line is proportional to the number of atomic contacts. The model built with 1ZSW as a template (model: CpsA.1) has a sequence identity of 37.82% while those built with 30AJ (model: CpsA.2) and 4HUZ (model: CpsA) have 38.87% and 30.0% sequence identities respectively. CpsA.1 has no ligands while CpsA.2 and CpsA have zinc (Zn) and iron (FE) metal ions respectively. However, sulphate (SO_4^{2-}) ion was present in the templates but was not automatically built into the model been a non-biologically functional ligand.

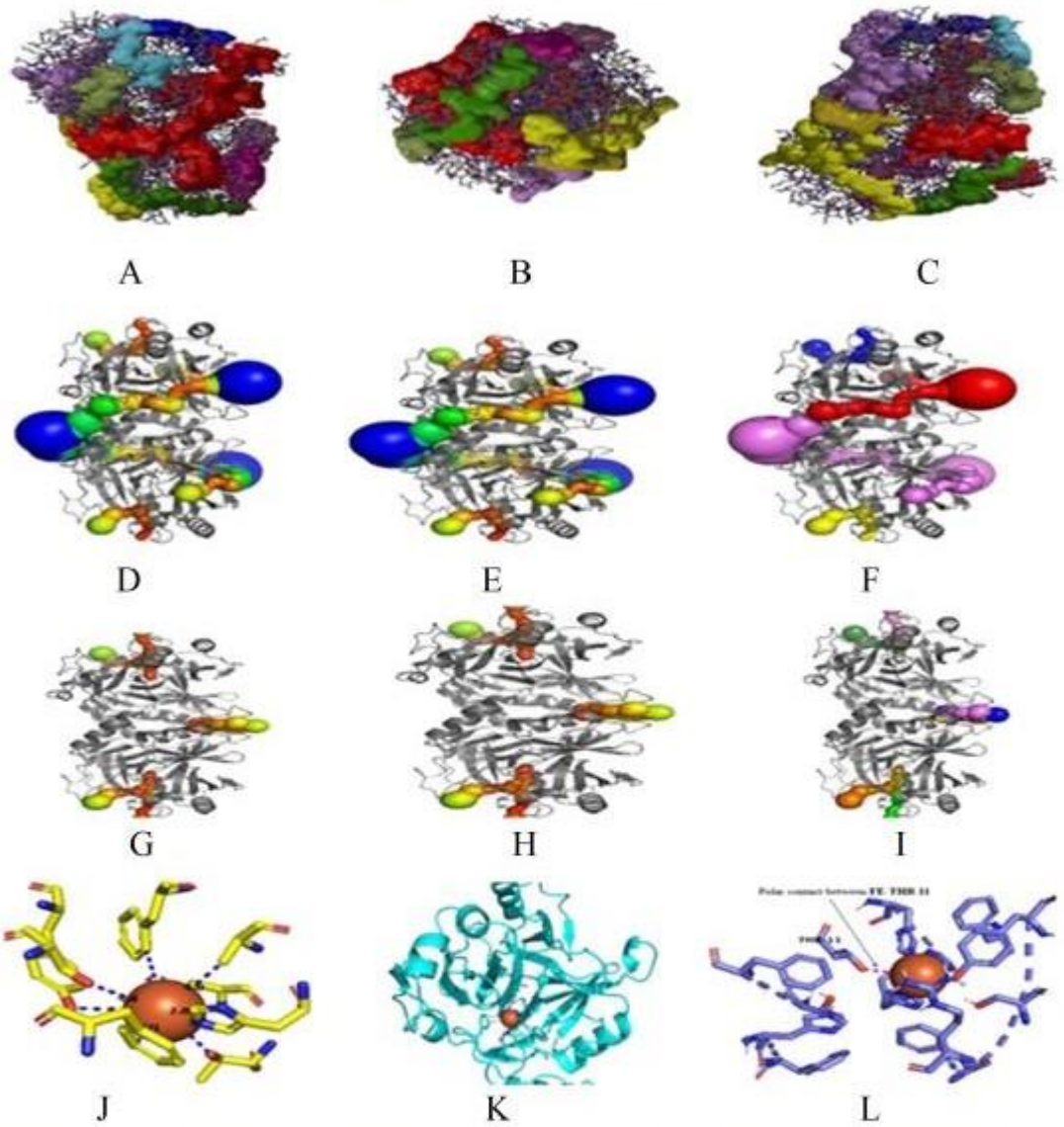


Fig. 5. Clefts (PDBSum), pores and tunnels (MOLE 2.0 program v2.5.13.11.08 and visualized on pymol 0.97rc) found in CpsA structure: The clefts (A, B and C); pore calculated on whole structure (D); pores calculated excluding ligands (E and F); tunnel calculated on whole CpsA structure (G); tunnels calculated excluding ligands (H and I) and the active site of CpsA (J, K and L). The pores are connected internal spaces within through the structure, there were 4 pores found in the protein structure, however, only pores longer than 25 Å were presented. The radius of the pores bottleneck is given in Å. The free radius of the pore bottleneck was calculated without sidechain atoms in order to consider also the side-chain flexibility.

Previous studies have shown that the ring-cleavage product of DiCHQ by PcpA was 2-CIMA [35,43,55]. The regio-specificity of ring cleavage and the isomeric forms of the cleavage products of DiCHQ had been studied using a proton nuclear magnetic resonance (^{1}H NMR). The ^{1}H NMR spectra of the ring-cleavage product of DiCHQ are 2-CIMA but the metabolite exists also in a keto form [33], suggesting that the ring cleavage products can exist in different isomeric forms, under different conditions [33,56,57]. For instance, at neutral pH in water, unsubstituted maleylacetate exists entirely in the keto form [56] while unsubstituted maleylacetone exists as a mixture of the keto and enol forms [58]. Under acidic conditions and subsequent extraction into organic solvents, these compounds condense or cyclize to form cyclic beta-keto ester or cyclic carboxylic esters [33].

3.7. Confirmation of the identity of purified protein as CpsA

The pure SDS-PAGE band excision followed by the tryptic digestion resulted in the generation of peptides of CpsA. The LC-MS analysis of the peptides and resolution with a multiple search engine (SearchGui v3.3.15) followed by virtualisation on PeptideShaker (version 1.16.45), revealed that the peptide structure matched the multi species glyoxalase family protein (UniProt accession number Q81H02) reported from *B. cereus* strains ATCC 14579/DSM 31/JCM 2152/NBRC 15305/NCIMB 9373/NRRL B-3711, with 71.38% coverage, precursor charges of 2 and a 100% peptide matching confidence. The protein Q81H02 was reported only at prediction level and no experimental evidence existed. However, the present study fills the gap by reporting CpsA at

of 325 residues of 2 β-sheets, 8 β-hairpins, 1 ψ-loop, 10 β-bulges, 16 strands, 9 α-helixes, 28 β-turns and 3 γ-turns (Fig. 3A–B). The predicted secondary structure of CpsA contains βαβββ motifs suggesting that the protein is a member of the vicinal oxygen chelate (VOC) protein superfamily. The βαβββ motifs in VOC superfamily provide a metal coordination domain with two or three open or readily accessible coordination sites to promote direct electrophilic participation of the metal ion in catalysis [61,62]. VOC has been found in different structurally related metalloproteins, including the type I extradiol dioxygenases (type 1 EDOs), methylmalonyl-CoA epimerase, glyoxylase I family proteins, and a group of antibiotic resistance proteins such as bleomycin/fosfomycin resistance proteins [61,62]. In addition, a 3D superimposition of 4HUZ, MhqA, 3OAJ and CpsA sequences showed that the sequence and the structure were conserved in the selected proteins (Fig. 4A–B).

The 3D graphical structures and models of CspA generated at PDBsum database by submitting Model.01 PDB file provided a significant insight into its structure (Fig. 4C). The model predicted the protein to be a homodimer (Fig. 4E) but confirmed in vitro to be a monomer based on the 12% SDS-PAGE of the pure protein (Fig. 1C). The clefts, pores, tunnels and the active site structure in the build model are shown in Fig. 5A–I. There were 8 clefts found in the structure, most of these clefts were large enough to enhance the effectiveness of the protein catalytic function (Fig. 5A–C). CpsA was shown to have 8 tunnels in the protein structure (Fig. 5D–I). The hydropathy index per each amino acid in the tunnels indicated that the amino acids in the channels are mostly hydrophilic. The metal ion-binding and residues in contacts with the metal ions along with their bond lengths were docked (Fig. 5J–L). The docking studies show the presence of amino acids residues (His9, Thr11 and His224) in contact with the metal ion in the catalytic site of CpsA.

3.9. The active site of CpsA based on structural and protein ligand modelling

The structure modelling studies suggest that the catalytic active site of CpsA has an Fe^{3+} metal ion-binding centre which is in contact with His 9 and His 224 (Fig. 4D) and in a polar contact with Thr 11 (3.51 Å) (Fig. S6a). Previous reports suggested that PcpA uses Fe^{2+} for catalysis [27,32,43,44,55], experimental evidence suggested that CpsA activity is enhanced only in the presence of Fe^{2+} contrary to the structural modelling studies. It is probable that Fe^{2+} was partially oxidized to Fe^{3+} during purification [27,43].

The 3D structure of CpsA somewhat indicated a possible catalytic mechanism of reaction which is distinctively different from that of PcpA model. The protein-ligands interactions between the residues of CpsA involved ten His with one Thr residues, and the metal ions. This indicate that the facial triad motif of CpsA (His 9, His 224 and Thr 11) is polar, positive and hydrophilic, meaning they can form hydrogen bonds as proton donors or acceptors, unlike that of PcpA which is mainly polar (Fig. S6a–c).

DiCHQ has two hydrogen bond donor and acceptor, each with one covalent bond unit and no formal charge, while Fe^{2+} neither has hydrogen bond donor nor acceptor but has two formal charges and one covalent bond unit. Substrate-ligands interactions in the protein can occur in two plausible ways: the facial triad motif of CpsA and its participating histidine (His 9 and His 224) residues can co-ordinate DiCHQ to the metal ion to form a covalent bond which can be significantly stabilized, by a release of two electrons from any of the participating residues (His 9, His 224 and Thr 11) in contacts with the metal ion (Fe^{2+}) to form a double H-bonds, since DiCHQ has the capacity to accept electrons to form H-bonds. Eight tunnels which are mostly hydrophilic and polar were present in the protein structure (Fig. S7a–c). These tunnels are located near the active site and might be responsible for substrate entry and product exit, been a hydrophilic environment [45].

3.10. The multiple sequences alignments and phylogenetic analysis

The multiple sequences alignments (Fig. 6) showed that the metal ions binding residues of CpsA (H9, T11, and H224) were conserved in other members (PDB: 3OAJ, 4HUZ; LinE, MnpC and 1ZSW) of the VOC superfamily. All the protein sequences aligned showed great similarities in their residues and orientations. Full lengths 3D domains structural alignment of CpsA, 4HUZ and 1ZSW (Fig. S8) revealed seemingly possible evolutionary events that could have accounted for the ring-cleaving ability of 4HUZ and CpsA to degrade hydroquinone. A thorough analysis of the aligned 3D domains structure showed a conserved 3D domains within the sequences. It was observed that a lot of insertions, additions, duplications and swapping of genes/amino acids have: (1) shortened the amino acids residues of both CpsA and 4HUZ compared to those of 1ZSW and 3OAJ; (2) insertion, addition, duplication and/or swapping of genes/amino acids noticed in both sequences occurred at different points and times on the amino acid residues of both CpsA and 4HUZ. These alterations in the amino acid residues might have individually favoured both 4HUZ and CpsA abilities to cleave the aromatic ring of 2,6-Dichloro-p-hydroquinone.

The phylogenetic tree (Fig. 7) showed that 4HUZ, LinE, 3OAJ, PcpA_N_like, VOC and CpsA are in the same monophyletic node and are all like the C-terminal domain of *S. chlorophenolicum* DiCHQ 1,2-dioxygenase (4HUZ). The 100% relatedness of CpsA with 'VOC', a multi-species ring-cleaving dioxygenase from oxygen chelate lineage of *Bacillus*, and similarity in function with PcpA shed more light on the possible

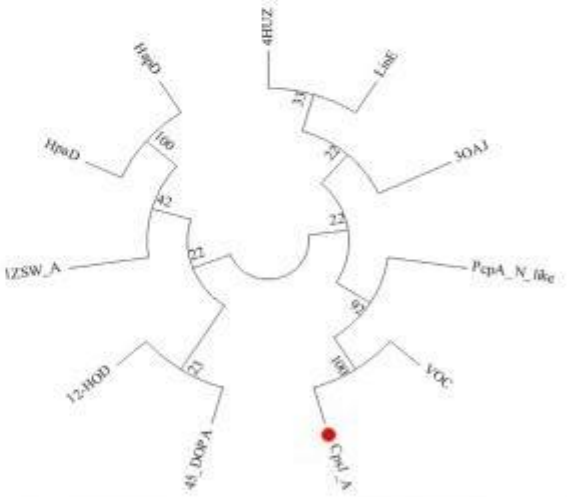


Fig. 7. Phylogenetic relatedness of CpsA with vicinal oxygen chelate (VOC) family, intradiol, extradiol and other meta-ring cleavage dioxygenases. A phylogenetic tree was constructed using: DiCHQ 1,2-dioxygenase (4HUZ) from *S. chlorophenolicum* L-1; 3,4-dihydroxyphenylacetate 2,3-dioxygenase (HapD and HpaD), an aromatic-ring metal-cleaving dioxygenase from *Pseudomonas putida* (BAJ19103.1) and *P. putida* strain TSN1 (BAJ19103.1) respectively; hydroxyquinol 1,2-dioxygenase (1,2-HQD), an intradiol ring-cleavage dioxygenase from *Cupriavidus necator* (WP_042877371.1); 4,5-DOPA dioxygenase (4,5_DOPA), a class III extradiol dioxygenase from *Rhodococcus opacus* (WP_005564099.1); ring-cleaving dioxygenase (PcpA_N_like) from *Cupriavidus pinatubonensis* (WP_011295765.1); hexachlorocyclohexane (HCH) ring cleavage dioxygenase (LinE) from *P. aeruginosa* strain IIFRC-5 (ABP93364.1); multispecies ring-cleaving dioxygenase of Vicinal oxygen chelate (VOC) family from *Bacillus* (WP_001072584.1); glyoxalase family protein (1ZSW) from *B. cereus* ATCC 14579 and a putative ring-cleaving dioxygenase Mhqo (3OAJ) from *B. subtilis* strain 168. The bootstrap consensus tree inferred from 1000 replicates and the branches corresponding to partitions reproduced in <50% bootstrap replicates were collapsed. The analysis involved 11 amino acid sequences and all positions containing gaps and missing data were eliminated. There was total of 261 positions in the final evolutionary dataset.

origin of PcpA (4HUZ), whose hydroquinone ring-cleaving ability was thought to have arisen twice from structurally unrelated proteins [44]. Since the VOC family shared the same phyletic node with 4HUZ, which cleaves aromatic rings with two hydroxyl group at para position, 4HUZ might have acquired ability to cleave aromatic ring from this group, by acquiring specificity for chlorohydroquinone and Fe^{2+} instead of Zn^{2+}. The evolutionary tree indicated that 1ZSW is more related to a meta-cleaving dioxygenase (HapD and HpaD); and 1,2-HOD and 4,5_DOPA_dioxygenase which belongs to the intradiol and class III extradiol dioxygenase respectively. The structurally unrelated protein that pcpA was thought to have evolved from might be as a result of the template (1ZSW) that was used for the structural homology modelling of PcpA. It is quite interesting that both PcpA (4HUZ) from *S. chlorophenolicum* and CpsA from BcAOA, which are bio-functionally the same, are also structurally related. By looking at this group of proteins, most of the organisms that expressed this protein are of *Bacillus* origin. Also, 2,6-dichloro-*p*-hydroquinone 1,2-dioxygenase (4HUZ) was model using 1ZSW as a template which is also a *Bacillus* multispecies family protein. Furthermore, 3OAJ is also a putative multispecies ring-cleaving dioxygenase from *Bacillus subtilis* strain 168. It could therefore be speculated that *S. chlorophenolicum* might have acquired a ring-cleavage ability from *Bacillus* lineage.

4. Conclusion

For years, PcpA, a DiCHQ 1,2-dioxygenase from *S. chlorophenolicum* has been the only enzyme reported to be capable of degrading DiCHQ, with limited understanding of its structure and functions. Cloning, over-expression and characterization of a metabolically active CpsA, a DiCHQ 1,2-dioxygenase from the indigenous PCP-degrading *Bacillus cereus* strain in this study has provided more insights on the properties of this group of enzymes. Structural analysis of CpsA residues showed a catalytic active site that is distinctively different from that of PcpA but has the same function. Analysis of the secondary and the 3D structure of CpsA buttress the Fe^{2+} metal ion binding preference of the enzyme while the presence of a βαβββ motifs repeats confirmed the protein to be a member of the special class of hydroquinone dioxygenase of the VOC protein superfamily. The findings of this study also provided information on the likely biological function of the multispecies ring cleaving dioxygenase, a multispecies family protein found in several species of *Bacillus*, whose function is unknown and remain putative to date. The information provided by this structure will be beneficial in understanding the evolutionary origin of the protein and its role in the bioremediation of persistent poly-aromatic organic compound.

Author contributions

O.A. and A.O. conceived and designed the project; O.A. and A.K. designed the experiments; O.A. performed the experiments; M.P. contributed reagents and materials; O.A., A.K., M.P. and A.O. wrote the manuscript; all the authors have read and approved the manuscript.

Funding

National Research Foundation, South Africa (Grant No: 94036 and 92803).

Ethical statement

This article does not contain any studies with human participants or animals performed by any of the authors.

Declaration of competing interest

All the authors declare no conflict of interest.

Acknowledgments

The authors thank Dr Mare Vlok of the central analytical facility, Stellenbosch, South Africa for the Liquid-Chromatography Mass-Spectrometry proteomics analysis of the protein.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2020.05.268.

References

[1] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. Shoemaker, P. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, E. Bolton, PubChem 2019 update: improved access to chemical data, Nucleic Acids Res. 47 (D1) (2019) D1102–D1109, https://doi.org/10.1093/nar/gky1033.

[2] E. Lopez-Echartea, T. Macek, K. Demnerova, O. Uhlik, Bacterial biotransformation of pentachlorophenol and micropollutants formed during its production process, Int. J. Environ. Res. Public Health 13 (2016) 1–21, https://doi.org/10.3390/ijerph13111146.

[3] E.O. Igbinosa, E.E. Odjadjare, V.N. Chigor, I.H. Igbinosa, A.O. Emoghene, F.O. Ekhaise, N.O. Igiehon, O.G. Idemudia, Toxicological profile of chlorophenols and their derivatives in the environment: the public health perspective, Sci. World J. (2013) 1–11, https://doi.org/10.1155/2013/460215.

[4] Centers for Disease Control, Fourth national report on human exposure to environmental chemicals updated table, http://www.cdc.gov/exposurereport/pdf/FourthReport_UpdatedTables_Sep2013.pdf 2018.

[5] ATSDR, Agency for Toxic Substances and Disease Registry, Substance Priority list (Candidates for Toxicological Profiles), 2017 190www.atsdr.cdc.gov/SPL/resources.%0Aiii.

[6] Stockholm convention, Stockholm convention on persistent organic pollutants., in: Draft Guid. Best Available Tech. Best Environ. Pract. Prod. Use Pentachlorophenol list with Specific exemptions under Stock. Conv. (UNEP/POPS/COP.9/INF/16)., Geneva, 2019y, http://chm.pops.int/TheConvention/ConferenceoftheParties/Meetings/COP9/tabid/7521/Default.aspx.

[7] IARC, Pentachlorophenol and some related compounds. IARC monographs on the evaluation of carcinogenic risks to humans, http://publications.iarc.fr/Book-And-Report-Series/Iarc-Monographs-On-The-Identification-Of-Carcinogenic-Hazards-To-Humans/Pentachlorophenol-And-Some-Related-Compounds-2019 2019.

[8] IARC, International Agency for Research on Cancer monographs on the identification of carcinogenic hazards to humans, Agents Classified by the IARC Monographs 2019, pp. 1–124https://monographs.iarc.fr/agents-classified-by-the-iarc/. (Accessed 28 May 2019).

[9] IARC, International Agency for Research on Cancer (IARC) monographs evaluate pentachlorophenol and some related compounds, Lyon, France, https://www.iarc.fr/wp-content/uploads/2018/07/volume-117_news-item.pdf 2016.

[10] A. Yahaya, O.O. Okoh, F.O. Agunbiade, A.J. Okoh, Occurrence of phenolic derivatives in Buffalo River of Eastern Cape South Africa: exposure risk evaluation, Ecotoxicol. Environ. Saf. 171 (2019) 887–893, https://doi.org/10.1016/j.ecoenv.2019.01.037.

[11] WRC, Water Research Commission, , Update http://www.wrc.org.za/ 2019.

[12] O.S. Olatunji, Evaluation of selected polychlorinated biphenyls (PCBs) congeners and dichlorodiphenyltrichloroethane (DDT) in fresh root and leafy vegetables using GC-MS, Sci. Rep. 9 (2019) 1–10, https://doi.org/10.1038/s41598-018-36996-8.

[13] L.N. Nthunya, N.P. Khumalo, A.R. Verliefde, B.B. Mamba, S.D. Mhlanga, Quantitative analysis of phenols and PAHs in the Nandoni Dam in Limpopo Province, South Africa: a preliminary study for dam water quality management, Phys. Chem. Earth 112 (2019) 228–236, https://doi.org/10.1016/j.pce.2019.02.003.

[14] E. Gakuba, B. Moodley, P. Ndungu, G. Birungi, Partition Distribution of Selected Organochlorine Pesticides in Water, Sediment Pore Water and Surface Sediment From Umngeni River, Kwazulu-Natal, South Africa, Water S.A. 44, 2018 232–249, https://doi.org/10.4314/wsa.v44i2.09.

[15] L.A. Thompson, W.S. Darwish, Y. Ikenaka, M.M.S. Nakayama, H. Mizukawa, M. Ishizuka, Organochlorine pesticide contamination of foods in Africa: incidence and public health significance, J. Vet. Med. Sci. 79 (2017) 751–764, https://doi.org/10.1292/jvms.16-0214.

[16] WRC, Persistent organic pollutants threats: threats posed by the organic pollutants to a tourist destination, Water Res. Comm. South Africahttp://www.wrc.org.za/wp-content/uploads/mdocs/PB_1977_PoPs%20in%20DBN%20estuaries.pdf 2015. (Accessed 19 October 2019).

[17] S. Adu-Kumi, M. Kawano, Y. Shiki, P.O. Yeboah, D. Carboo, J. Pwamang, M. Morita, N. Suzuki, Organochlorine pesticides (OCPs), dioxin-like polychlorinated biphenyls (dl-PCBs), polychlorinated dibenzo-p-dioxins and polychlorinated dibenzo furans (PCDD/Fs) in edible fish from Lake Volta, Lake Bosumtwi and Weija Lake in Ghana, Chemosphere 81 (2010) 675–684, https://doi.org/10.1016/j.chemosphere.2010.08.018.

[18] M.S. Nandish, Microbial degradation of phenol and pentachlorophenol, http://krishikosh.egranthac.in/bitstream/1/80650/1/th8401.pdf 2005. (Accessed 25 May 2018).

[19] M.B. Cassidy, H. Lee, J.T. Trevors, R.B. Zablotowicz, Chlorophenol and nitrophenol metabolism by *Sphingomonas* sp UG30, J. Ind. Microbiol. Biotechnol. 23 (1999) 232–241, https://doi.org/10.1038/sj.jim.2900749.

[20] M. Chanama, S. Chanama, Expression of pentachlorophenol-degradative genes of *Sphingobium chlorophenolica* ATCC 39723 in *Escherichia coli*, Asian J. Public Heal. 2 (2011) 78–83.

[21] R.W. Ammeri, I. Mehri, S. Badi, W. Hassen, A. Hassen, Pentachlorophenol degradation by *Pseudomonas fluorescens*, Water Qual. Res. J. Canada. 52 (2017) 99–108, https://doi.org/10.2166/wqrj.2017.003.

[22] H. Tong, M. Hu, Y. Chen, M. Lv, *Burkholderiales* participating in pentachlorophenol biodegradation in iron-reducing paddy soil as identified by stable isotope probing, Env. Sci Process Impacts. 17 (2015) 1282–1289.

[23] V.V. Joshi, M.L. Prewitt, D.-P. Ma, H. Borazjani, Enhanced remediation of pentachlorophenol (pcp)-contaminated groundwater by bioaugmentation with known PCP-degrading bacteria, Bioremediat. J. 19 (2015) 160–170, https://doi.org/10.1080/10889868.2014.995369.

[24] A. Khessairi, I. Fhoula, A. Jaouani, Y. Turki, A. Cherif, A. Boudabous, A. Hassen, H. Ouzari, Pentachlorophenol degradation by *Janibacter* sp., a new *Actinobacterium* isolated from saline sedimenbt of arid land, Biomed. Res. Int. 2014 (2014) 1–9, https://doi.org/10.1155/2014/296472.

[25] G. S.K. Kam, Pentachlorophenol remediation by *Enterobacter* sp. SG1 isolated from industrial dump site, Pakistan J. Biol. Sci. 17 (2014) 388–394, https://doi.org/10.3923/pjbs.2014.388.394.

[26] A.K. Marihal, K.S. Jagadeesh, S. Sinha, Biodegradation of pentachlorophenol by endophytic bacteria isolated from PCP-tolerant plant species, J. Pure Appl. Microbiol. 8 (2014) 1283–1288.

[27] Sun Wanpeng, Sammynaiken Ramaswami, Chen Lifeng, Maley Jason, Schatte Gabriele, *Sphingobium Chlorophenolicum* dichlorohydroquinone dioxygenase (PcpA) is alkaline resistant and thermally stable, J Env. Qual. 7 (2011) 1171–1179.

[28] L. Chen, J. Yang, Biochemical characterization of the tetrachlorobenzoquinone reductase involved in the biodegradation of pentachlorophenol, Int. J. Mol. Sci. 9 (2008) 198–212, https://doi.org/10.3390/ijms9030198.

[29] S.M. Belchik, L. Xun, S-Glutathionyl-(chloro) hydroquinone reductases: a new class of glutathione transferases functioning as oxidoreductases, Drug Metab. Rev. 43 (2012) 307–316, https://doi.org/10.3109/03602532.2011.552909.

[30] L. Chen, Ed S. Krol, M.K. Sakharkar, H.A. Khan, A.S. Alhomida, J. Yang, Residues His172 and Lys238 are essential for the catalytic activity of the maleylacetate reductase from *Sphingobium chlorophenolicum* strain L-1, Sci. Rep. 7 (2017) 1–11, https://doi.org/10.1038/s41598-017-18475-8.

[31] L. J. Chen, K. Maloney, E. Krol, B. Zhu, Yang, Cloning, overexpression, purification, and characterization of the maleylacetate reductase from *Sphingobium chlorophenolicum* strain ATCC 53874, Curr. Microbiol. 58 (2009) 599–603, https://doi.org/10.1007/s00284-009-9377-z.

[32] R.P. Hayes, A.R. Green, M.S. Nissen, K.M. Lewis, L. Xun, C. Kang, Structural characterization of 2,6-dichloro-*p*-hydroquinone 1,2-dioxygenase (PcpA) from *Sphingobium chlorophenolicum*, a new type of aromatic ring-cleavage enzyme, Mol. Microbiol. 88 (2013) 523–536, https://doi.org/10.1111/mmi.12204.

[33] T.E. Machonkin, A.E. Doerner, Substrate specificity of *Sphingobium chlorophenolicum* 2,6-dichlorohydroquinone 1,2-dioxygenase, Biochemistry 50 (2011) 8899–8913, https://doi.org/10.1021/bi200855m.

[34] F.J. Enguita, A.L. Leitão, Hydroquinone: environmental pollution, toxicity, and microbial answers, Biomed. Res. Int. 2013 (2013) 1–15, https://doi.org/10.1155/2013/542168.

[35] Y. Ohtsubo, K. Miyauchi, K. Kanda, T. Hatta, H. Kiyohara, T. Senda, Y. Nagata, Y. Mitsui, M. Takagi, PcpA, which is involved in the degradation of pentachlorophenol in *Sphingomonas chlorophenolica* ATCC39723, is a novel type of ring-cleavage dioxygenase, FEBS Lett. 459 (1999) 395–398, https://doi.org/10.1016/S0014-5793(99)01305-8.

[36] Y. Yin, N.Y. Zhou, Characterization of MnpC, a hydroquinone dioxygenase likely involved in the *meta*-nitrophenol segradation by *Cupriavidus necator* JMP134, Curr. Microbiol. 61 (2010) 471–476, https://doi.org/10.1007/s00284-010-9640-3.

[37] S. Dai, F.H. Vaillancourt, H. Maaroufi, N.M. Drouin, D.B. Neau, V. Snieckus, J.T. Bolin, L.D. Eltis, Identification and analysis of a bottleneck in PCB biodegradation, Nat. Struct. Biol. 9 (2002) 934.

[38] A.Y. Chang, V.W.Y. Chau, J.A. Landas, Y. Pang, Preparation of calcium competent *Escherichia coli* and heat-shock transformation, JEMI Methods. 1 (2017) 22–25https://jemi.microbiology.ubc.ca/sites/default/files/Chang%20et%20al%20JEMI-methods%20Vol%201%20pg%2022-25.pdf. (Accessed 2 June 2019).

[39] J.R. Marchesi, T. Sato, A.J. Weightman, T.A. Martin, J.C. Fry, S.J. Hiom, W.G. Wade, Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA, Appl. Environ. Microbiol. 64 (1998) 795–799.

[40] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T.L. Madden, BLAST+: architecture and applications, BMC Bioinformatics 10 (2009) 421, https://doi.org/10.1186/1471-2105-10-421.

[41] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal. Biochem. 72 (1976) 248–254, https://doi.org/10.1016/0003-2697(76)90527-3.

[42] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature 227 (1970) 680–685.

[43] L. Xun, J. Bohuslavek, M. Cai, Characterization of 2,6-dichloro-*p*-hydroquinone 1,2-dioxygenase (PcpA) of *Sphingomonas chlorophenolica* ATCC 39723, Biochem. Biophys. Res. Commun. 266 (1999) 322–325.

[44] T.E. Machonkin, P.L. Holland, K. N. Smith, J.S. Liberman, A. Dinescu, T.R. Cundari, S.S. Rocks, Determination of the active site of *Sphingobium chlorophenolicum* 2,6-dichlorohydroquinone dioxygenase (PcpA), J. Biol. Inorg. Chem. 15 (2010) 291–301.

[45] S. Liu, T. Su, C. Zhang, W. Zhang, D. Zhu, J. Su, T. Wei, K. Wang, Y. Huang, L. Guo, S. Xu, N. Zhou, L. Gu, Crystal structure of PnpCD, a two-subunit hydroquinone 1,2-dioxygenase, reveals a novel structural class of Fe^{2+}-dependent dioxygenases, J. Biol. Chem. 290 (2015) 24547–24560, https://doi.org/10.1074/jbc.M115.673558.

[46] B. Setlhare, A. Kumar, M.P. Molooena, B. Pillay, A.O. Olaniran, Phenol hydroxylase from *Pseudomonas* sp. KZNSA: purification, characterization and prediction of three-dimensional structure. - PubMed - NCBI, Int. J. Biol. Macromol. 146 (2020) 1000–1008, https://doi.org/10.1016/j.ijbiomac.2019.09.224.

[47] R.A. Laskowski, J. Jabłońska, L. Pravda, R.S. Vařeková, J.M. Thornton, PDBsum: structural summaries of PDB entries, Prot. Sci. 27 (2018) 129–134.

[48] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425.

[49] J. Felsenstein, Confidence limits on phylogenies: an approach using the bootstrap, Evolution (N. Y.). 39 (1985) 783–791.

[50] E. Zuckerkandl, L. Pauling, in: V. Bryson, H.J. Vogel (Eds.), Evolutionary Divergence and Convergence in Proteins. Edited in Evolving Genes and Proteins, Academic Press, New York 1965, pp. 97–166.

[51] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: molecular evolutionary genetics analysis version 6.0, Mol. Biol. Evol. 30 (12) (2013) 2725–2729, https://doi.org/10.1093/molbev/mst197.

[52] D.L. Daubaras, K. Saido, A.M. Chakrabarty, Purification of hydroxyquinol 1,2-dioxygenase and maleylacetate reductase: the lower pathway of 2,4,5-trichlorophenoxyacetic acid metabolism by *Burkholderia cepacia* AC1100, Appl. Environ. Microbiol. 62 (1996) 4276–4279.

[53] J-g. Lee, L. Xee, Purification and characterization of 2, 6-dichloro-*p*-hydroquinone chlorohydrolase from *Flavobacterium* sp, J. Bacteriol. 179 (1997) 1521–1524.

[54] Y. Bellin, C.A. O'Connor, G.A. Jin, Sorption and degradation of pentachlorophenol in sludge-amended soils, J Env. Qual. 19 (1990) 603–608.

[55] L. Xu, K. Resing, S.L. Lawson, P.C. Babbitt, S.D. Copley, Evidence that pcpA encodes 2,6-dichloro hydroquinone dioxygenase, the ring cleavage enzyme required for pentachlorophenol degradation in *Sphingomonas chlorophenolica* strain ATCC 39723, Biochemistry 38 (1999) 7659–7669, https://doi.org/10.1021/bi990103y.

[56] D.H. Pieper, K. Pollmann, P. Nikodem, B. Gonzalez, V. Wray, Monitoring key reactions in degradation of chloroaromatics by in situ H-1 nuclear magnetic resonance: solution structures of metabolites formed from *cis*-dienelactone, J. Bacteriol. 184 (2002) 1466.

[57] S. Schmidt, G.W. Kirby, Dioxygenative cleavage of C-methylated hydroquinones and 2,6-dichlorohydroquinone by *Pseudomonas* sp. HH35, Biochim. Biophys. Acta - Gen. Subj. 1568 (2001) 83–89, https://doi.org/10.1016/S0304-4165(01)00204-5.

[58] S. Seltzer, Nuclear magnetic resonance, paramagnetic ion induced relaxation method to differentiate between 1,3-diketo and 1,3-keto-enol isomers, J. Org. Chem. 46 (1981) 2643–2650.

[59] L. Holm, L.M. Laakso, Dali server update, Nucleic Acids Res. 44 (2016) W351–W355, https://doi.org/10.1093/nar/gkw357.

[60] A. Drozdetskiy, C. Cole, J. Procter, G.J. Barton, JPred4: a protein secondary structure prediction server, Nucleic Acids Res. 43 (2015) W389–W394, https://doi.org/10.1093/nar/gkv332.

[61] A. Marchler-Bauer, Y. Bo, L. Han, J. He, C.J. Lanczycki, S. Lu, F. Chitsaz, M.K. Derbyshire, R.C. Geer, N.R. Gonzales, M. Gwadz, D.I. Hurwitz, F. Lu, G.H. Marchler, J.S. Song, N. Thanki, Z. Wang, R.A. Yamashita, D. Zhang, C. Zheng, L.Y. Geer, S.H. Bryant, CDD/SPARCLE: functional classification of proteins via subfamily domain architectures, Nucleic Acids Res. 45 (2017) D200–D203, https://doi.org/10.1093/nar/gkw1129.

[62] A. Marchler-Bauer, J.B. Anderson, M.K. Derbyshire, C. DeWeese-Scott, N.R. Gonzales, M. Gwadz, L. Hao, S. He, D.I. Hurwitz, J.D. Jackson, Z. Ke, D. Krylov, C.J. Lanczycki, C.A. Liebert, C. Liu, F. Lu, S. Lu, G.H. Marchler, M. Mullokandov, J.S. Song, N. Thanki, R.A. Yamashita, J.J. Yin, D. Zhang, S.H. Bryant, CDD: a conserved domain database for interactive domain family analysis, Nucleic Acids Res. 35 (2007) D237–D240, https://doi.org/10.1093/nar/gkl951.

[63] J. Huang, J. Villemain, R. Padilla, R. Sousa, Mechanisms by which T7 lysozyme specifically regulates T7 RNA polymerase during different phases of transcription, J. Mol. Biol. 293 (1999) 457–475, https://doi.org/10.1006/jmbi.1999.3135.

[64] P.T. Wingfield, Overview of the purification of recombinant proteins, Curr. Protoc. Protein Sci. 2015 (2015) 6.1.1–6.1.35, https://doi.org/10.1002/0471140864.ps0601s80.

SUPPLEMENTARY MATERIAL

Journal: International Journal of Biological Macromolecules

Cloning, overexpression, purification, characterization, and structural modelling of a metabolically active Fe^{2+} dependent 2,6-dichloro-*p*-hydroquinone 1,2-dioxygenase (CpsA) from *Bacillus cereus* strain AOA-CPS1.

Oladipupo A. Aregbesola, Ajit K. Kanwal, Mduduzi P. Mokoena and Ademola O. Olaniran*

Discipline of Microbiology, School of Life Sciences, College of Agriculture, Engineering and Science, University of KwaZulu-Natal (Westville Campus), Private Bag X54001, Durban 4000, Republic of South Africa.

***Corresponding author:**
Ademola O. Olaniran
Phone: +27 31260 7400
Fax: +27 31260 7809
E-mail: olanirana@ukzn.ac.za

Supplementary material 1 (Fig. S1)

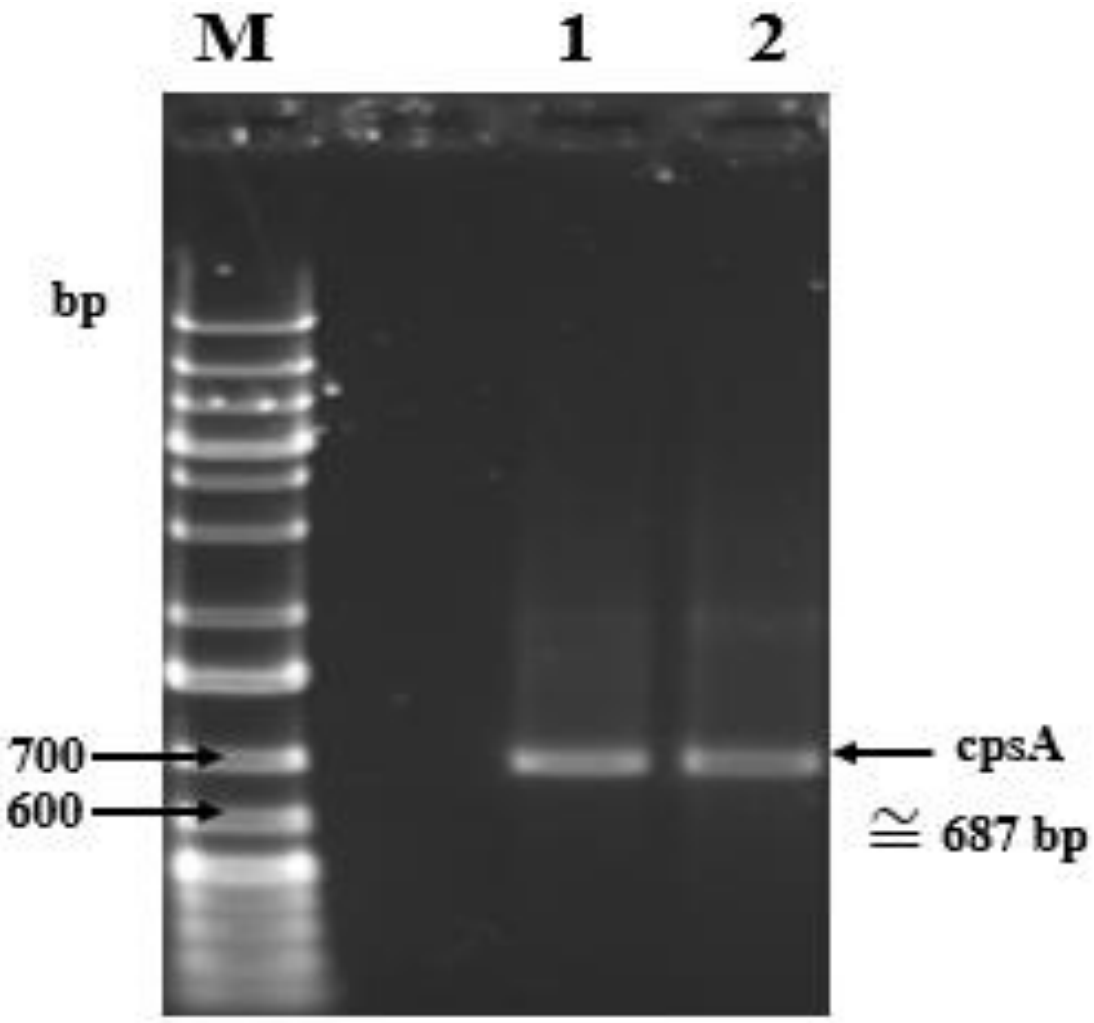


Fig. S1: PCR detection of cpsA gene using degenerate primer pair (forward: 5'-TCTGTYMCRAWTTCAAAWA-3' and reverse: 5'-ATRADRVAGGWARNSCHGGWA-3'). Lane M: DNA ladder; lanes 1-2 amplified cpsA fragment detected using the primer designed from the consensus sequences of *B. thuringiensis* strains ATCC 10792, L-7601, and *Bacillus* sp. FDAARGOS_235.

Supplementary material 2 (Fig. S2)

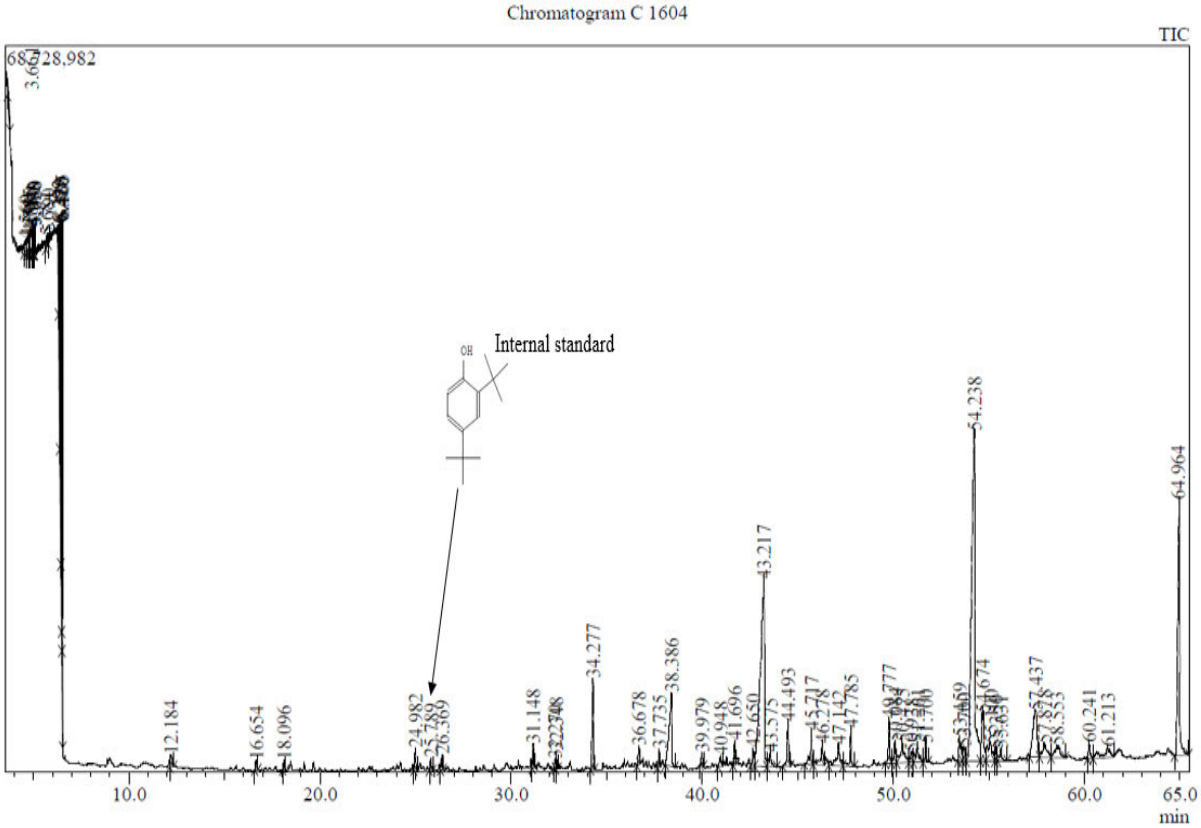


Fig. S2a: GC-MS chromatogram of 2,6-Dichloro-p-hydroquinone transformation by 2,6-Dichloro-p-hydroquinone 1,2-dioxygenase (CpsA) from *Bacillus cereus* strain OAO-CPS_1

Supplementary material 2

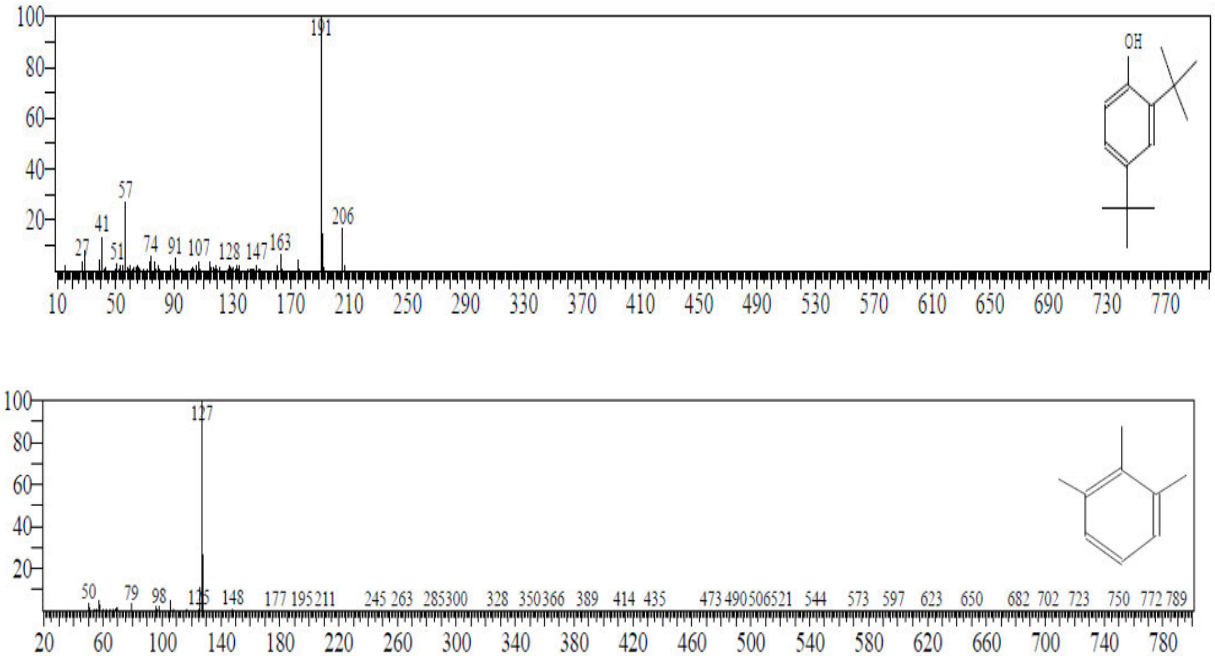


Fig. S2b: Internal standard

Line#:22 R. Time: 25.790 (Scan#:4459) Mass Peaks: 435

Raw Mode: Averaged 25.785 - 25.795 (4458 - 4460) Base Peak: 191.10 (320262)

BG Mode:Calc. from Peak Group 1 - Event 1 Scan

Hit#:2 Entry: 18377 Library:NIST11s.lib

Comp Name: Phenol, 2,4-bis(1,1-dimethylethyl)- $$ Phenol, 2,4-di-tert-butyl- $$ 2,4-Di-tert-butylphenol $$ 2,4-di-t-Butylphenol $$ 1-Hydroxy-2,4-di-tert-buty

<< Target >>
Line#:14 R. Time: 6.420 (Scan#:585) Mass Peaks: 387

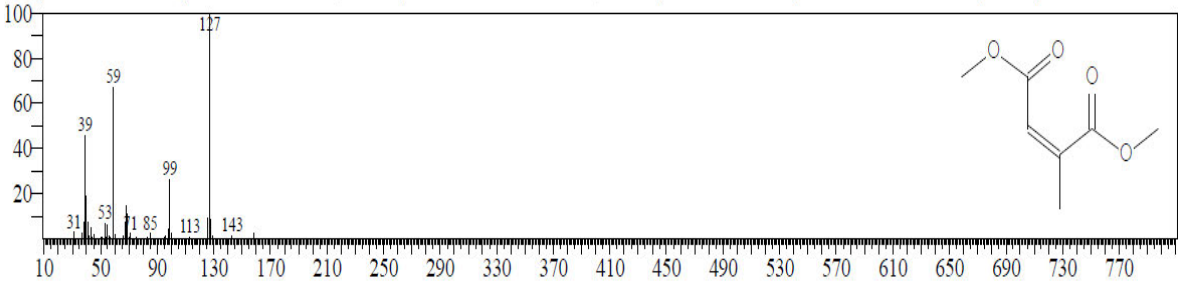
Raw Mode: Averaged 6.415 - 6.425 (584-586) Base Peak: 127.05 (159213)

BG Mode: Calc. from Peak Group 1 - Event 1 Scan

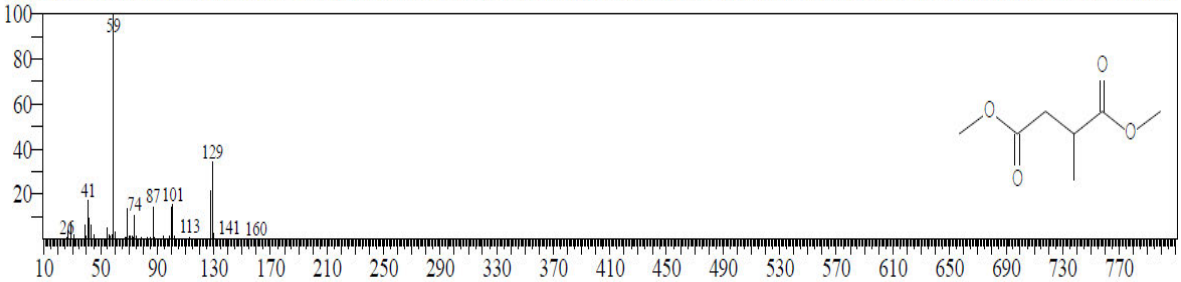
CompName:2,6-DiChlorotoluene $$ 1,3-DiChloro-2-methylbenzene # $$

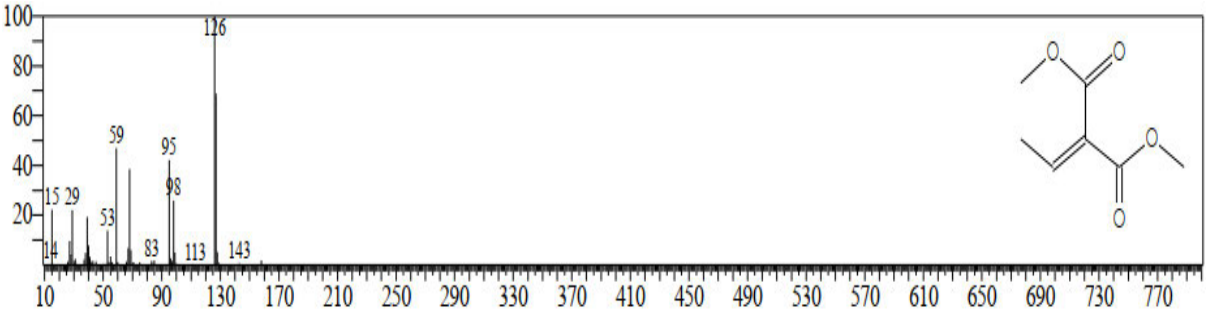
Supplementary material 2

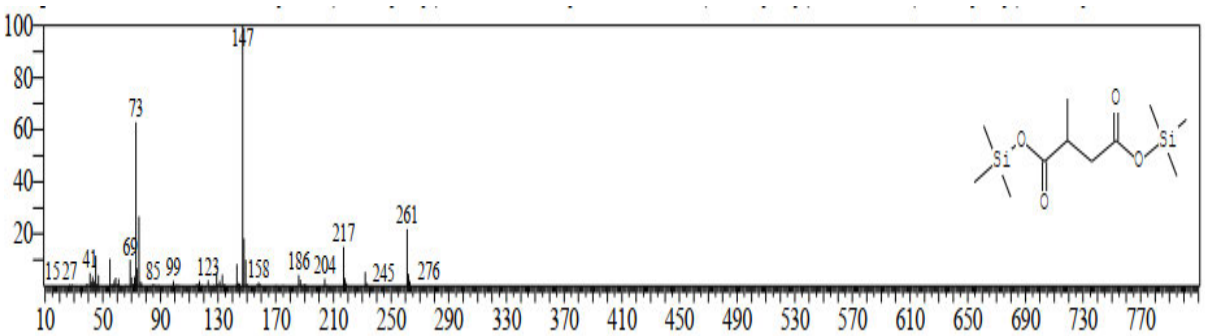
Hit#:3 Entry:10547 Library:NIST11s.lib
SI:67 Formula:C7H10O4 CAS:617-54-9 MolWeight:158 RetIndex:1037
CompName:cis-2-Methyl-2-butenedioic acid, dimethyl ester $$ Citraconic acid, dimethyl ester $$ Dimethyl citraconate $$ Dimethyl methyl maleate $$ Metl



Hit#:4 Entry:10840 Library:NIST11s.lib
SI:67 Formula:C7H12O4 CAS:1604-11-1 MolWeight:160 RetIndex:988
CompName:Butanedioic acid, methyl-, dimethyl ester $$ Succinic acid, methyl-, dimethyl ester $$ Dimethyl methylsuccinate $$ Dimethyl 2-methylsuccinate







methyl-, bis(trimethylsilyl) ester.

Fig. S2c: Isoforms of the transformation metabolite

Supplementary material 3 (Fig. S3)

Fig. S3 (a): Protein, Peptides and PSMs identification summary

Dataset properties	Protein	Peptides	PSMs
Total number of true positives	15.68	82.67	441.67
Resolution (%)	20.0	7.69	5.26
Validated hits	16.0	72.0	323.0
Confidence (%)	73.68	76.92	90.91
False positive (%)	0.0	0.0	3.0
False discovery rate (%)	0.0	0.0	0.93
True positive	16.0	72.0	320
False negative rate (%)	0.0	14.08	27.5

Supplementary material 3

Peptide Spectrum Matches

	ID	Sequence	Protein(s)	Confidence
1		NH2-GIHHVTAITSSAEK-COOH	Q81H02	100

Spectrum Identification Results

Validated

	SE	Rnk	Sequence	Charge	Confidence
1		1	NH2-GIHHVTAITSSAEK-COOH	2	100
2		1	NH2-GIHHVTAITSSAEK-COOH	2	100
3		1	NH2-GIHHVTAITSSAEK-COOH	2	99
4		1	NH2-GIHHVTAITSSAEK-COOH	2	95
5		1	NH2-GIHHVTAITSSAEK-COOH	2	94
6		1	NH2-GIHHVTAITSSAEK-COOH	2	93
7		2	NH2-GIHHVTAITSSAEK-COOH	2	87
8		8	NH2-FLIIGTAIVLFIM-COOH	2	74
9		1	NH2-GIHHVTAITSSAEK-COOH	2	65

OMSSA X!Tandem Andromeda MS Amanda MS-GF+ DirecTag Comet MyriMatch Tide Novor PeptideShaker

Fig. S3 (b): Peptides structure matches and spectrum Identification results

Supplementary material 3

Fig. S3 (c): Spectrum identification results

S/N	SE	Rank	Sequence	Charge	Confidence (%)
1	5	1	NH2-GIHHVTAITSSAEK-COOH	2	100.0
2	7	1	NH2-GIHHVTAITSSAEK-COOH	2	100.0
3	2	1	NH2-GIHHVTAITSSAEK-COOH	2	99.328
4	10	1	NH2-GIHHVTAITSSAEK-COOH	2	94.736
5	28	1	NH2-GIHHVTAITSSAEK-COOH	2	94.117
6	13	1	NH2-GIHHVTAITSSAEK-COOH	2	93.333
7	8	2	NH2-GIHHVTAITSSAEK-COOH	2	87.234
8	8	8	NH2-FLIIGTAIVLFIM-COOH	2	74.468
9	4	1	NH2-GIHHVTAITSSAEK-COOH	2	64.705
10	1	1	NH2-GIHHVTAITSSAEK-COOH	2	55.555

Supplementary material 3

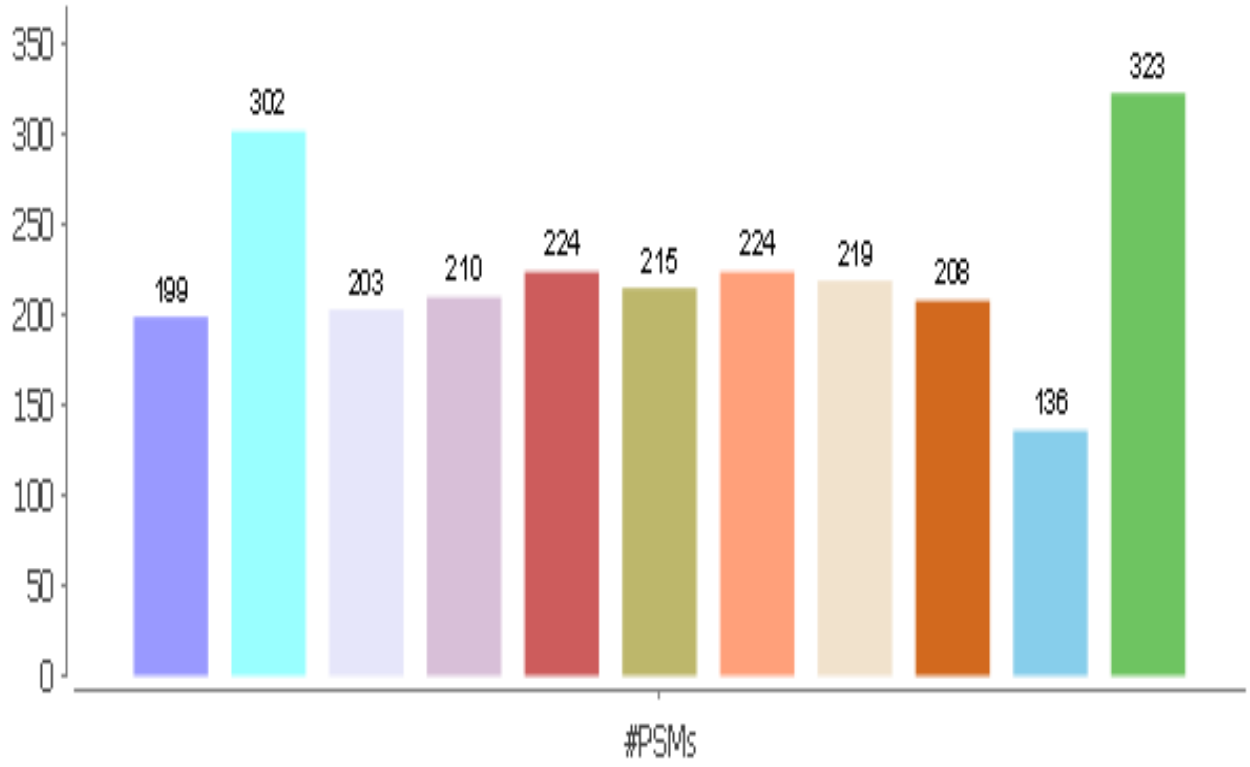


Fig. S3 (d): Spectrum identification overview for Peptide-Spectrum Matches (PSMs)

Supplementary material 3

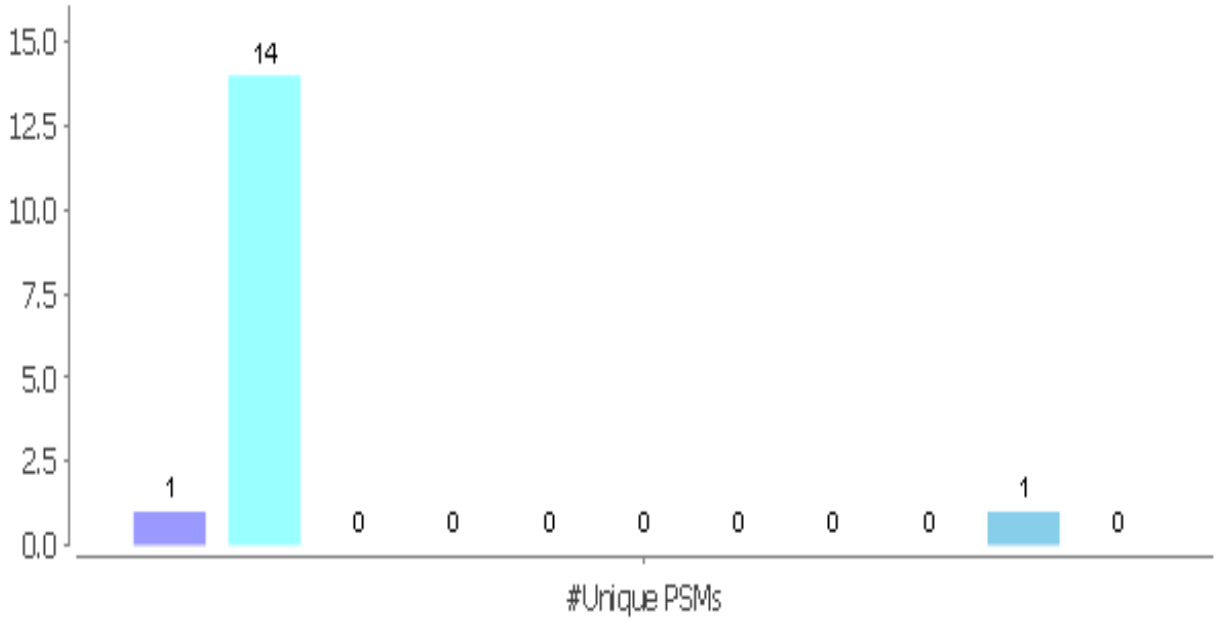


Fig. S3 (e): Spectrum identification overview for Peptide-Spectrum Matches (PSMs)

Supplementary material 3

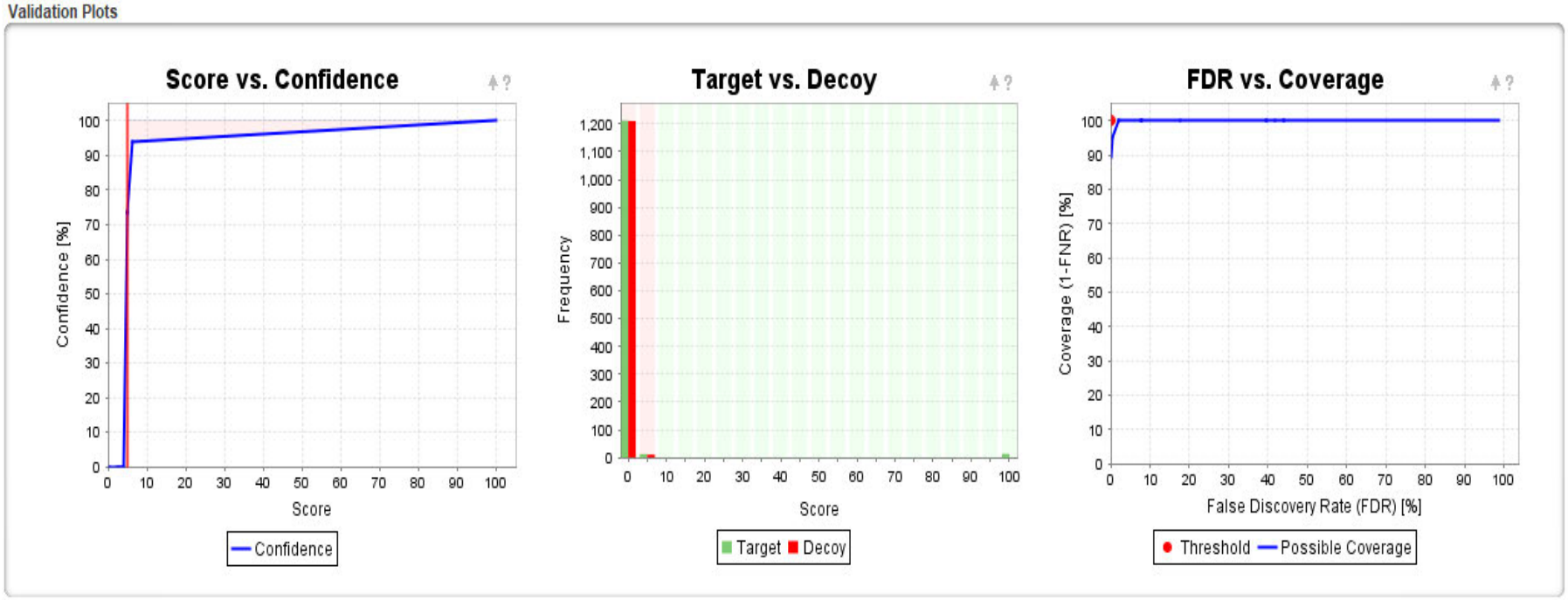


Fig. S3 (f): Protein validation plots

Supplementary material 3

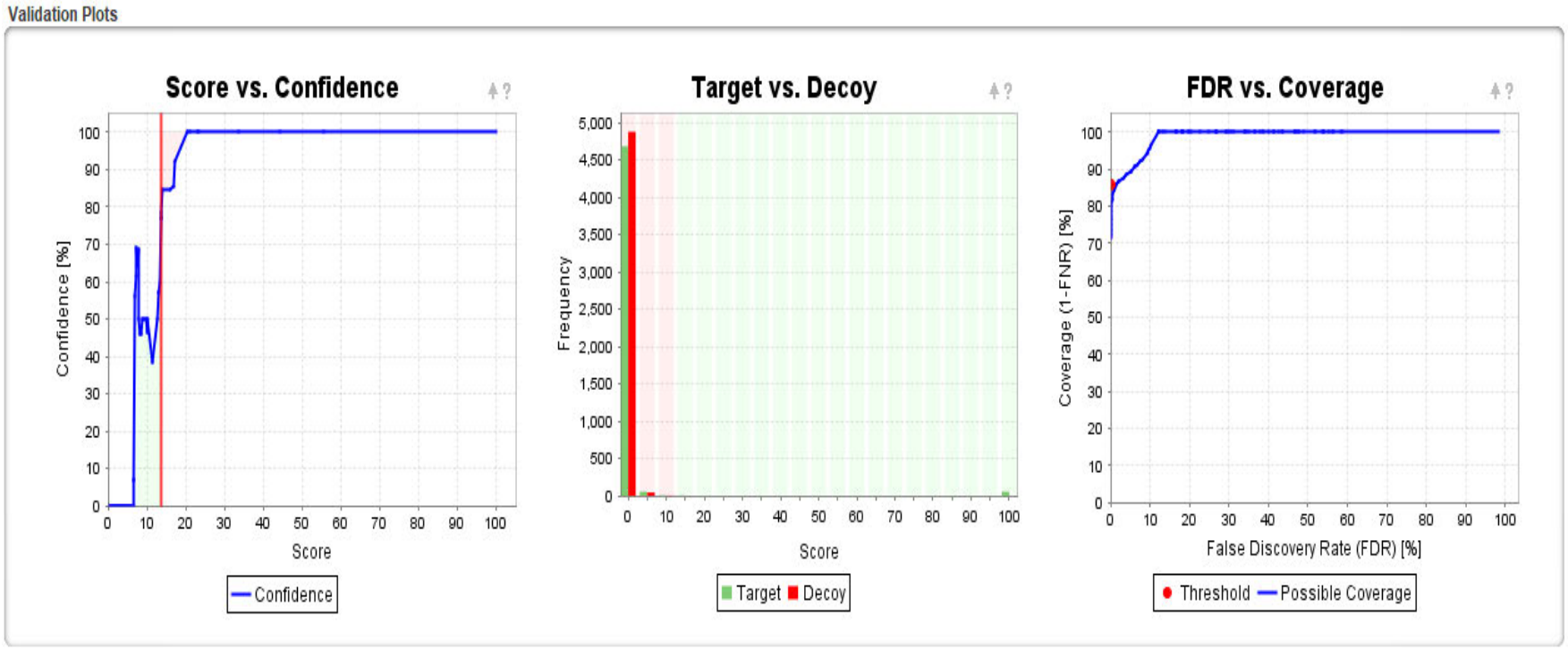


Fig. S3 (g): Peptides validation plots

Supplementary material 3

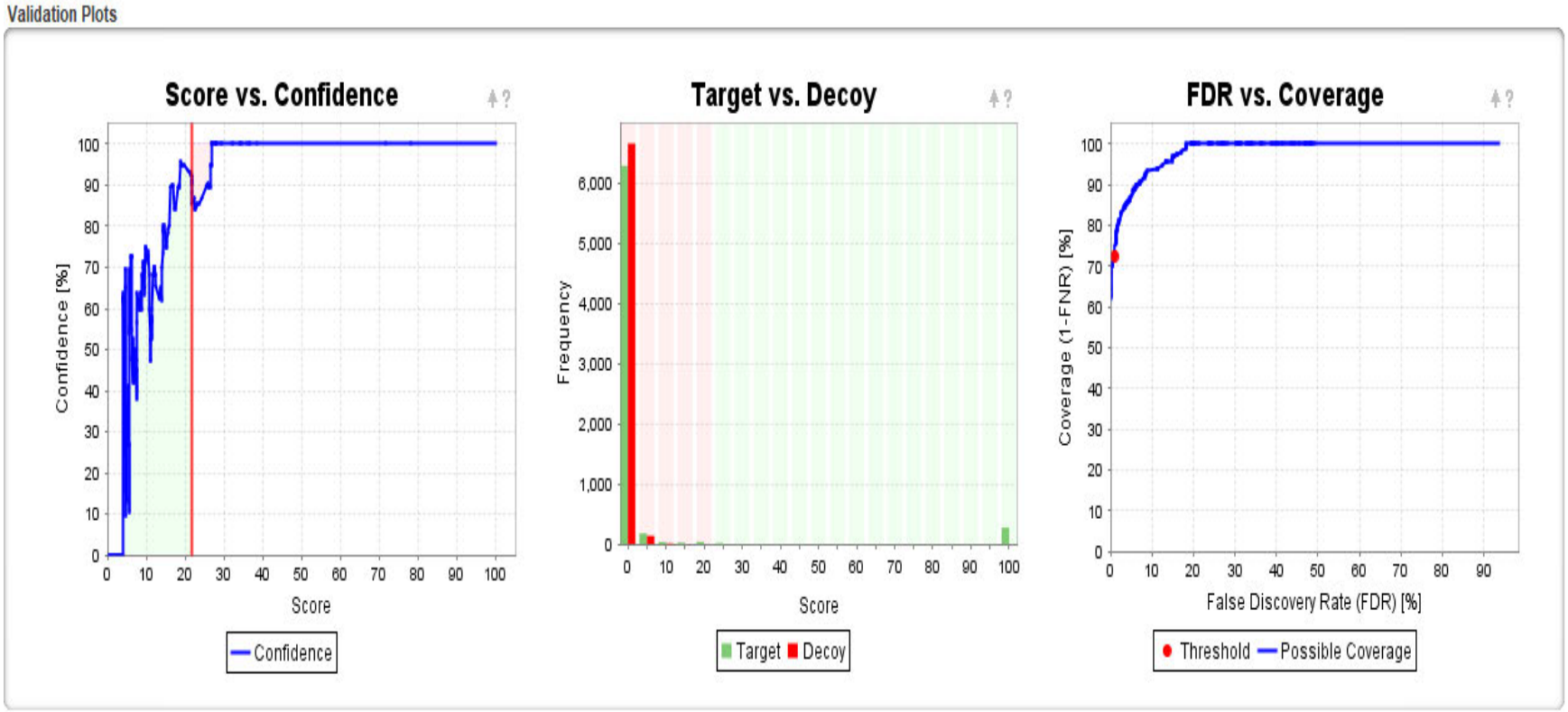


Fig. S3 (h): Target-Decoy PSMs validation plots

Supplementary material 3

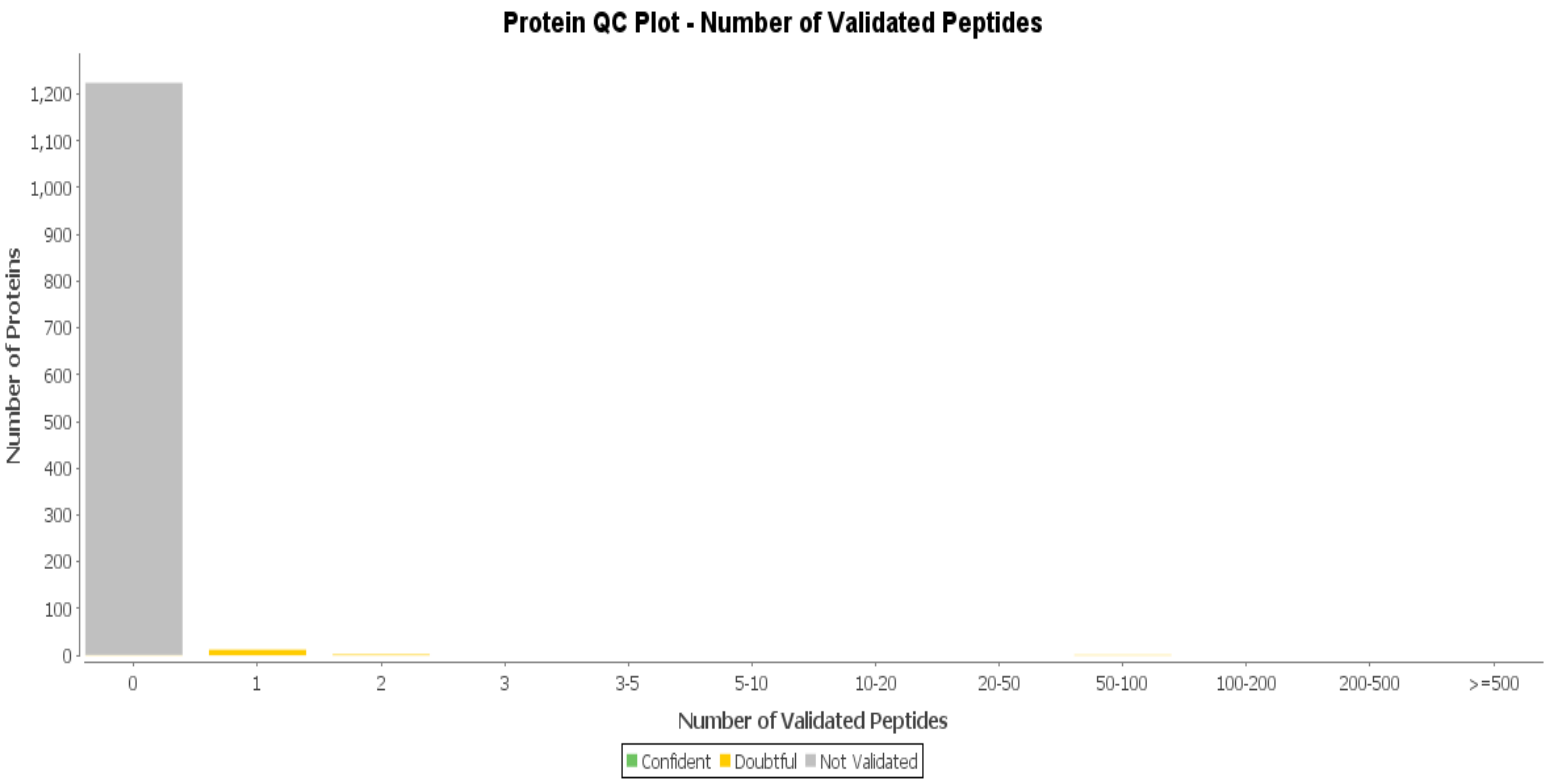


Fig. S3 (i): Protein quality control vs number of validated peptides

Supplementary material 3

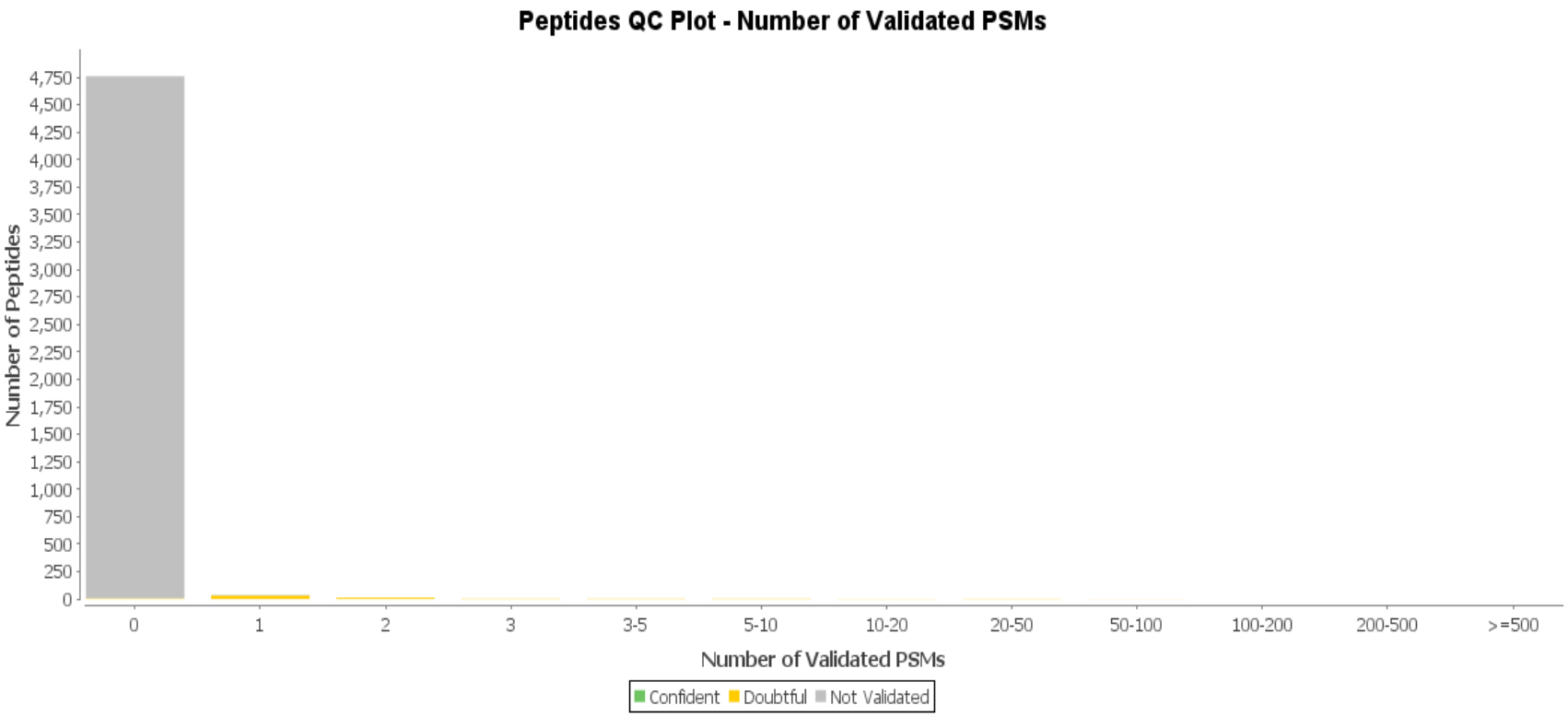


Fig. S3 (j): Peptides quality control plot vs number of validated peptides-spectrum matches (PSMs)

Supplementary material 3

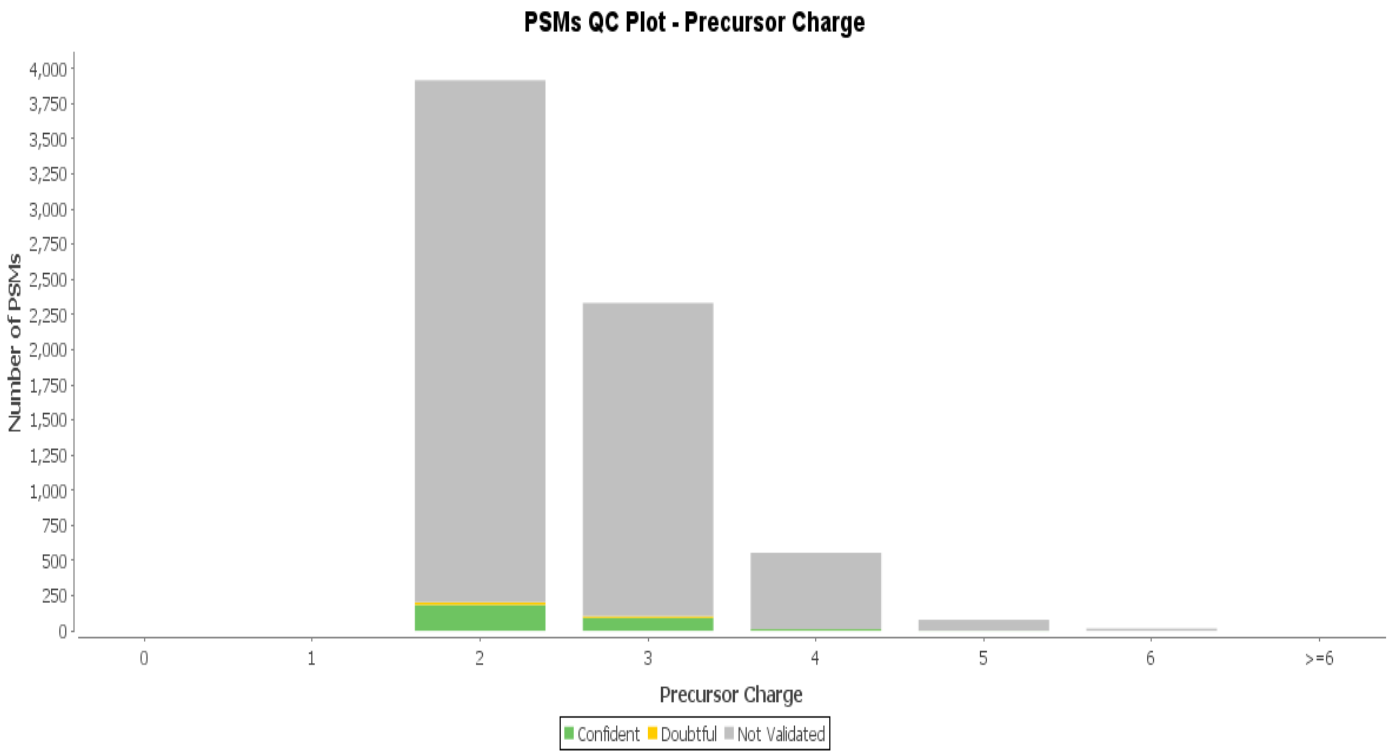


Fig. S3 (k): Peptides-Spectrum matches quality control plot vs precursor charge

Supplementary material 3

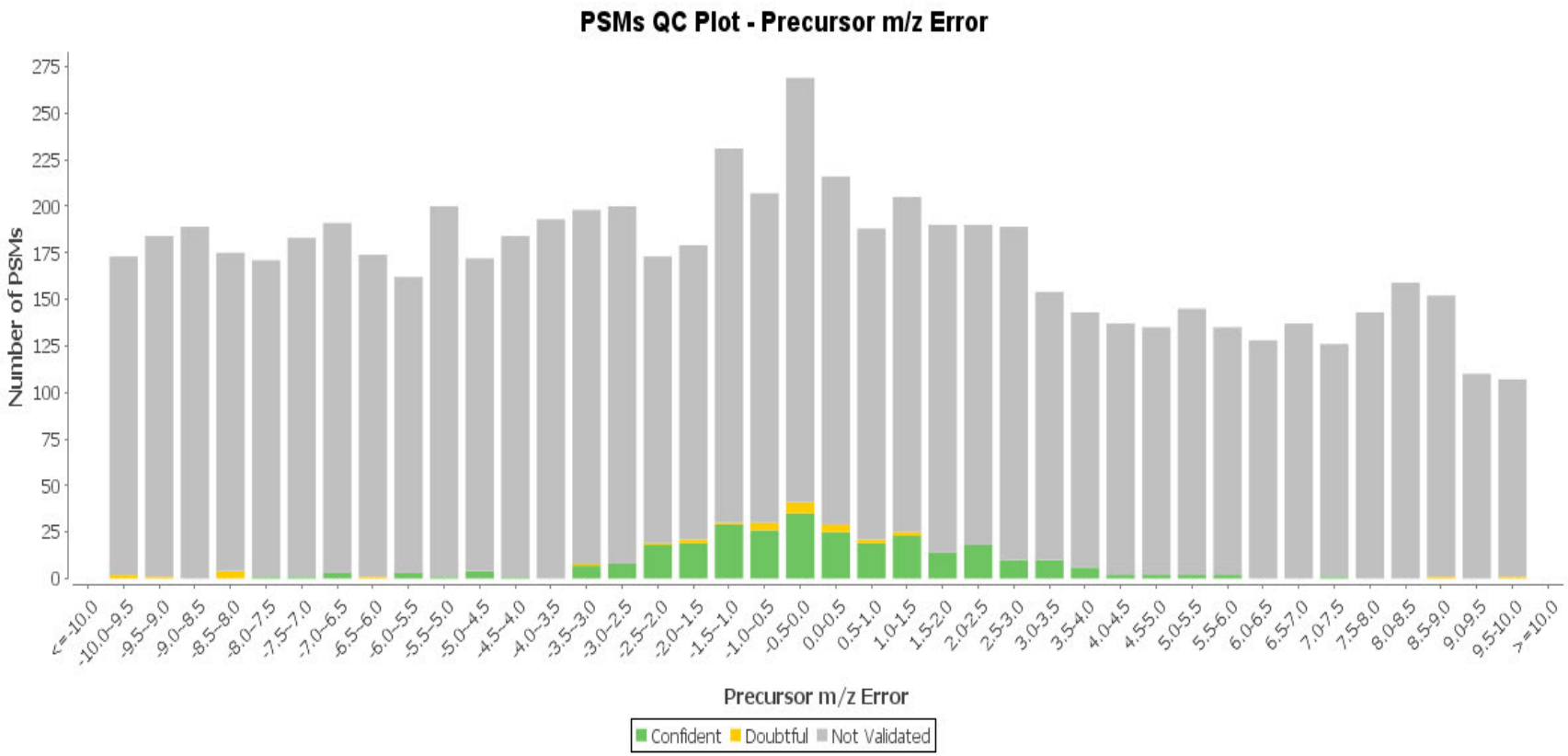


Fig. S3 (l): Peptides-Spectrum matches quality control vs m/z Error

Supplementary material 4

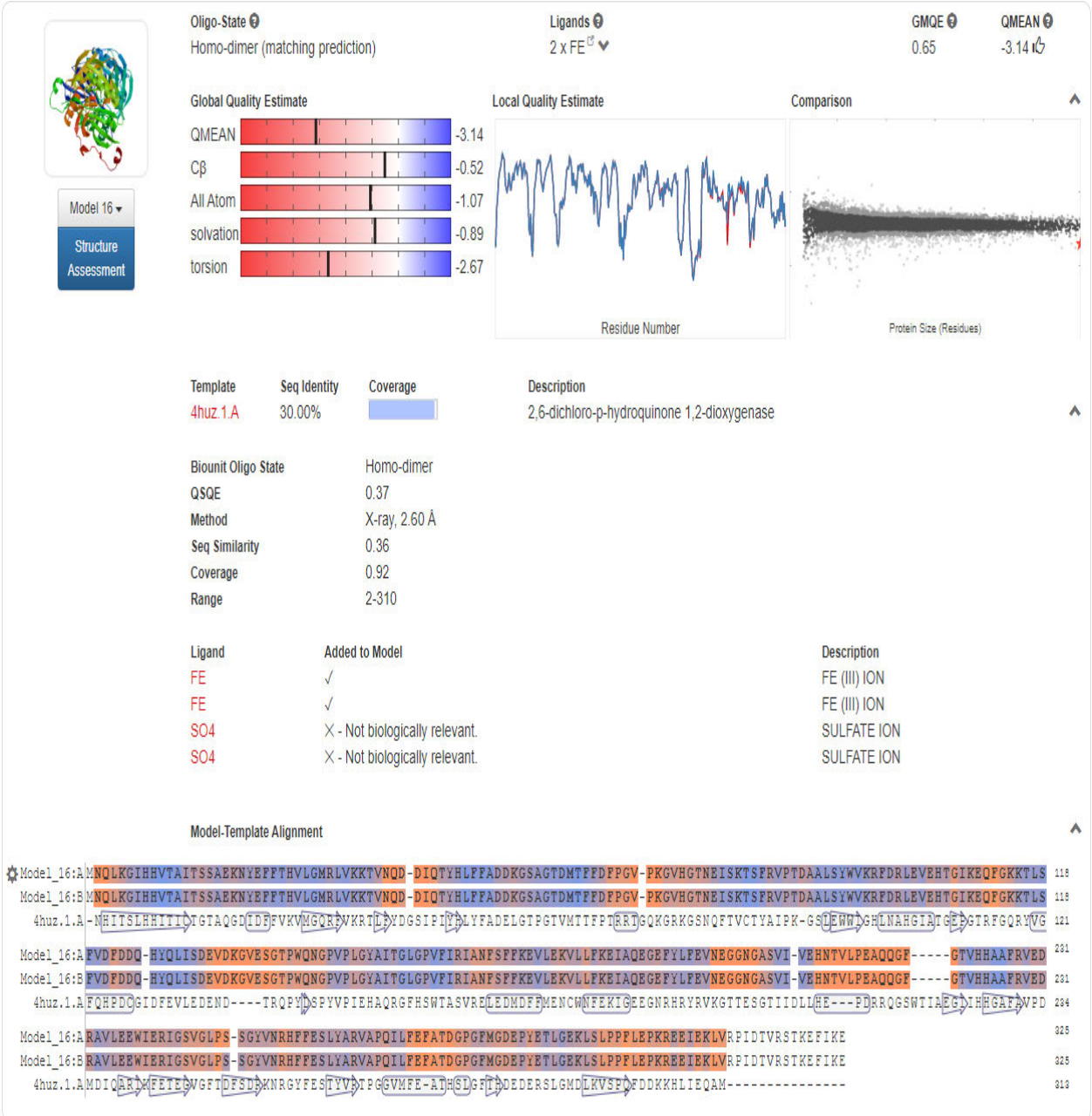


Fig. S4 a: Model built with 4HUZ as a template .

Supplementary material 4

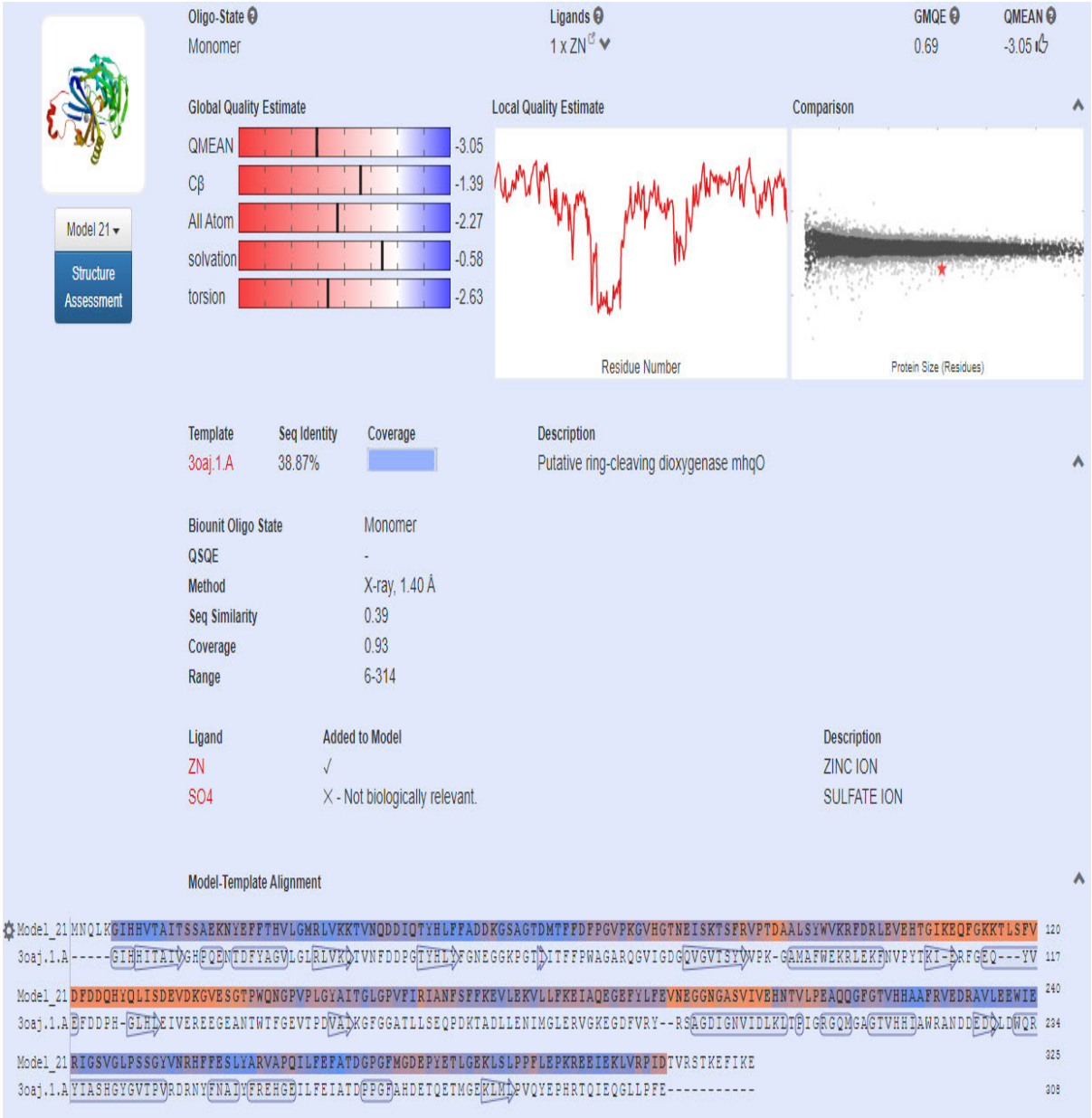


Fig. S4 b: Model built with 3OAJ as a template

Supplementary material 4

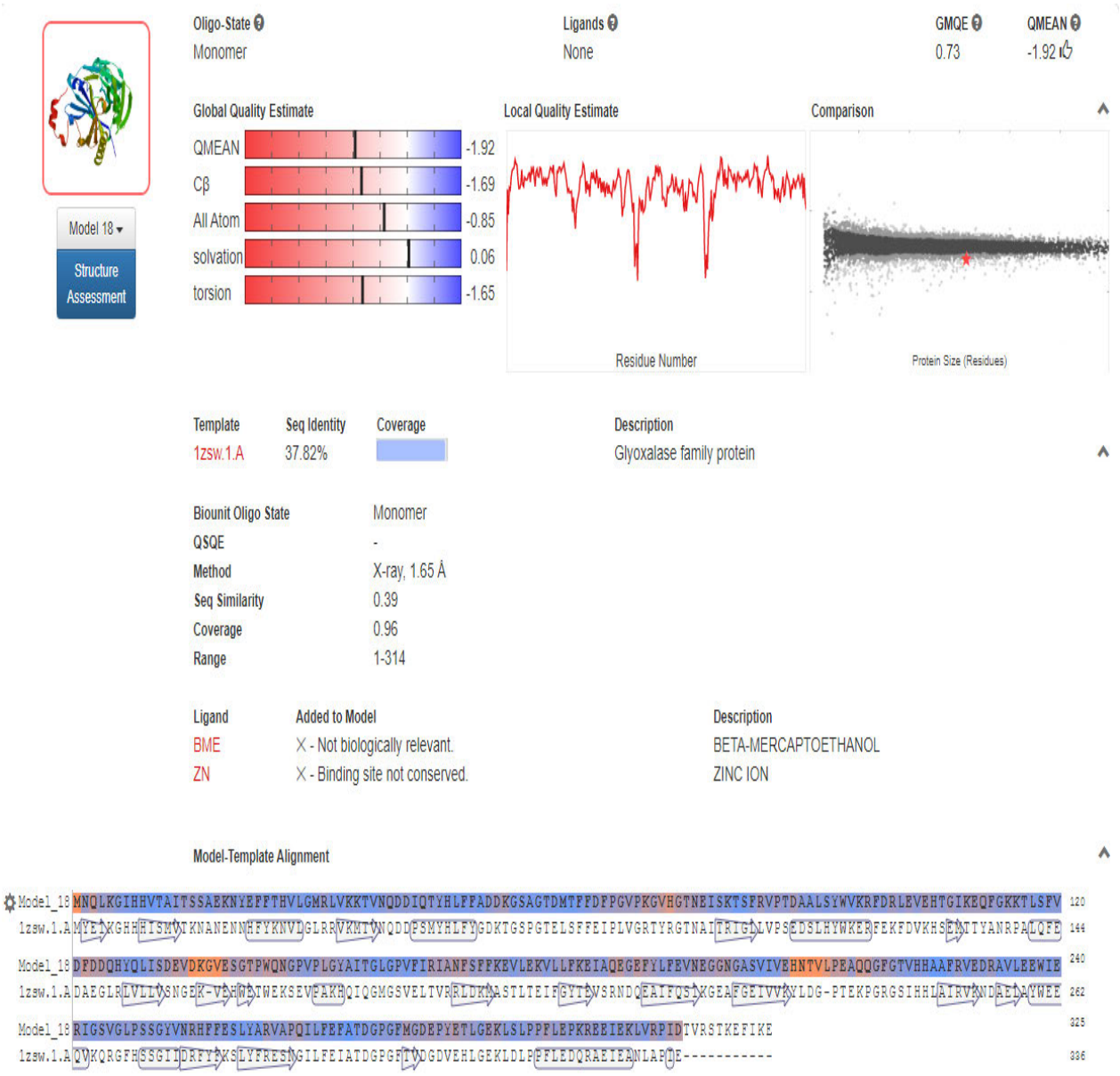


Fig. S4 c: Model built with 1ZSW as a template.

Supplementary material 5

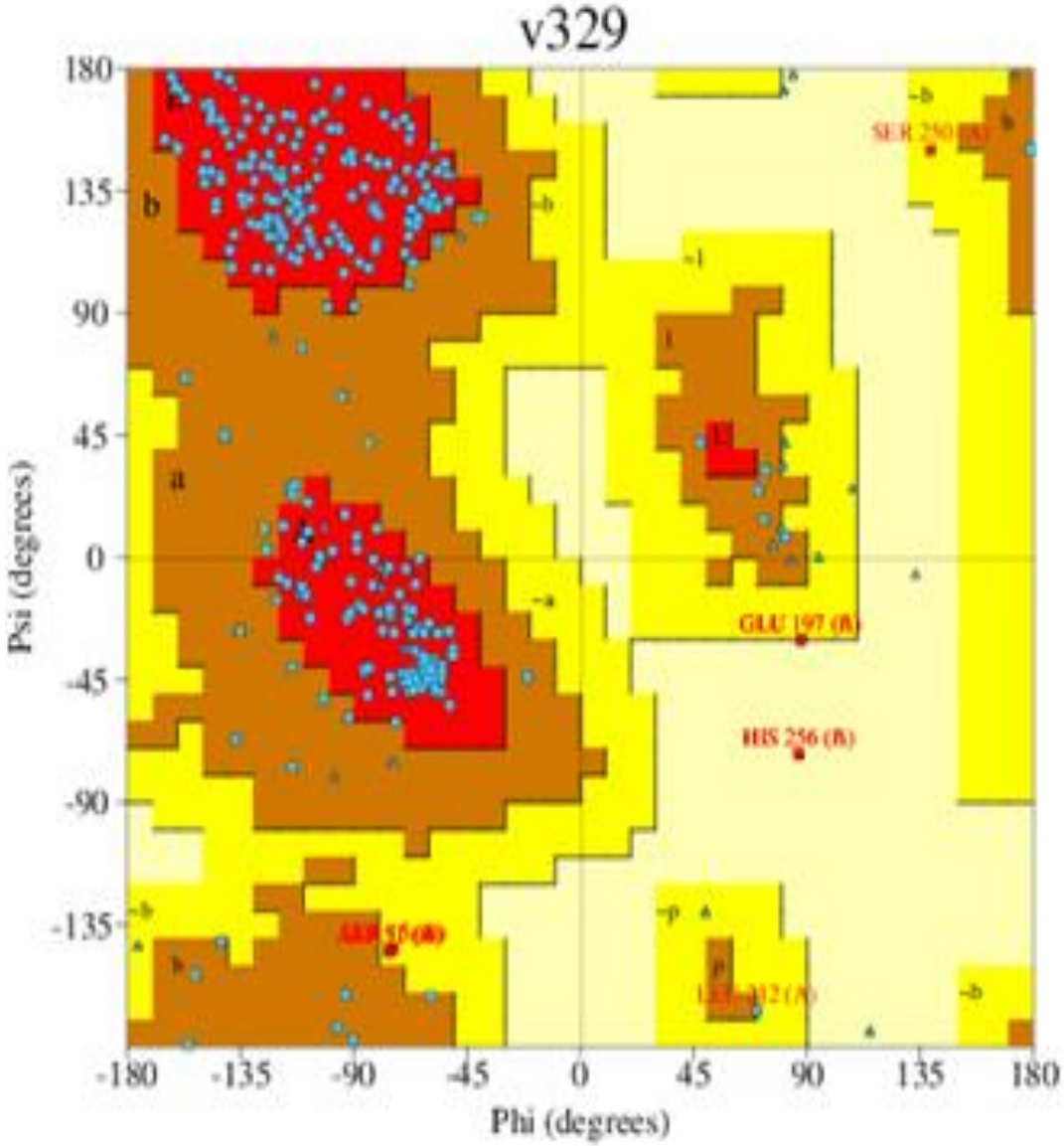


Fig. S 5a: Ramachandran Plot statistics of CpsA

Supplementary material 5

Fig. S5 (b): Table of beta sheets

Sheet	View	No. of strands	Type	Barrel	Topology
A		8	Mixed	No	-3X 1 1 2 3X -1 -1
B		8	Mixed	No	3X -1 -1 -2 -3X 1 1

		Strand 1			Strand 2			Loop
No.	View	Start	End	Length	Start	End	Lenght	Lenght
1		His 126	Ser 131	6	Leu 158	Ile 165	8	26

Fig. S5 (c): Table of psi Loops. The table gives the beginning and end residues, and lengths, of the two strands involved in the hairpin together the with hairpin's class as defined above.

Supplementary material 5

	Strand 1			Strand 2			Hairpin
No.	Start	End	Length	Start	End	Lenght	Class
* 1.	Arg31	Gln39	9	Gln43	Ala50	8	1:3
* 2.	Gln43	Ala50	8	Asp59	Asp64	6	10:10
* 3.	Lys109	Gln111	3	Lys114	Val120	7	2:2 IP
* 4.	Lys114	Val120	7	His126	Ser131	6	3:5
5.	Phe181	Glu187	7	Phe190	Val195	6	2:2 IIP
6.	Phe190	Val195	6	Ser203	His208	6	6:8
7.	Val253	Asn254	2	Glu259	Leu261	3	2:4
8.	Glu259	Leu261	3	Phe273	Asp276	4	12:12

Fig. S5 (d) : Table of beta hairpins. The table gives the beginning and end residues, and lengths, of the two strands involved in the hairpin together the with hairpin's class.

Supplementary material 5

No.	Bulge type	Res X	Res 1	Res 2	Res 3	Res 4
* 1.	Antiparallel classic	Thr11A	His223A	His224A		
* 2.	Antiparallel classic	Phe48A	Val33A	Lys34A		
* 3.	Antiparallel classic	Ser82A	Gly159A	Pro160A		
* 4.	Antiparallel classic	Arg84A	Thr156A	Gly157A		
5.	Antiparallel G1	Asp121A	Asp124A	Gln125A		
6.	Antiparallel classic	Phe162A	Ser79A	Lys80A		
7.	Antiparallel classic	Leu192A	Ile184A	Ala185A		
8.	Antiparallel classic	Ala226A	His8A	His9A		
9.	Antiparallel classic	Arg228A	Lys5A	Gly6A		
10.	Antiparallel classic	Arg255A	Phe257A	Phe258A		

Fig. S5 (e): Table of beta bulges. The table gives the bulge type, and residues X, 1, 2, 3 and 4. Most of the bulges will have just residues X, 1 and 2; the special bulges may have one or more of residues 3 and 4, and the bent bulges will just have residues 1 and 2. A beta bulge is a region of irregularity in a beta sheet, where the normal pattern of hydrogen bonding is disrupted e.g. by the insertion of an extra residue.

Supplementary material 5

No.	Start	End	Sheet	Edge	Sequence
1.	Gly6	Thr14	A	No	GIHHVTAIT
2.	Arg31	Gln39	A	Yes	RLVKKTVNQ
3.	Gln43	Ala50	A	No	QTYHLFFA
4.	Asp59	Asp64	A	No	DMTFFD
5.	Ile78	Val85	B	No	ISKTSFRV
6.	Lys109	Gln111	B	Yes	KEQ
7.	Lys114	Val120	B	No	KKTLSFV
8.	His126	Ser131	B	No	HYQLIS
9.	Leu158	Ile165	B	No	LGPVFIRI
10.	Phe181	Glu187	B	Yes	FKEIAQE
11.	Phe190	Val195	B	No	FYLFEV
12.	Ser203	His208	B	No	SVIVEH
13.	Val222	Arg228	A	No	VHHAAFR
14.	Val253	Asn254	A	Yes	VN
15.	Glu259	Leu261	A	No	ESL
16.	Phe273	Asp276	A	No	FATD

Fig. S5 (f): Table of beta strands. Note: for each beta strand in the protein chain the table gives the strand number (assigned sequentially from the N-terminus of the protein), the start and end residues, the letter corresponding to the sheet in which the strand is involved, the number of residues in the strand, whether the strand is at the edge of a beta sheet or not and its amino acid sequence.

Supplementary material 5

No.	Start	End	Type	Number of Residues	Length	Unit rise	Residues per turn	Pitch	Deviation from ideal	Sequence
*1.	Ala17	Phe24	H	8	12.03	1.46	3.73	5.43	13.5	AEKNYEFF
*2.	Asp88	Asp99	H	12	18.54	1.52	3.56	5.43	10.7	DAALSYWVKRFD
*3.	Asp135	Gly137	G	3	-	-	-	-	-	DKG
*4.	Glu139	Gly141	G	3	-	-	-	-	-	ESG
5.	Phe168	Lys177	H	10	14.95	1.47	3.64	5.36	5.3	FSFFKEVLEK
6.	Pro213	Ala215	G	3	-	-	-	-	-	PEA
7.	Arg232	Ser244	H	13	19.93	1.49	3.60	5.36	7.0	RAVLEEWIERIGS
8.	Pro297	Leu299	G	3	-	-	-	-	-	PFL
9.	Glu304	Leu309	H	6	10.23	1.63	3.43	5.58	5.3	EEIEKL

Fig. S5 (g): Table of helices. Note: Table for each helix includes the helix number (assigned sequentially starting with 1 at the N-terminus of the protein), the residue numbers corresponding to the start and end of the helices, the helix type (**H** (alpha helix) or **G** (3, 10) helix. This is followed by the number of residues in the helix and information about the geometry of the helix as follows: length and unit rise (both in Angstroms), the number of residues per turn (ideally 3.6) for alpha helices), the helix pitch in Angstroms and a measure of the deviation of the helix geometry from an ideal helix (in degrees). This latter value should be 0 for a perfect helix. The geometrical parameters are not calculated for helices with fewer than four residues. The final column in the table gives the helix's amino acid sequence.

Supplementary material 5

				Residue i+1			Residue i+2			i to i+3	
No.	Turn	Sequence*	Turn type	Phi	Psi	Chi1	Phi	Psi	Chi1	CA-dist	H-bond
* 1.	Gln39-Ile42	QDDI	IV	-91.9	-58.9	-63.4	-137.0	-67.1	-79.0	5.7	Yes
* 2.	Asp40-Gln43	DDIQ	IV	-137.0	-67.1	-79.0	-77.9	163.0	65.7	6.3	Yes
* 3.	Asp51-Gly54	DDKG	I	-66.2	-46.9	-171.8	-125.6	10.8	56.7	5.6	Yes
* 4.	Ser55-Thr58	SAGT	II	-62.6	131.4	-	94.6	-0.4	-	5.8	No
* 5.	Ala56-Asp59	AGTD	IV	94.6	-0.4	-	-120.4	-15.4	79.5	5.9	Yes
6.	Phe65-Val68	FPGV	I	-69.5	-4.1	38.3	-79.5	-3.6		5.6	No
7.	Pro66-Pro69	PGVP	IV	-79.5	-3.6	-	-142.6	-141.3		6.8	Yes
8.	Glu110-Gly113	EQFG	IV	-130.9	122.9	-64.5	47.4	42.1		5.5	Yes
9.	Gln111-Lys114	QFGK	I'	47.4	42.1	-71.0	76.4	4.3		5.3	No
10.	Asp121-Asp124	DFDD	I	-82.1	-0.8	-69.9	-117.9	11.7		5.3	No
11	Asp132-Asp135	DEVD	I	-75.5	-21.0	57.3	-78.7	-26.9	177.7	5.4	Yes
12	Asn146-Val149	NGPV	IV	-74.5	-75.8	-	-21.2	-43.9	-40.2	6.5	Yes
13	Gly147-Pro150	GPVP	VIII	-21.2	-43.9	-40.2	-123.8	118.3	175.2	6.8	Yes
14	Pro150-Tyr153	PLGY	I	-62.2	-28.0	-157.7	-65.9	-20.7	-	5.8	No
15	Leu151-Ala154	LGYA	I	-65.9	-20.7	-	-116.7	-9.4	-65.1	5.1	Yes
16	Glu187-Phe190	IANF	VIII	-85.5	-42.2	57.3	-100.6	91.9	-168.6	6.6	Yes
17	Val195-Gly198	EGEF	II'	49.7	-130.8	-	-93.6	15.9	-65.1	5.3	No

				Residue i+1			Residue i+2			i to i+3	
No.	Turn	Sequence*	Turn type	Phi	Psi	Chi1	Phi	Psi	Chi1	CA-dist	H-bond
17	Val195-Gly198	EGEF	II'	49.7	-130.8	-	-93.6	15.9	-65.1	5.3	No
18	Gly198-Gly201	VNEG	II	-57.0	128.5	-40.2	87.6	-30.3	-173.1	5.6	No
19	Gly220-His223	GGNG	IV	-112.4	-14.0	-157.7	-75.7	-3.6	-71.8	6.5	Yes
20	Glu187-Phe190	GTVH	VIII	-75.1	-18.9	-	-81.5	127.8	-171.6	5.3	Yes
21	Arg255-Phe258	RHFF	IV	86.7	-72.5	-64.5	-108.0	9.9	-47.8	5.3	Yes
22	Val265-Gln268	VAPQ	I	-51.3	-35.9	-	-64.2	-0.3	34.6	5.4	No
23	Leu270-Phe273	LFEF	IV	-41.2	125.3	163.8	-106.6	115.6	-70.5	6.8	Yes
24	Gly279-Gly282	GFMG	IV	-110.6	5.5	-68.6	-103.6	0.8	-63.4	6.8	Yes
25	Met281-Glu284	MGDE	I	-79.1	-11.3	-	-119.8	-7.4	70.6	5.1	Yes
26	Thr288-Glu291	TLGE	II	-52.4	131.4	-60.6	83.5	-1.3	-	5.8	No
27	Leu299-Lys302	LEPK	I	-59.1	-41.7	-175.4	-88.8	7.6	31.1	4.8	Yes
28	Glu300-Arg303	EPKR	IV	-88.8	7.6	31.1	-157.0	66.3	48.4	5.8	Yes

Fig. S5 (h): Table of beta turns. Note: A beta turn is defined for 4 consecutive residues (denoted by i, i+1, i+2 and i+3), if the distance between the Calpha atom of residue i and the Calpha atom of residue i+3 is less than 7Å and if the central two residues are not helical. The residue numbers of residues i and i+3 in the turn, the one-letter amino acid code of residues i, i+1, i+2 and i+3 in the turn, and the turn type. For each of the central two residues (i+1 and i+2) phi, psi and chi1 are recorded. The final colums show the distance between the Calpha atoms of residues i and i+3 and whether a hydrogen bond exists between these two residues. The phi and psi angles can vary by +/- 30 degrees from these ideal values with the added flexibility of one angle being allowed to deviate by as much as 40 degrees. Types VIa1, VIa2 and VIb turns are subject to the additional condition that residue i must be a cis-proline. Turns which do not fit any of the above criteria are classified as type IV.

Supplementary material 5

No.	Start	End	Sequence*	Turn type	Residue i+1 Phi	Psi	Chi1	i to i+2 CA-dist
1.	Trp144	Asn146	W Q N	INVERSE	-90.0	92.6	-168.3	6.0
2.	Leu247	Ser249	L P S	INVERSE	-84.6	41.9	24.9	5.7
3.	Arg255	Phe257	R H F	CLASSIC	86.7	-72.5	-64.5	5.8

Fig. S5 (i): Table of gamma turns. Note: The start and end residues of the gamma turn (residues i and i+2), the amino acid sequence of the residues in the turn and the turn type. Phi, psi and chi1 dihedral angles are given for the central residue (i+1). The final column gives the distance between the Calpha atoms of residues i and i+2.

Supplementary material 6

S/N	Atom no.	Atom no	Res name	Res no.	Chain		Atom no.	Atom no.	Res name	Res no.	Chain	distance
						Hydrogen bonds						
1	2534	NE2	HIS	9	B	< - >	4938	FE	FE	3	-	2.08
2	4241	NE2	HIS	224	B	< - >	4938	FE	FE	3	-	1.73
						Non-bonded contacts						
1	2532	CD2	HIS	9	B	:-:	4938	FE	FE	3	_	3.24
2	2533	CE1	HIS	9	B	:-:	4938	FE	FE	3	_	2.80
3	2534	NE2	HIS	9	B	:-:	4938	FE	FE	3	_	2.08
4	2547	OG1	THR	11	B	:-:	4938	FE	FE	3	_	3.51
5	4237	CG	HIS	224	B	:-:	4938	FE	FE	3	_	3.88
6	4238	ND1	HIS	224	B	:-:	4938	FE	FE	3	_	3.71
7	4239	CD2	HIS	224	B	:-:	4938	FE	FE	3	_	2.92
8	4240	CE1	HIS	224	B	:-:	4938	FE	FE	3	_	2.53
9	4241	NE2	HIS	224	B	:-:	4938	FE	FE	3	_	1.73

Fig. S6 (a): List of protein-metal interactions. The interactions shown are those mediated by hydrogen bonds and by hydrophobic contacts. Hydrogen bonds are indicated by dashed lines between the atoms involved, while hydrophobic contacts are represented by an arc with spokes radiating towards the ligand atoms they contact. The contacted atoms are shown with spokes radiating back.

Supplementary material 6

	- - - - - -ATOM 1- - - - - - -						- - - - - -ATOM 2- - - - - - -					
S/N	Atom No	Atom name	Res name	Res No.	Chain		Atom No	Atom name	Res name	Res No.	Chain	Distance
1	1720	NE2	GLN	217	A	<-->	4580	O	ALA	266	B	2.77
2	1763	NE2	HIS	223	A	<-->	4417	O	GLY	246	B	2.82
3	1949	O	GLY	246	A	<-->	4231	NE2	HIS	223	B	2.95
4	1970	OG	SER	249	A	<-->	4545	O	TYR	262	B	2.81
5	1976	OG	SER	250	A	<-->	4429	O	PRO	248	B	2.64
6	2112	O	ALA	266	A	<-->	4188	NE2	GLN	217	B	2.77

Fig. S6 (b): Lists of hydrogen bond between atom-atom interactions across protein-protein.

Supplementary material 6

	- - - - - -ATOM 1- - - - - -						- - - - - -ATOM 2- - - - - -					
S/N	Atom No	Atom name	Res name	Res No.	Chain		Atom No	Atom name	Res name	Res No.	Chain	Distance
1	1316	CG	ASN	167	A	<-->	4179	NE2	GLN	216	B	3.81
2	1317	OD1	ASN	167	A	<-->	4179	NE2	GLN	216	B	3.71
3	1318	ND2	ASN	167	A	<-->	4179	NE2	GLN	216	B	3.81
4	1318	ND2	ASN	167	A	<-->	4183	O	GLN	217	B	3.87
5	1318	ND2	ASN	167	A	<-->	4184	CB	GLN	217	B	3.44
6	1318	ND2	ASN	167	A	<-->	4185	CG	GLN	217	B	3.57
7	1318	ND2	ASN	167	A	<-->	4186	CD	GLN	217	B	3.45
8	1318	ND2	ASN	167	A	<-->	4188	NE2	GLN	217	B	3.15
9	1318	ND2	ASN	167	A	<-->	4202	CE2	PHE	219	B	3.78
10	1340	CB	PHE	170	A	<-->	4202	CE2	PHE	219	B	3.89
11	1341	CG	PHE	217	A	<-->	4202	CE2	PHE	219	B	3.72
12	1711	NE2	GLN	216	A	<-->	3784	CG	ASN	167	B	3.86
13	1711	NE2	GLN	216	A	<-->	3785	OD1	ASN	167	B	3.71
14	1711	NE2	GLN	216	A	<-->	3786	ND2	ASN	167	B	3.81
15	1715	O	GLN	217	A	<-->	3786	ND2	ASN	167	B	3.87
16	1716	CB	GLN	217	A	<-->	3786	ND2	ASN	167	B	3.44
17	1717	CG	GLN	217	A	<-->	3786	ND2	ASN	167	B	3.57
18	1718	CD	GLN	217	A	<-->	3786	ND2	ASN	167	B	3.45
19	1718	CD	GLN	217	A	<-->	4580	O	ALA	266	B	3.61
20	1719	OE1	GLN	217	A	<-->	4580	O	ALA	266	B	3.63
21	1720	NE2	GLN	217	A	<-->	3786	ND2	ASN	167	B	3.15
22	1720	NE2	GLN	217	A	<-->	4579	C	ALA	266	B	3.74
23	1720	NE2	GLN	217	A	<-->	4580	O	ALA	266	B	2.77
24	1734	CE2	PHE	219	A	<-->	3786	ND2	ASN	167	B	3.77
25	1734	CE2	PHE	219	A	<-->	3808	CB	PHE	170	B	3.89

Fig. S6 (c): Lists of non-bonded atom-atom interactions across protein-protein interface

	- - - - - -ATOM 1- - - - - - -						- - - - - -ATOM 2- - - - - - -					
S/N	Atom No	Atom name	Res name	Res No.	Chain		Atom No	Atom name	Res name	Res No.	Chain	Distance
26	1734	CE2	PHE	219	A	<-->	3809	CG	PHE	170	B	3.72
27	1735	CZ	PHE	219	A	<-->	4581	CB	ALA	266	B	3.62
28	1762	CE1	HIS	223	A	<-->	4417	O	GLY	246	B	3.66
29	1763	NE2	HIS	223	A	<-->	4415	CA	GLY	246	B	3.60
30	1763	NE2	HIS	223	A	<-->	4416	C	GLY	246	B	3.59
31	1763	NE2	HIS	223	A	<-->	4417	O	GLY	223	B	2.82
32	1947	NE2	GLY	246	A	<-->	4231	NE2	HIS	223	B	3.48
33	1948	C	GLY	246	A	<-->	4231	NE2	HIS	223	B	3.54
34	1949	O	GLY	246	A	<-->	4229	CD2	HIS	223	B	3.89
35	1949	O	GLY	246	A	<-->	4231	NE2	HIS	223	B	2.95
36	1957	CD2	LEU	247	A	<-->	4569	NE2	ARG	264	B	3.77
37	1960	C	PRO	248	A	<-->	4546	CB	TYR	262	B	3.89
38	1961	O	PRO	248	A	<-->	4443	CB	SER	250	B	3.54
39	1961	O	PRO	248	A	<-->	4444	OG	SER	250	B	3.74
40	1961	O	PRO	248	A	<-->	4546	CB	TYR	262	B	3.16
41	1961	O	PRO	248	A	<-->	4547	CG	TYR	262	B	3.41
42	1961	O	PRO	248	A	<-->	4548	CD1	TYR	262	B	3.65
43	1962	CB	PRO	248	A	<-->	4547	CG	TYR	262	B	3.90
44	1962	CB	PRO	248	A	<-->	4549	CD2	TYR	262	B	3.72
45	1963	CG	PRO	248	A	<-->	4230	CE1	HIS	223	B	3.78
46	1963	CG	PRO	248	A	<-->	4549	CD2	TYR	262	B	3.35
47	1964	CD	PRO	248	A	<-->	4612	CD1	LEU	270	B	3.90
48	1966	CA	SER	249	A	<-->	4443	CB	SER	250	B	3.58
49	1969	CB	SER	249	A	<-->	4545	O	TYR	262	B	3.57
50	1970	OG	SER	249	A	<-->	4429	O	PRO	248	B	3.79

Fig. S6 (c): Lists of non-bonded atom-atom interactions across protein-protein interface

	- - - - - -ATOM 1- - - - - - -						- - - - - -ATOM 2- - - - - - -					
S/N	Atom No	Atom name	Res name	Res No.	Chain		Atom No	Atom name	Res name	Res No.	Chain	Distance
51	1970	OG	SER	249	A	<-->	4434	CA	SER	249	B	3.88
52	1970	OG	SER	249	A	<-->	4435	C	SER	249	B	3.73
53	1970	OG	SER	249	A	<-->	4439	N	SER	250	B	3.19
54	1970	OG	SER	249	A	<-->	4440	CA	SER	250	B	3.73
55	1970	OG	SER	249	A	<-->	4443	CB	SER	250	B	3.58
56	1970	OG	SER	249	A	<-->	4544	C	TYR	262	B	3.65
57	1970	OG	SER	249	A	<-->	4545	O	TYR	262	B	2.81
58	1971	N	SER	250	A	<-->	4440	CA	SER	250	B	3.40
59	1971	N	SER	250	A	<-->	4443	CB	SER	250	B	3.36
60	1972	CA	SER	250	A	<-->	4440	CA	SER	250	B	3.82
61	1973	C	SER	250	A	<-->	4440	CA	SER	250	B	3.80
62	1974	O	SER	250	A	<-->	4440	N	GLY	251	B	3.75
63	1975	CB	SER	250	A	<-->	4445	O	PRO	248	B	3.73
64	1975	CB	SER	250	A	<-->	4429	C	SER	249	B	3.69
65	1975	CB	SER	250	A	<-->	4435	O	SER	249	B	3.26
66	1975	CB	SER	250	A	<-->	4436	N	SER	250	B	3.87
67	1975	CB	SER	250	A	<-->	4439	CA	SER	250	B	3.71
68	1976	OG	SER	250	A	<-->	4440	C	PRO	248	B	3.56
69	1976	OG	SER	250	A	<-->	4428	O	PRO	248	B	2.64
70	1976	OG	SER	250	A	<-->	4429	C	SER	249	B	3.55
71	1976	OG	SER	250	A	<-->	4435	O	SER	249	B	3.47
72	1976	OG	SER	250	A	<-->	4436	N	SER	250	B	3.81
73	2078	CB	TYR	262	A	<-->	4439	CB	PRO	248	B	3.79
74	2078	CB	TYR	262	A	<-->	4430	CG	PRO	248	B	3.19
75	2079	CG	TYR	262	A	<-->	4431	CG	PRO	248	B	3.58

Fig. S6 (c): Lists of non-bonded atom-atom interactions across protein-protein interface

	- - - - - -ATOM 1- - - - - - -						- - - - - -ATOM 2- - - - - - -					
S/N	Atom No	Atom name	Res name	Res No.	Chain		Atom No	Atom name	Res name	Res No.	Chain	Distance
76	2081	CD2	TYR	262	A	<-->	4431	CG	PRO	248	B	3.70
77	2095	CB	ARG	264	A	<-->	4431	CB	ARG	264	B	3.72
78	2096	CG	ARG	264	A	<-->	4431	CG	ARG	264	B	3.62
79	2105	O	VAL	265	A	<-->	4596	OE1	GLN	268	B	3.90
80	2111	C	ALA	266	A	<-->	4188	NE2	GLN	217	B	3.74
81	2112	O	ALA	266	A	<-->	4186	CD	GLN	217	B	3.61
82	2112	O	ALA	266	A	<-->	4187	OE1	GLN	217	B	3.63
83	2112	O	ALA	266	A	<-->	4188	NE2	GLN	217	B	2.77
84	2112	O	ALA	266	A	<-->	4596	OE1	GLN	268	B	3.72
85	2113	CB	ALA	266	A	<-->	4203	CZ	PHE	219	B	3.62
86	2121	N	GLN	268	A	<-->	4596	OE1	GLN	268	B	3.88
87	2122	CA	GLN	268	A	<-->	4596	OE1	GLN	268	B	3.66
88	2125	CB	GLN	268	A	<-->	4596	OE1	GLN	268	B	3.48
89	2128	OE1	GLN	268	A	<-->	4580	O	ALA	266	B	3.73
90	2128	OE1	GLN	268	A	<-->	4589	N	GLN	268	B	3.88
91	2128	OE1	GLN	268	A	<-->	4590	CA	GLN	268	B	3.66
92	2128	OE1	GLN	268	A	<-->	4593	CB	GLN	268	B	3.48
93	2128	OE1	GLN	268	A	<-->	4596	OE1	GLN	268	B	3.21
94	2144	CD1	LEU	270	A	<-->	4417	O	GLY	246	B	3.79

Fig. S6 (c): Lists of non-bonded atom-atom interactions across protein-protein interface

Supplementary material 7

	Volume	R1 ratio	Accessible vertices		Buried vertices		Average depth		Residue..type							Ligands
1	10475.58	4.23	72.31	3	16.25	1	20.38	1	29	23	25	41	17	18	0	
2	2475.14	0.00	63.49	9	9.21	7	9.95	3	10	10	8	8	6	4	0	
3	2465.02	0.00	63.43	10	9.13	8	9.78	4	10	10	8	8	6	4	0	
4	1815.75	0.00	64.58	8	8.00	10	8.55	8	4	7	8	8	6	4	0	
5	1814.48	0.00	64.59	7	8.03	9	8.64	7	4	8	8	9	6	4	0	
6	1400.20	0.00	68.76	4	14.33	2	13.22	2	2	0	15	8	2	2	0	
7	1105.73	0.00	72.59	2	14.02	4	8.69	5	3	3	6	8	0	6	0	
8	1105.73	0.00	73.01	1	14.16	3	8.66	6	3	3	6	8	0	6	0	
9	1177.45	0.00	67.49	6	9.53	6	6.80	10	4	6	6	4	5	6	0	
10	1173.66	0.00	67.62	5	9.67	5	7.56	9	4	6	5	4	5	6	0	

Protein structure

Residue-type_colouring						
Positive	Negative	Neutral	Aliphatic	Aromatic	Pro & Gly	Cysteine
H,K,R	D,E	S,T,N,Q	A,V,L,I,M	F,Y,W	P,G	C

Fig. S7 a: Table of Clefts found in CpsA

Supplementary material 7

Pores

	Radius	Free R	Length	HPathy	HPhob	Polar	Rel Mut	Residue.type							Ligands
1	1.71	3.00	93.5	-0.43	0.04	18.5	78	9	6	3	6	6	3	0	
2	1.65	2.67	136.0	-0.59	-0.06	18.3	79	10	11	4	8	7	3	0	
3	1.16	1.41	31.5	-0.80	-0.61	10.8	81	1	2	1	3	1	0	0	
4	1.16	1.41	31.5	-0.80	-0.61	10.8	81	1	2	1	3	1	0	0	

Residue-type_colouring

Positive	Negative	Neutral	Aliphatic	Aromatic	Pro & Gly	Cysteine
H,K,R	D,E	S,T,N,Q	A,V,L,I,M	F,Y,W	P,G	C

Fig. S7 b: Table of Pores found in CpsA. Note: Free R: Free radius; Hpathy: hydropathy; HPhob: hydrophobicity; Rel Mut: Relative mutability

Supplementary material 7

Tunnels

	Radius	Free R	Length	HPathy	HPhob	Polar	Rel Mut	Residue..type							Ligands
1	1.52	1.83	17.2	-1.12	-0.67	2.6	96	0	0	5	3	0	0	0	
2	1.52	1.83	17.2	-1.12	-0.67	2.6	96	0	0	5	3	0	0	0	
3	1.54	1.92	21.4	-0.19	-0.49	2.3	97	0	0	4	5	0	0	0	
4	1.55	1.55	21.4	-0.23	-0.49	2.3	97	0	0	4	5	0	0	0	
5	1.17	1.40	27.6	1.09	0.09	2.0	85	0	0	1	8	1	0	0	
6	1.36	1.58	27.7	-0.07	-0.28	9.9	86	1	2	0	5	1	0	0	
7	1.17	1.40	27.6	1.09	0.09	2.0	85	0	0	1	8	1	0	0	
8	1.36	1.58	27.7	-0.07	-0.28	9.9	86	1	2	0	5	1	0	0	

Residue-type colouring

Positive	Negative	Neutral	Aliphatic	Aromatic	Pro & Gly	Cysteine
H,K,R	D,E	S,T,N,Q	A,V,L,I,M	F,Y,W	P,G	C

Fig. S7 c: Tunnels found in CpsA. Free R: Free radius; Hpathy: hydropathy; HPhob: hydrophobicity; Rel Mut: Relative mutability

Supplementary material 8 (Fig. S8)

```
     ....*....|....*....|....*....|....*....|....*....|....*....|
CpsA_A   1 NQLKGIHHVTAITS--SAEKNYEFFTHVLG----MRLVKKTVNQ----------D-D--I 41
1ZSW_A  26 YEIKGHHHISMVTK--NANENNHFYKNVLG----LRRVKMTVn----------qD-D--P 66
4HUZ_A   4 NHITSLHHITICTG--TAQGDIDFFVKVMG----QRFVKRTLFY----------DgS--I 45
           ....*....|....*....|....*....|....*....|....*....|....*....|
CpsA_A  42 QTYHLFFADDKG---------------SAGTDMTFF-DF--PGV-P---KGVH----GTN 75
1ZSW_A  67 SMYHLFYGDKTG---------------SPGTELSFF-EI--PLV-G---RTYR----GTN 100
4HUZ_A  46 PIYHLYFADELG---------------TPGTVMTTF-PT--RRTgQ---KGRK----GSN 80
           ....*....|....*....|....*....|....*....|....*....|....*....|
CpsA_A  76 EISKTSFRVPTDAALSYWVKRFDRLEV-EH-TG----IKEQ--F-----GKKTLSFVD-F 121
1ZSW_A 101 AITRIGLLVPSEDSLHYWKERFEKFDV-KH-SE----MTTY--A-----NRPALQFED-A 146
4HUZ_A  81 QFTVCTYAIPK-GSLEWWIGHLNAHGI-AT-GE----PGTR--F-----GQRYVGFQH-P 125
           ....*....|....*....|....*....|....*....|....*....|....*....|
CpsA_A 122 -DDQHYQLISDEVDKGVESGTP--WQNGPVPLGYAITGLGPVFIRIAN--FSFFKEVLEK 176
1ZSW_A 147 -EGLRLVLLVSN-GEKVEHWET--WEKSEVPAKHQIQGMGSVELTVRR--LDKMASTLTE 200
4HUZ_A 126 dCGIDFEVLEDENDT----RQP--YDSPYVPIEHAQRGFHSWTASVRE--LEDMDFFMEN 177
           ....*....|....*....|....*....|....*....|....*....|....*....|
CpsA_A 177 VLLFKEI--AQE------G-----------------EFYLFEVNEGGNG-ASVIVEHNTV 210
1ZSW_A 201 IFGYTEV--SRN------D-----------------QEAIFQSIKGEAF-GEIVVKYLD- 233
4HUZ_A 178 CWNFEKI--GEE------G-----------------NRHRYRVKGTTESgTIIDLLHEPD 212
           ....*....|....*....|....*....|....*....|....*....|....*....|
CpsA_A 211 ----LPE-AQQG-------------F-G----TVHHAAFRVEDRAV-LEEWIERIGSV-G 245
1ZSW_A 234 --------GPTE-------------K-Pgr-gSIHHLAIRVKNDAE-LAYWEEQVKQR-G 268
4HUZ_A 213 rrqgsw----------------------tiaegIIHHGAFAVPDMDI-QARIKFETEGV-G 249
           ....*....|....*....|....*....|....*....|....*....|....*....|
CpsA_A 246 L---------------------------PS--------S--G--Y----VN--R---H-- 255
1ZSW_A 269 F---------------------------HS--------S--G--I----ID--R---F-- 278
4HUZ_A 250 F---------------------------TDf-------S--D--R----KN--R---G-- 260
           ....*....|....*....|....*....|....*....|....*....|....*....|
CpsA_A 256 F---FE-SL-Y-ARVA--PQILFEFATDGPGFMGDEP----------------------- 284
1ZSW_A 279 Y---FK-SL-Y-FRES--NGILFEIATDGPGFTVDGD----------------------- 307
4HUZ_A 261 Y---FE-ST-Y-VRTP--GGVM-FEATHSLGFTHDED----------------------- 288
           ....*....|....*....|....*....|....*....|....*....|....*....|
CpsA_A 285 ---------------------------------------------YETLGEKLSL-PPFLEP 300
1ZSW_A 308 ---------------------------------------------VEHLGEKLDLpPFLEDQ 324
4HUZ_A 289 ---------------------------------------------ERSLGMDLKV-SPQFDD 304
           ....*....
CpsA_A 301 KREEIEKLV 309
1ZSW_A 325 RAEIEANLA 333
4HUZ_A 305 KKHLIEQAM 313
```

Fig. S8. A pairwise 3D domains structural alignments of CpsA with 4HUZ and 1ZSW. Three-state secondary structure definitions by DSSP (reduced to H=helix, E=sheet, L=coil) are shown above the amino acid sequence. Structurally equivalent residues are in uppercase, structurally non-equivalent residues (e.g. in loops) are in lowercase. Amino acid identities are marked by vertical bars. The points were addition, deletion and duplication have occurred in 4HUZ and CpsA were highlighted yellow shadings.

CHAPTER NINE

This chapter ***has been submited to the Journal*****:** ***Functional and Integrated Genomics***

Whole-genome sequencing, genome mining, metabolic reconstruction and evolution of pentachlorophenol and other xenobiotics degradation pathways in *Bacillus tropicus* strain AOA-CPS1

Oladipupo A. Aregbesola, Ajit Kumar, Mduduzi P. Mokoena, Ademola O. Olaniran*

Discipline of Microbiology, School of Life Sciences, College of Agriculture, Engineering and Science, University of KwaZulu-Natal (Westville Campus), Private Bag X54001, Durban 4000, Republic of South Africa.

***Corresponding author:**
Ademola O. Olaniran
Phone: +27 31 260 7400
Fax: +27 31 260 7809
E-mail: olanirana@ukzn.ac.za

Running title: Genome annotation of *Bacillus tropicus* strain AOA-CPS1.

Abstract

A pentachlorophenol degrading bacterium was isolated from effluent of a wastewater treatment plant in Durban, South Africa and identified as *Bacillus tropicus* strain AOA-CPS1 (*Bt*AOA). The isolate degraded 29% of pentachlorophenol (PCP) within 9 days at initial PCP concentration of 100 mg L^{-1} and 62% of PCP when the initial concentration was set at 350 mg L^{-1}. The whole-genome of *Bt*AOA was sequenced using Pacific Biosciences RS II sequencer with the Single Molecule, Real-Time (SMRT) Link (version 7.0.1.66975) and analysed using the HGAP4-*de-novo* assembly application. The contigs were annotated at NCBI, RASTtk and PROKKA prokaryotic genome annotation pipelines. *Bt*AOA genome comprises of 5,246,860 bp chromosome and a 58,449 bp plasmid with GC content of 35.4%. The metabolic reconstruction for *Bt*AOA showed that the organism has been naturally exposed to various chlorophenolic compounds including PCP and other xenobiotics. The chromosome encodes genes for core processes, stress response and PCP catabolic genes. Analogues of PCP catabolic gene (cpsBDCAE, and p450) sequences were identified from the NCBI annotation data, PCR-amplified from the whole genome of *Bt*AOA, cloned into pET15b expression vector, overexpressed in *E. coli* BL21 (DE3) expression host, purified and characterized. Sequence mining and comparative analysis of the metabolic reconstruction of *Bt*AOA genome with closely related strains suggests that the operon encoding the first two enzymes in the PCP degradation pathway were acquired from a pre-existing pterin-carbinolamine dehydratase subsystem. The other two enzymes were recruited via horizontal gene transfer (HGT) from the pool of hypothetical proteins with no previous specific function, while the last enzyme was recruited from pre-existing enzymes from the TCA or serine-glyoxalase cycle via HGT events. This study provides a comprehensive understanding of the role of *Bt*AOA in PCP degradation and its potential exploitation for bioremediation of other xenobiotic compounds.

Keywords: *Bacillus tropicus;* Whole-genome sequencing; phenylalanine-4-monooxygenase; pterin-carbinolamine dehydratase; ring-cleaving dioxygenase.

10.0 Introduction:

The first agricultural revolution witnessed massive use of noxious agrochemicals such as lindane, atrazine, 1,1,1-trichloro-2,2-bis(4-chlorophenyl)ethane (DDT), polychlorinated biphenyls (PCBs), pentachlorophenol (PCP) and other multi-halogenated organic compounds (to control pests and diseases causing microorganisms) with little or no plan for the long-term consequences in the environment and human health. Most of these chemicals are not environmentally friendly and have been listed as priority pollutants (ATSDR 2017) because of their toxicological profiles (Igbinosa et al. 2013).

Pentachlorophenol was specifically used as a wood preservative to combat wood pests and fungi (IARC 2019a). Not only has PCP been listed as a priority pollutant, it has also been recognized as a human carcinogen (IARC 2019a, b; Stockholm convention 2019). In spite of the health hazards, PCP is still been used by certified industries (Kim et al. 2019) to protect products (such as utility poles) that would in one way or the other come in contact with humans. For instance, PCP, polychlorinated dibenzo-p-dioxins and polychlorinated dibenzofurans were recently found in surface soil surrounding PCP-treated utility poles on the Kenai National Wildlife Refuge, Alaska USA and Montreal, Quebec in Canada (Verbrugge et al. 2018).

The presence of these anthropogenic agrochemicals in a natural environment often triggers adaptive and selective pressures geared towards evolution of *de novo* genes, enzymes and pathways that allow microorganisms to circumvent the toxic effects of these xenobiotics and possibly use them as carbon, nitrogen, or phosphorus sources and/or degrade them to (none- or less-toxic) metabolites that could be channelled into the central metabolic pathways. However, biodegradation of some of these persistent organic compounds is sometimes not efficient because they are recalcitrant, and microorganisms have not yet evolved metabolically efficient enzymes that can efficiently catalyse the steps required to convert the compounds into intermediates of central carbon metabolism (Copley 2010).

Although PCP is resistant to microbial degradation, some microorganisms capable of transforming PCP has been reported (Lopez-Echartea et al. 2016; El-naas et al. 2017; Niesler and Surmacz-Górska 2018; Xu et al. 2019), including the recently isolated *Bacillus tropicus* strain

AOA-CPS1 (*Bt*AOA). *Sphingobium chlorophenolicum* has been extensively studied and found to have assembled a poorly functional pathway for PCP degradation (Copley et al. 2012), with low degradation efficiency and inability to grow at high concentrations of PCP (Dai and Copley 2004). In this study, *Bt*AOA was recently isolated from wastewater and characterized for its ability to degrade PCP and some of its congeners.

The catabolic genes (cpsBDCAE), encoding PCP degradation enzymes (CpsBDCAE), in *Bt*AOA were amplified, cloned, expressed heterologously, purified to homogeneity and most of them have been characterized. To exploit the full potential of *Bt*AOA xenobiotics' degradation and its possible application in the bioremediation of polluted soil and wastewater, its whole genome was sequenced, annotated, and analysed. Comparative analysis of the sequence to that of the closely related strains was carried out to ascertain the evolutionary route of the PCP degrading enzymes. Metabolic reconstruction of *Bt*AOA was also compared to the closest strains to determine the distribution of PCP degrading enzymes among *Bacillus* spp.

10.1 Materials and methods

10.1.1 Materials

Pentachlorophenol (PCP) (98%), Isopropyl β-D-1-thiogalactopyranoside (IPTG, ≥99.0%), Luria Bertani (LB) agar and Broth (Vegitone), Sodium Dodecyl Sulphate (SDS), phenylmethylsulfonyl fluoride (PMSF ≥98.5%), ampicillin (Amp) sodium salt, 2-Mercaptoethanol (2-ME, ≥99.0%) were purchased from Merck (Merck & Company, Inc., NJ, USA). Dithiothreitol (DTT, ≥99.5%), N,N,N′,N′-Tetramethyl ethylenediamine (TEMED, ≥99.0%), Acrylamides/bis-acrylamide (40.0%) solution, ethylenediaminetetraacetic acid (EDTA, ACS 99.4%), FastDigest restriction endonucleases (*XhoI*, *BamHI*, *NotI* and *NdeI*), T4 DNA ligase, DNA ladders (100 bp plus, 1 kb and 1 kb plus), page ruler plus pre-stained protein ladder (10–250kDa), *Taq* polymerase and PCR reaction mix were obtained from ThermoFisher Scientific (Waltham, MA, USA). *Escherichia coli* strains DH5α and BL21 (DE3) (Invitrogen, ThermoFisher Scientific, Waltham, USA) were used as cloning and protein overexpression hosts, respectively.

Plasmid vector: pET15b DNA (Novagen, Merck, Germany), was used as a cloning and expression vector. Chemically competent cells of *E.coli* strains DH5α and BL21(DE3) were prepared as previously described (Chang et al. 2017). All other chemicals and reagents used in this study were of analytical grade standards unless otherwise stated. Primers synthesis and DNA sequencing were done by Inqaba Biotech, South Africa. Biomass was produced in an Erlenmeyer flask at 30°C, with shaking at 200 rpm in a shaking incubator (Innova 44 series, New Brunswick Scientific, UK). All centrifugation was done with either Eppendorf centrifuge 5415D (Hamburg, Germany) or Avanti J-26 XPI centrifuge (Beckman Coulter, USA).

10.1.2 Isolation and preliminary identification of *Bacillus tropicus* strain AOA-CPS1 (*Bt*AOA)

*Bt*AOA was isolated from wastewater effluent, via culture enrichment and purified through successive sub-culturing on sterile nutrient agar plates until distinct colonies were obtained. Briefly, after three successive sub-culturing, 0.1 ml of the enriched culture was spread inoculated on minimal salt agar plates supplemented with 50 mg L^{-1} of PCP. The plates were incubated at 30 °C until visible growths were observed. The isolate was purified via successive sub-culturing on sterile nutrient agar (NA) plates until distinct colonies were obtained. The 16S rDNA fragment was amplified from the genomic DNA of the isolate via PCR using 63F and 1387R universal primers pair (Marchesi et al. 1998), sequenced and submitted at NCBI BLASTn server (Camacho et al. 2009) for the early identification of the pure culture.

10.1.3 PCP biotransformation study

Biotransformation of PCP by *Bt*AOA was performed using a low buffered MSM (Saber and Crawford 1985), with some modifications. The MSM contained (g L^{-1}): K_2HPO_4 (0.065); KH_2PO_4 0.019; $MgSO_4.7H_2O$ (0.1); $NaNO_3$ (0.5); 0.02 M $FeSO_4$ (2 ml); pH 7.0, and 2 ml of micronutrients. The micronutrients contained (mg L^{-1}): $ZnSO_4{\cdot}7H_2O$ (4.0); $MnSO_4{\cdot}4H_2O$ (0.2); H_3BO_3 (0.15) and EDTA (2.5) (Ammeri et al. 2017). The reconstituted MSM was supplemented with 350 mg L^{-1} PCP for biotransformation study (PMSM). The isolate was grown in nutrient broth overnight, harvested by centrifugation (8000 rpm for 10 min), washed twice with MSM, resuspended in the same medium and standardized (OD_{600nm} of 1.0). About 90 ml of PMSM media were inoculated with 10 ml of standardized inoculum and incubated for 9 days. Proper

positive and negative controls were used along with the experiment to check for growth and abiogenic loss of PCP during biotransformation processes (Patel and Kumar 2016a).

Cell growth and PCP biotransformation were monitored spectrophotometrically daily. The metabolites during the PCP degradation were identified by GC-MS analysis. The isolate was grown in MSM supplemented with 50 mg L^{-1} of PCP for induction and incubated for 48 h. Metabolites were extracted and derivatized as previously described (Li et al. 2001; Smith, R. M. 2003). Both derivatized and underivatized samples were analysed in the Agilent 7890A GC System (Agilent Technologies, USA), equipped with a 5975C MS detector. The system was run at 80 °C, raised at 5 °C min^{-1} to 160 °C for 3 min, then raised at 10 °C min^{-1} to 260 °C and held at 260 ^{o}C for another 3 min. Mass ranges (m/z) were set at 50-700, ionization energy was set at 70 eV and injection volume of 1 µl. The peaks of the PCP and its metabolites were compared to the National Institute of Standard and Technology (NIST) library database (Sharma et al. 2009).

10.1.4 Isolation of total genomic DNA from *Bt*AOA and Whole-genome sequencing

The genomic DNA of the isolate was extracted using a Quick-DNA^{TM} fungal/bacterial miniprep kit (Zymo Research Corporation, CA, USA) from 5 ml of Luria-Bertani (LB) inoculated with a pure single colony. The draft genome data was generated (Inqaba Biotech, Pretoria, South Africa) using a combination of Sequel II System (Biosciences 2019), PacBio Single-Molecule Real-Time (SMRT Link Version 7.0.1.66975) sequencing technology (Moine-Scientist & Applications Support, 2019), FALCON assembler (Kingan et al. 2018) and Hierarchical Genome Assembly Process 4 (HGAP4) *de novo* assembly analysis application. The sequencing procedure including the construction of SMRTbell libraries are elaborated in the PacBio infographics at (https://www.pacb.com/wp-content/uploads/Infographic_SMRT-Sequencing-How-it-Works.pdf).
The dataset generated were trimmed, assembled, and analysed via HGAP4 *de novo* assembly analysis. The parameters set up for the HGAP4 analysis is shown in supplementary material (Table S1).

The resulting contigs from the *de novo* assembly of the draft genome were polished and submitted to the National Centre for Biotechnological Information (NCBI) GenBank as whole-genome shotgun (WGS) under the BioSample and BioProject accession numbers

SAMN13384248 and PRJNA591551, respectively. Annotations were added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP), using the best-placed reference protein set and GeneMarkS-2+ (Tatusova et al. 2016; Haft et al. 2018). Pieces of information about the annotation pipeline can be found at https://www.ncbi.nlm.nih.gov/genome/annotation_prok/. Furthermore, gene prediction and annotation of *Bt*AOA whole genome were obtained using the Rapid Annotations via the Subsystems Technology (RASTtk pipeline v.2.0) server (Brettin et al., 2015). Additionally, genes including tRNA and rRNAs were annotated at PROKKA annotation server using Prodigal v.2.6 (Hyatt et al. 2010).

10.1.5 Cloning, overexpression, purification, and characterization of catabolic enzymes

Analogues of PCP catabolic gene sequences for cpsBDCAE and p450 were retrieved from NCBI annotation data and amplified from the whole genome of *Bt*AOA using manually designed primers according to the gene sequences provided at NCBI database, targeting expression vector pET15b (Table 1). The amplification of genes was performed using High-fidelity DNA polymerase master mix using standard PCR conditions as described previously (Kumar et al. 2018). The cloning and overexpression procedures were performed by following standard molecular cloning protocols (Sambrook et al. 2012). The recombinant 6xHis-tagged proteins were purified by loading the cell lysate on Pierce HisPur Cobalt Chromatography Cartridge (#90094, Thermo Scientific, IL, USA) connected to ÄKTA purifier system (GE Healthcare Life Sciences, IL, USA) following the manual instructions. The enzymes properties i.e., optimum pH, optimum temperature, pH stability, temperature stability, K_m, v_{max} as well as the biophysical and chemical properties were evaluated as described previously (Kumar et al. 2018; Setlhare et al. 2020).

10.1.6 Cluster analysis and evolutionary relationships of taxonomic group

Multiple sequences alignments were constructed using Cluster Omega v.1.2.4 (Madeira et al. 2019). The evolutionary history of taxa was inferred using the Neighbour-Joining method (Zhang & Sun 2008), evolutionary analyses were conducted in MEGA7 (Kumar et al. 2016). The bootstrap consensus tree inferred from 1000 replicates, evolutionary history of the taxa analysed, and evolutionary distances computed using the Poisson correction method are in the units of the number of amino acid substitutions per site (Felsenstein 1985).

Table 1: Primers used for amplification of PCP-degradation genes in *Bt*AOA.

Genes	**NCBI GeneBank Accession no.**	**Primers sequences (5'-3')**	**T_m (ºC) used for PCR**	**Amplicon size (bp)**
cpsA	QIE35765.1	F-GGG***CATATG***ATGAACCAATTAAAAGGA	58	978
		R-ACG***GGATCC*** TTACTCTTTAATAAATTCCTT		
cpsB	QIE38732.1	F-AGA ***CTC GAG***ATGACAAAGAAAACAGAA ATT	58	1755
		R-AAA***GGATCC***TCAGTTAATCTTAGCATCATT		
cpsC	QIE36766.1	F-TTT***CTCGAG***ATGTCGCAATACATAAGGGAT	56	1236
		R-AGT***GGATCC***TCATTTATTTTCCCCCTTCTT		
cpsD	QIE38733.1	F-TAA***CATATG***ATGATGCTAAGATTAACTGAA	54	315
		R-GTA***GGATCC***TTATTTTCTTATAATTGC		
cpsE	QIE38957.1	F-TGT***CATATG***ACAATCAAACGCAAGAAAG	54	939
		R-AGA***GGATCC***TTAAACAAGAACTTTCATTAC		
P450	QIE37123.1	F-GAG**GGATCC**CATGTCAATGAAAAACAAAGT	58	1237
		R-TAT**GCGGCCGC**GAAAGTTAAAGGCAATTCC		

F-: forward, R-: reverse, underline, bold, italics: restriction sites

10.1.7 Nucleotide sequence accession numbers

The Whole Genome Shotgun project of *Bacillus tropicus* strain AOA-CPS1 has been deposited in GenBank under accession number CP049019 (version CP049019.1) for chromosome and CP049020 (version CP049019.1) for plasmid. The sequence and annotation data are publicly available at https://www.ncbi.nlm.nih.gov/nuccore/CP049019.1 (chromosomal) and https://www.ncbi.nlm.nih.gov/nuccore/CP049020.1 (plasmid).

10.2 Results and Discussion

10.2.1 Identification of isolate based on whole-genome sequencing

Primarily, the isolate was identified as *Bacillus cereus* strain AOA-CPS1 (*Bc*AOA) based on the 16S rDNA sequence analysis (submitted to NCBI as accession number MH504118.1). However, a quality control test by NCBI for the submitted whole genome sequence of the strain, using an average nucleotide identity (ANI), which compares the submitted genome sequence against the whole genomes of the type strains that are already in GenBank (Federhen et al. 2016; Ciufo et al. 2018), resulted in the renaming of *Bc*AOA as *Bacillus tropicus* strain AOA-CPS1 (*Bt*AOA). The ANI analysis indicated that the genome sequences of *Bc*AOA are 96.61% identical to the genome of the type strain of *Bacillus tropicus*, with 89.9% coverage of the genome. Consequently, *Bc*AOA was renamed as *Bt*AOA (based on the whole genome data submitted at NCBI under accession number CP049019). The annotation data for the plasmid sequence at NCBI (accession number: CP049020) did not show the presence of any gene involved in PCP degradation pathway and hence was not included in this study.

10.2.2 PCP biotransformation, degradation kinetics and metabolites detection

*Bt*AOA transformed 74% of 350 mg L^{-1} of PCP within 9 days in unoptimized and 98.2% of 500 mg L^{-1} of PCP in 9 days at optimized conditions respectively (data not shown). However, no significant reduction in PCP concentration was observed in PMSM without the inoculum after 9 days of incubation. Apart from PCP, *Bt*AOA also transformed 2,4,6-trichlorophenol (2,4,6-TCP), 2,4-dichlorophenol, 4-chlorophenol and 2- chlorophenol. The strain also co-metabolized different concentrations of PCP and 2,4,6-TCP mix. The specific PCP biotransformation rate increased with increase in PCP concentrations up to 250 mg L^{-1}, remains at equilibrium/saturated between 250 and 300 mg L^{-1}, and then increased at 350 mg L^{-1} after which the removal rates decreased. A reciprocal plot, between PCP removal rate ($1/R_s$), and substrate concentration ($1/S$), at low PCP conc ($\leq$ 250 mg L^{-1}), yielded a linear graph with a slope of K_s/R_m and an intercept of $1/R_m$ (Fig. S1). From a linear line of the best fit plot, the biodegradation kinetic parameters obtained were: R_m (0.996 mg L^{-1} h^{-1}); K_s (171.19 mg L^{-1}; K_{si} (723.75 mg L^{-1}); and R^2 (0.98), where R_m is the specific PCP transformation rate, K_s is the half-saturation or substrate utilization constant and K_{si} is the second order substrate inhibition constant.

*Bt*AOA transformed PCP and its congeners, singly and in co-metabolism. Bacterial capacity to mineralize mixtures of environmental pollutants have also been reported (Durruty et al. 2011; Patel and Kumar 2016a). Such strains have great potentials for bio-based environmental remediation. The observed increase in PCP transformation by the isolate with increase in PCP concentration agrees with previous reports (Khessairi et al. 2014; Lv et al. 2014). The maximum PCP transformation rates also increased with an increase in PCP concentration which is in consonance with other earlier reports indicating that biotransformation of PCP increases with increase in PCP concentration and that transformation is not growth-dependent but substrate concentration-dependent (Ammeri et al. 2017; El-Bialy et al. 2018).

Although the compound exerts inhibitory effect on the growth of the bacterium, synthesis of the enzymes required for the transformation of the compound was stimulated as previously observed (Yang et al. 2006). Half-saturation or substrate utilization constant (K_s) is an affinity coefficient of an organism to the substrate (Patel and Kumar, 2016). At high substrate concentrations, degradation kinetics are independent of substrate concentration but depends on maximum specific transformation rate. However, at low concentrations, the substrate becomes rate-limiting factor, which is mainly influenced by K_s (Arnaldos et al. 2015). A low K_s obtained showed that *Bt*AOA has a high affinity for the compound. Inhibition constant symbolizes the inhibitory effects of the compound on an organism. A higher K_{si} indicates the tolerance level of the organism to the compound (Patel and Kumar 2016b). The high K_{si} value obtained showed that *Bt*AOA has a high tolerance to the substrate, and therefore has high potential for the bioremediation of PCP at a short incubation time.

The GC-MS analysis of the PCP degradation products showed that PCP was degraded by the organism. The retention time(s), mass spectra and relative intensity of both derivatized (TMS) and underivatized products are shown in Table S2. Based on the retention time of metabolites in GC-MS analysis, two pathways (Fig. S2) were proposed for PCP transformation in *Bt*AOA. In the first pathway, 2,6-bis(1,1-dimethylethyl)phenol; 1-methoxy-5-trimetylsilyoxyhexane, trimethylsilyl 2-butoxyacetate, and Methyl 2-hydroxyl-3-methylbutanoate fragments were found.

The metabolites detected for the proposed second pathway include 1,3-dimethyl-4,6-diisopropylbenzene; 2,4-Dimethylbenzenecarboxaldehyde and 2,5-Dimethylbenzaldehyde. The

metabolites detected in the first pathway were similar to those earlier described in the *ortho* pathway for PCP degradation by *S. chlorophenolicum* ATCC 39723 (Cai and Xun 2002), while the second pathway is similar to PCP degradation pathway by *Rhodococcus chlorophenolicus* PCP-1 (Uotila et al. 1995) and *Desulfitobacterium hafniense* strain PCP-1 (Villemur 2013). Undoubtedly, both 2,6-Di-tert-butylbenzoquinone and 1,3-Dimethyl-4,6-diisopropylbenzene cannot be metabolites from the same pathway, it could be that one of the pathways is initiated by PcpB and the other by Cytochrome P450 monooxygenase. The presence of these two monooxygenases in the organism further stresses the potential biotechnological applications of this organism.

10.2.3 Cloning, purification, and characterization of catabolic enzymes

Analogues of PCP catabolic genes (*cpsABCDE*, and p450) were successfully amplified from the whole genome of *Bt*AOA using the primers designed based on the gene sequences retrieved from the whole genome annotation data submitted at NCBI. The approximate size of the amplified genes (bp) was: *cps*A (936), *cps*B (≅1775), *cps*C (≅1236), *cps*D (≅315) *cps*E (≅939) and p450 (≅1236) (Fig. 1A), while the size (kDa) of the protein expressed were found to be CpsA (≅33.0), CpsB (≅68.0), CpsC (≅38.0), CpsD (≅12.5) CpsE (≅31.3) and p450 (≅48.0) (Fig. 1B).

The genes shared >99% sequences homologies with the corresponding genes in the genomes of their progenitors but their biological functions were only putatively derived by automated gene prediction method to date. The isolate seems to have assembled a metabolically competent pathway for complete PCP degradation by recruiting inherent genes from its ancestors and from already existing pathways into its PCP degradation pathway. The enzymes properties i.e. optimum pH, optimum temperature, pH stability, temperature stability, K_m and v_{max} are shown in Table 2. The biophysical and chemical characteristics of each of the enzyme as predicted by the bioinformatics tools are presented in Table 3.

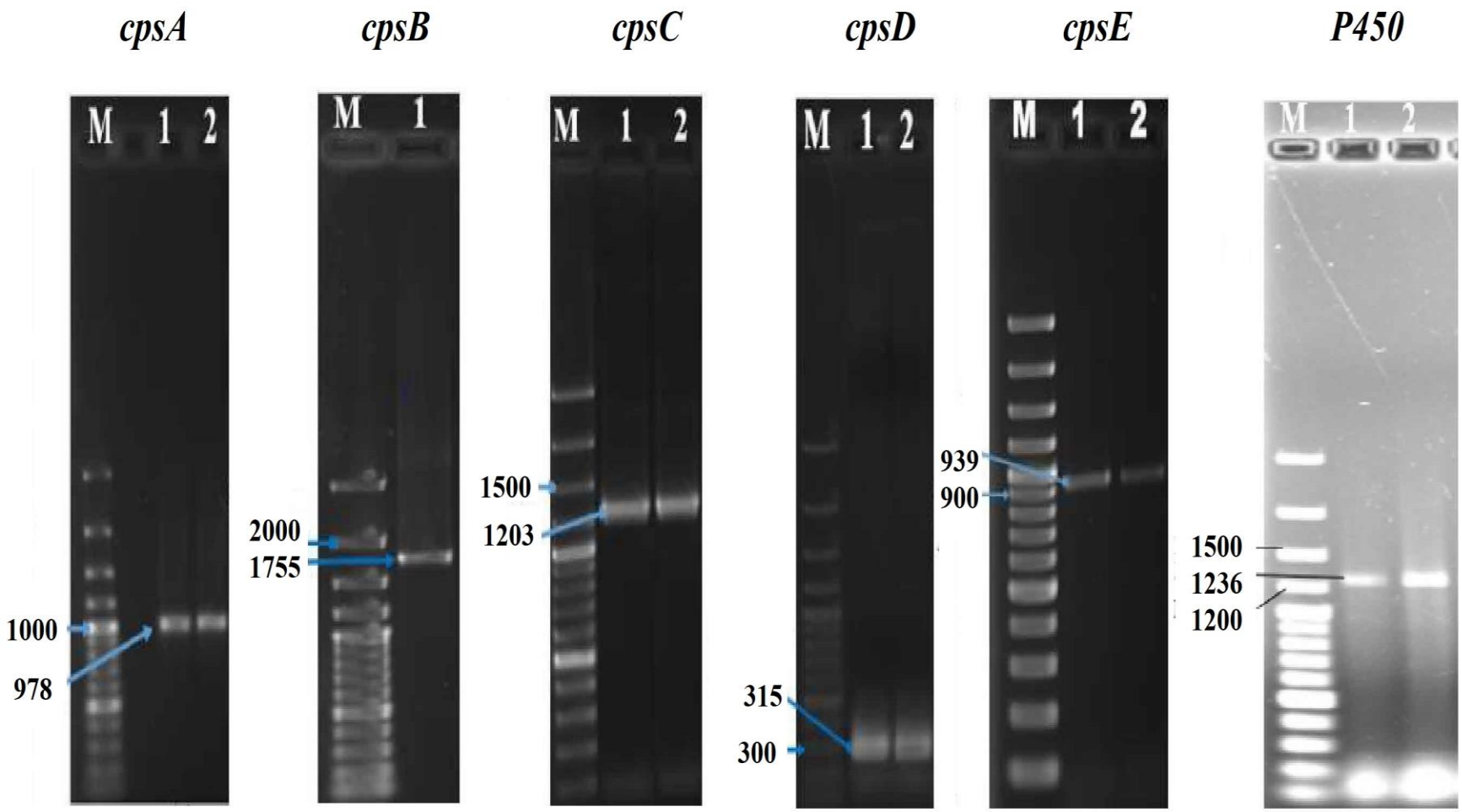


Figure 1A: Amplification of pentachlorophenol catabolic genes of *Bt*AOA. Lane M: DNA marker; Lane 1:amplified genes with their respective molecular weights.

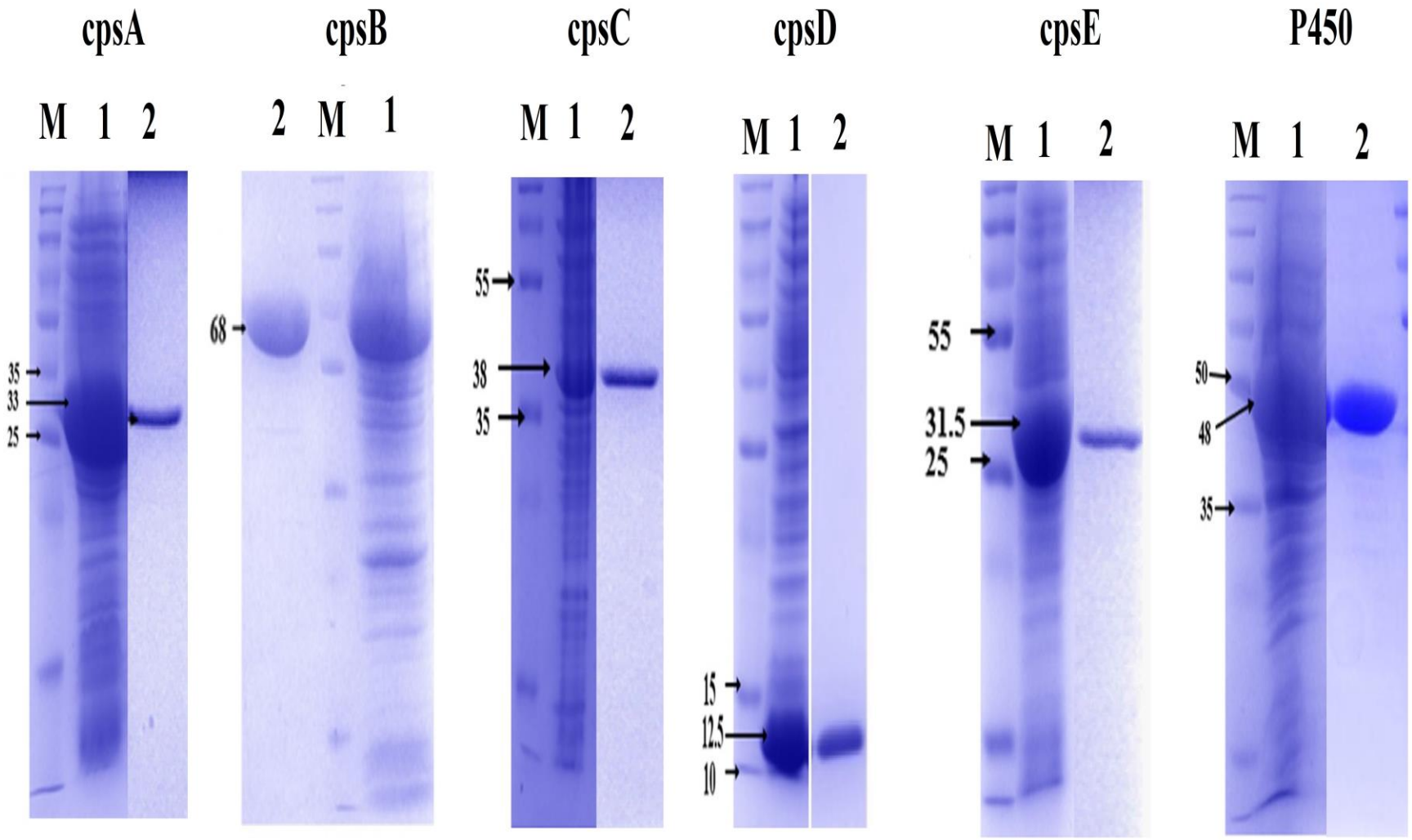


Figure 1B: Purification and SDS-PAGE of pentachlorophenol degrading enzymes of *Bt*AOA. Lane M: protein ladder; Lanes 1: expressed proteins in cell lysate; Lane 2: purified proteins.

Table 2: Characteristics and kinetics properties of PCP degrading enzymes of *Bt*AOA.

Properties	**Pentachlorophenol degrading enzymes**					
	CpsA	**CpsB**	**CpsC**	**CpsD**	**CpsE**	**P450**
Optimum pH	9.0	7.0	6.5	7.5	6.5	7.5
pH stability range	8-12[a]	6.0-7.5[a]	6.0-7.0[a]	6.0-7.5[a]	6.0-7.0[a]	7.0-9.0[b]
Optimum temperature (^{o}C)	30	30	40	40	40	40
Temperature stability range (^{o}C)	20-40[a]	25-30[b]	20-50[a]	25-50[a]	20-50[a]	25-45[a]
Substrate	DiCHQ	PCP	Tet-CHQ	Tet-CBQ	2-ClMA	PCP
Product	2-ClMA	THTH	DiCHQ	Tet-CHQ	CO	THTH
K_m (mM)	0.99	0.17	ND	0.99	ND	0.31
v_{max} (s^{-1})	27.77	0.034	ND	0.63	ND	0.08
K_{cat} (s^{-1})	4.20	4.58	ND	0.24	ND	0.013
K_{cat}/K_m ($s^{-1}mM^{-1}$)	4.24	26.94	ND	2.43	ND	0.043

[a]above 90% residual activity for 2h

[b]above 90% residual activity for 3h

DiCHQ -2,6-Dichloro-*p*-hydroquinone, 2-ClMA-2-chloromaleylacetate, PCP-pentachlorophenol, THTH-2,3,5,6-tetrakis[(2-hydroxyethyl)thio]-1,4-hydroquinone, Tet-CBQ- tetrachloro-1,4-benzoquinone, Tet-CHQ-tetrachloro-*p*-hydroqinone. CO- 5-Chloro-3-oxoadipate, ND-not determined

Table 3: Biophysical properties of PCP degrading enzymes of *Bt*AOA.

Biophysical and chemical properties	**Pentachlorophenol degrading enzymes**					
	CpsA	**CpsB**	**CpsC**	**CpsD**	**CpsE**	**P450**
No of amino acids residues	325	584	377	104	312	411
Molecular weight (kDa)	36.71	64.20	42.07	12.51	33.55	47.12
Theoretical pI	5.16	5.10	5.84	5.64	5.13	5.51
Total number of negatively charged residues (Asp + Glu)	49	76	45	19	41	61.49
Total number of positively charged residues (Arg + Lys)	33	48	38	16	34	2114
Total number of carbon (C)	1677	2874	1901	573	1500	3372
Total number of hydrogen (H)	2540	4504	2933	896	2441	564
Total number of nitrogen (N)	430	752	487	144	395	628
Total number of oxygen (O)	491	883	549	161	455	12
Total number of sulphur (S)	4	15	21	4	8	
Total number of atoms	5142	9028	5891	1778	4799	6690
Extinction coefficient (M^{-1} cm^{-1}), if all pairs of Cys residue form cystine	29910	54235	42330	20970	*14900	43890
Extinction coefficient (M^{-1} cm^{-1}) if all cystine residues are reduced	-	53860	41830	-	-	43890
Estimated half-life in *E. coli* in vivo	>10 h	>10 h	>10 h	>10 h	>10 h	>10 h
Yeast *in vivo*	>20 h	>20 h	>20 h	>20 h	>20 h	>20 h
Mammalian reticulocytes, *in vivo*	>30 h	>30 h	>30 h	>30 h	>30 h	>30 h
Instability index (II)	26.63	43.38	25.43	44.49	27.92	41.25
Classification based on instability index	stable	unstable	stable	unstable	stable	unstable
Aliphatic index	79.42	91.73	81.62	97.50	107.18	100.85
GRAVY	-0.288	-0.133	-0.229	-0.377	0.114	-0.319

(-) all cystine residues are oxidized; (*): The protein does not contain tryptophan residue and could result in more than 10% error in its computed extinction coefficient; GRAVY: grand average of hydropathicity.

10.2.4 Genome sequence characteristics

The draft whole genome sequences *de novo* assembly represented two contigs. The contigs were polished and the analysis metric of the polished contigs is shown in (Table S3). The final draft genome of *Bt*AOA consists of one chromosomal DNA and one plasmid DNA. The total length of the chromosomal DNA was 5,246,860 bp with an average GC content of 35.4%, while that of the plasmid length was 58,449 with an average GC content of 35.4% (Table 4). The mean GC contents for both the genetic elements; chromosomal and plasmid were found to be low and with same percentage which is in consonance with another report on the significant correlation between the G+C contents of plasmids and that of their bacterial host. The low GC content of the organism genome reflected on the annealing temperature (Tm ^{o}C) of all PCP catabolic genes (54°C–58°C) detected and amplified from the genome of the organism. Moreover, selective advantages drive the nucleotide composition of a cell since A+T contents of nucleotides are more abundant and are energetically less expensive than G+C contents (Almpanis et al. 2018). The features of the *Bt*AOA genome based on NCBI, RASTtk and PROKKA prokaryote genome annotation pipelines are shown in Table 4. The NCBI annotation indicated that *Bt*AOA chromosome encodes a total of 5400 genes, 5248 CDSs (total), 5138 coding genes, 5138 CDSs with protein, 152 RNA genes, 14 each of 5s, 16s, and 23s complete rRNAs, 105 tRNAs, 5 ncRNAs, 110 pseudo genes (total), and 110 CDSs without protein. Out of 110 pseudo genes, one gene showed the presence of ambiguous residues, 53 mutated by frameshift, 45 were found incomplete, 36 had internal stop codons and 20 were found with multiple problems. The RASTtk and PROKKA annotations also showed that the chromosomal DNA encodes for 5647 features, which include 5426 (96.09%) coding sequences (CDS), 133 (2.36%) RNAs and 88 (1.56%) of total number of repeats sequences. Furthermore, of the 5426 CDS found on the chromosome, 1306 (25%) were featured in subsystems while 4120 (75%) were not in any subsystem (Table 5). Of the 1306 CDS that were featured in a subsystem, 1232 (94.3%) encoded non-hypothetical protein while 74 (5.7%) were hypothetical. Also, of the 4120 CDS that were not in any subsystem, 2374 (57.6%) were non-hypothetical while 1746 (42.4%) were hypothetical proteins. The closest neighbour genome of *Bt*AOA as determined by the RASTtk server is *Bacillus cereus* ATCC 14579 with a similarity score of 516, followed by *B. anthracis* str. Ames, by *B. anthracis* str. Ames Ancestor, *B. thuringiensis* serovar Konkukian str. 87-27 and *B. cereus* biovar *anthracis* str. CI with similarities scores of 494, 477, 465 and 355, respectively (Table 6).

Table 4: Properties of *Bt*AOA genome based on annotation methods.

	Annotation pipelines		
Genome properties	**NCBI**	**RASTtk**	**PROKKA**
Genome Size (bp)	5,246,960	5,246,860	5,246,860
DNA G+C content (%)	35.4	35.4	35.4
L50	NA	1	NA
Number of contigs (with PEGs)	1	1	1
Total number of features	5,400	5647	5647
Number of coding sequences (CDS)	5,248	5426	5426
Number of RNAs	152	133	NA
Total number of repeats sequences	NA	88	88
Number of subsystems	NA	334	NA
Features in subsystems	NA	1306 (25%)	NA
Features not in subsystems	NA	4120 (75%)	NA
Genes (total)	5,400	NA	NA
Genes (coding)	5,138	NA	NA
CDSs (with protein)	5,138	NA	NA
Genes	152	NA	NA
rRNAs	14, 14, 14 (5S, 16S, 23S)	NA	NA
tRNAs	105	NA	NA
ncRNAs	5	NA	NA
Pseudo Genes (total)	110	NA	NA
CDSs (without protein)	110	NA	NA
Pseudo Genes (ambiguous residues)	1 of 110	NA	NA
Pseudo Genes (frameshifted)	53 of 110	NA	NA
Pseudo Genes (incomplete)	43 of 110	NA	NA
Pseudo Genes (internal stop)	36 of 110	NA	NA
Pseudo Genes (multiple problems)	20 of 110	NA	NA

NA: not analysed

Table 5: Subsystem category distribution of *Bt*AOA whole genome annotated at RASTtk.

S/N	Subsystem features	Count	Count %
01	Cofactors, vitamins, prosthetic groups, pigments	185	9.9
02	Cell wall and capsule	85	4.5
03	Virulence, disease and defense	61	3.3
04	Potassium metabolism	10	0.5
05	Phages, prophages, transposable elements, plasmids	10	0.5
06	Membrane transport	39	2.1
07	Iron acquisition and metabolism	52	2.8
08	RNA metabolism	57	3.0
09	Nucleosides and nucleotides	110	5.9
10	Protein metabolism	160	8.6
11	Cell division and cell cycle	6	0.3
12	Motility and chemotaxis	7	0.4
13	Regulation and cell signaling	33	1.8
14	Secondary metabolism	10	0.5
15	DNA metabolism	66	3.5
16	Fatty acids, lipids, and isoprenoids	65	3.5
17	Nitrogen metabolism	16	0.9
18	Dormancy and sporulation	108	5.8
19	Respiration	79	4.2
20	Stress response	38	2.0
21	Metabolism of aromatic compounds	12	0.6
22	Amino acids and derivatives	350	18.7
23	Sulphur metabolism	6	0.3
24	Phosphorus metabolism	22	1.2
25	Carbohydrates	255	13.6
26	Miscellaneous	27	1.4
27	Photosynthesis	0	0
	Total	**1869**	**99.8**

Table 6: Whole-genomes neighbours similar to *Bt*AOA.

S/N	Genome ID	Similarity score	Genome name
01	226900.1	516	*Bacillus cereus* ATCC 14579
02	198094.1	494	*Bacillus anthracis* str. Ames
03	261594.1	477	*Bacillus anthracis* str. 'Ames Ancestor'
04	281309.3	465	*Bacillus thuringiensis* serovar konkukian str. 97-27
05	637380.6	355	*Bacillus cereus* biovar *anthracis* str. CI
06	444177.5	328	*Lysinibacillus sphaericus* C3-41
07	491915.4	290	*Anoxybacillus flavithermus* WK1
08	592022.4	290	*Bacillus megaterium* DSM319
09	279010.5	284	*Bacillus licheniformis* ATCC 14580
10	235909.3	282	*Geobacillus kaustophilus* HTA426
11	420246.5	278	*Geobacillus thermodenitrificans* NG80-2
12	221109.1	257	*Oceanobacillus iheyensis* HTE831
13	224308.1	250	*Bacillus subtilis* subsp. *subtilis* str. 168
14	272558.1	212	*Bacillus halodurans* C-125
15	272626.1	191	*Listeria innocua* Clip11262
16	169963.1	188	*Listeria monocytogenes* EGD-e
17	439292.5	181	*Bacillus selenitireducens* MLS10
18	66692.3	170	*Bacillus clausii* KSM-K16
19	350688.3	169	*Alkaliphilus oremlandi oremlandii* OhILAs
20	649639.5	168	*Bacillus cellulosilyticus* DSM 2522
21	262543.4	162	*Exiguobacterium sibiricum* 255-15
22	358681.3	155	*Brevibacillus brevis* NBRC 100599
23	158878.1	150	*Staphylococcus aureus* subsp. *aureus* Mu50
24	158879.1	147	*Staphylococcus aureus* subsp. *aureus* N315
25	324057.4	137	*Paenibacillus* sp. JDR-2
26	458233.11	137	*Macrococcus caseolyticus* JCSC5402
27	176279.3	120	*Staphylococcus epidermidis* RP62A
28	282458.1	102	*Staphylococcus aureus* subsp. *aureus* MRSA252
29	413999.4	101	*Clostridium botulinum* A str. ATCC 3502
30	195103.9	98	*Clostridium perfringens* ATCC 13124

Table was computed using the genome annotated at RASTtk

The RASTtk ModelSEED database was used to identify the metabolic pathways in *Bt*AOA, using the Kyoto Encyclopedia of Gene and Genomes (KEGG) and to plot a comprehensive metabolic reconstruction for *Bt*AOA. The metabolic model for *Bt*AOA showed the presence of core bioprocesses such as respiration, secondary metabolite synthesis, stress response, cell division and cell cycle, metabolism of aromatic compounds, dormancy and sporulation, DNA metabolism, regulation and signalling, protein metabolism, carbohydrates metabolism, phosphorus metabolism, nitrogen metabolism, xenobiotic biodegradation and others (Fig. 2, Table 5). The subsystem category distribution of *Bt*AOA genome annotated at RASTtk showed that 1869 functioning features were present in subsystems. Of this, 350 (18.7%) feature were involve in amino acid and derivatives subsystems, 185 (9.9%) were involve in cofactors, vitamins, prosthetic group and pigments, protein metabolism 160 (8.6%), nucleotides and nucleotides 110 (5.9%), carbohydrate 255 (13.6%), cell wall and capsule 85 (4.5%) and virulence, disease and defence 61 (3.3%) and remaining features are not describes here (Table 5). The presence of dormancy and sporulation, motility and chemotaxis, cell wall and capsule subsystems with their proteins showed that the organism is motile and spore forming.

10.2.5 Metabolic reconstruction of *Bt*AOA xenobiotics biodegradation pathways

The various xenobiotics degradation pathways for *Bt*AOA and metabolic reconstruction model is shown in Table 7. The *Bt*AOA metabolic model showed that the strain can degrade benzoate via CoA ligation or hydroxylation and encoded for 9 (20.5%) and 6 (12.0%) of the unique enzymes involved in the benzoate degradation via CoA ligation and hydroxylation pathways, respectively (Table 4). The model also showed that the organism could degrade gamma-hexachlorocyclohexane, 1,2-dichloroethane, 1,4-dichlorobenzene, 2,4-dichlorobenzene, 3-chloroacrylic acid, trinitrotoluene, ethylbenzene, atrazine biphenyl, bisphenol, naphthalene, anthracene and styrene which are all listed as priority pollutants (IARC 2019a, b; Stockholm convention 2019). The presence of genes encoding the enzyme involved in the degradation of chlorophenol and benzoates via hydroxylation also justify its ability to degrade pentachlorophenol and other congeners. The organism has been found to degrade PCP, 2,4,6-trichlorophenol, 2,4-dichlorophenol and other *orthos*- and *para-substitute*d phenolic compounds via hydroxylation the para-position (data not shown). The model, however, showed that *Bt*AOA lacks the capacity to degrade DDT and tetrachloroethane.

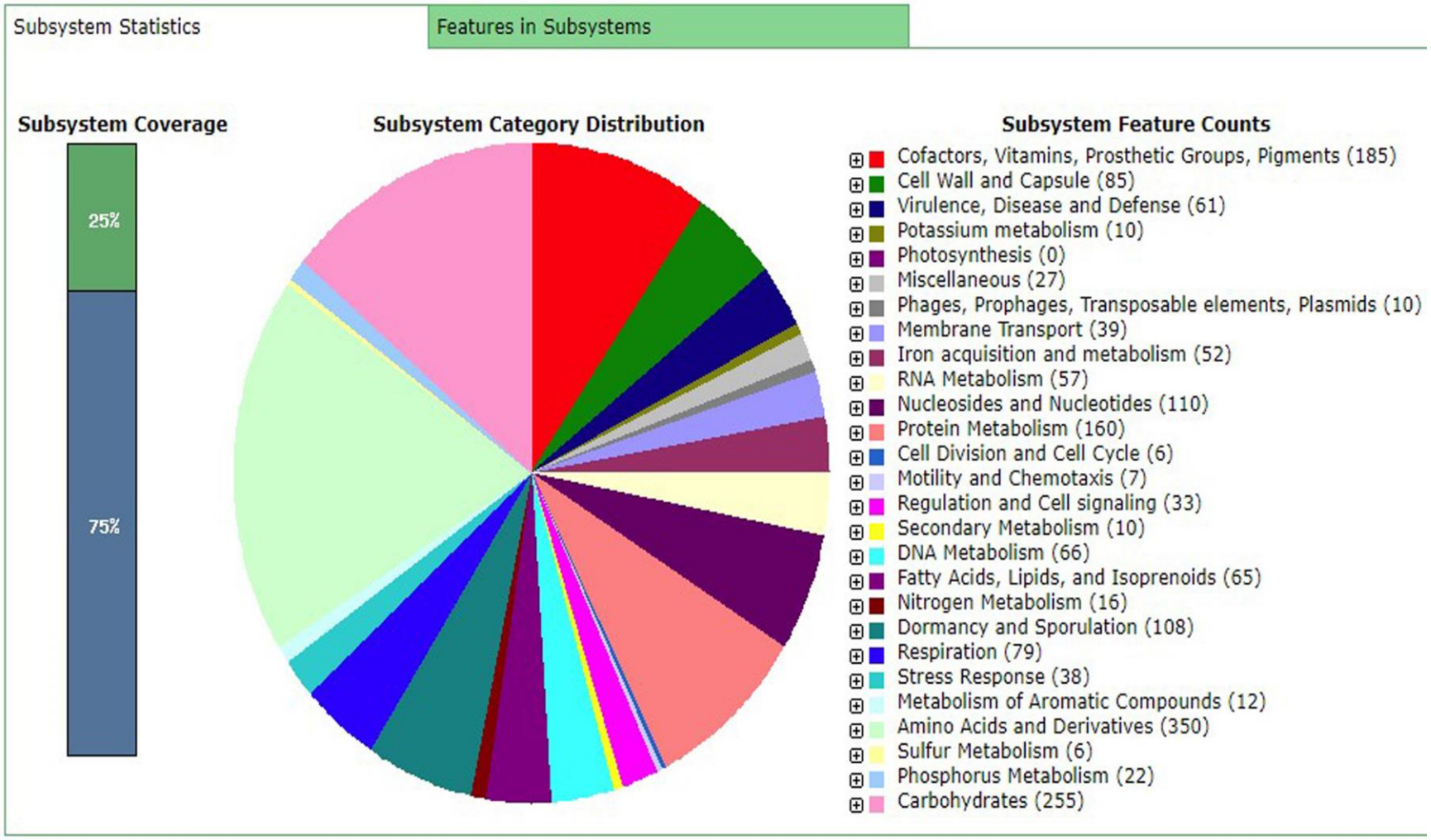


Figure 2: Subsystem coverage, subsystem category distribution and subsystem feature counts of annotated genome of *Bt*AOA. Graphics picture copied from RAST SEED Viewer v.2.0.

Table 7: Xenobiotic biodegradation pathways in *Bt*AOA metabolic reconstruction model deduced from RAST Model SEED v.2 using KEGG annotation pipeline.

S/N	Xenobiotics	Distinct ECs	In BtAOA	(%)
01	1- and 2-Methylnaphthalene degradation	17	03	17.6
02	DDT degradation	10	00	00.0
03	1,2-Dichloroethane degradation	05	02	40.0
04	1,4-Dichlorobenzene degradation	22	03	13.6
05	2,4-Dichlorobenzoate degradation	29	01	3.40
06	3-Chloroacrylic acid degradation	04	02	50.0
07	Atrazine degradation	12	01	8.30
08	Benzoate degradation via CoA ligation	44	09	20.5
09	Benzoate degradation via hydroxylation	50	06	12.0
10	Biphenyl degradation	13	01	7.70
11	Bisphenol A degradation	09	01	11.1
12	Caprolactam degradation	21	05	23.8
13	Carbazole degradation	11	01	9.10
14	Ethylbenzene degradation	11	03	27.3
15	Fluorene degradation	13	03	23.1
16	gamma-Hexachlorocyclohexane degradation	28	05	17.9
17	Geraniol degradation	12	03	25.0
18	Glycosaminoglycan degradation	16	01	6.20
19	Glycosphingolipid biosynthesis - ganglio series	11	03	27.3
20	Limonene and pinene degradation	22	04	18.2
21	Lysine degradation	54	11	20.4
22	Naphthalene and anthracene degradation	22	02	9.10
23	Other glycan degradation	09	01	11.1
24	Phenylalanine metabolism	51	11	21.6
25	Styrene degradation	21	04	19.0
26	Synthesis and degradation of ketone bodies	06	03	50.0
27	Synthesis and degradation of ketone bodies	06	03	50.0
28	Tetrachloroethene degradation	07	00	00.0
29	Toluene and xylene degradation	23	03	13.0
30	Trinitrotoluene degradation	07	01	14.3
31	Valine, leucine and isoleucine degradation	34	17	50.0

1,1,1-Trichloro-2,2-bis(4-chlorophenyl)ethane (DDT)

10.2.6 PCP degradation genes in *Bt*AOA

The NCBI, RASTtk and PROKKA annotations showed the presence of PCP catabolic genes encoding the enzymes CpsBDCAE and P450 which were further individually detected, amplified, cloned, overexpressed in expression host, purified, and characterized (Fig. 1A and 1B, Table 2) in *Bt*AOA genome. The PCP-degrading enzymes (CpsB, CpsD, CpsC, CpsA, CpsE and P450) in *Bt*AOA genome were annotated by NCBI as aromatic amino acid hydroxylase (gene locus_tag=GM610_18120), 4α-hydroxytetrahydrobiopterin dehydratase (EC:4.2.1.96; gene locus_tag=GM610_18125), alcohol dehydrogenase EutG (EC:1.1.1.1, gene locus_tag =GM610_07555), ring-cleaving dioxygenase (gene locus_tag=GM610_02225), malate dehydrogenase (EC:1.1.1.37, gene locus_tag=GM610_19285) and cytochrome P450 (EC:1.14.14.1, gene locus_tag=GM610_09540), respectively.

RASTtk and PROKKA annotation data also showed the presence of all the enzymes involved in PCP degradation pathway. CpsB, CpsD, CpsC, CpsA, CpsE and cytochrome P450 were found by the entry name as phenylalanine-4-hydroxylase (EC 1.14.16.1)/N-terminal regulatory domain, pterin-4-alpha-carbinolamin dehydratase, glutathione-dependent formaldehyde dehydrogenase, glyoxalase family protein/putative ring cleaving dioxygenase and malate dehydrogenase (EC. 1.1.1.37) and cytochrome P450, respectively.

In PROKKA annotated data, CpsB, CpsD, CpsC, CpsA, CpsE and cytochrome P450 were found by the entry name as hypothetical protein, putative pterin-4-alpha-carbinolamin dehydratase, putative zinc-binding alcohol dehydrogenase, putative ring cleaving dioxygenase and malate dehydrogenase (EC. 1.1.1.37) and cytochrome P450, respectively.

The presence of these catabolic enzymes is not surprising based on the metabolic reconstruction of xenobiotic degradation model presented (Table 7). These PCP degradation genes are found to be located in the chromosome. The genes are found in proximity to each other, CpsB and CpsD are located on the same operon while CpsC, CpsA and CpsE are located in proximity to the CpsBD operon (Figure 3).

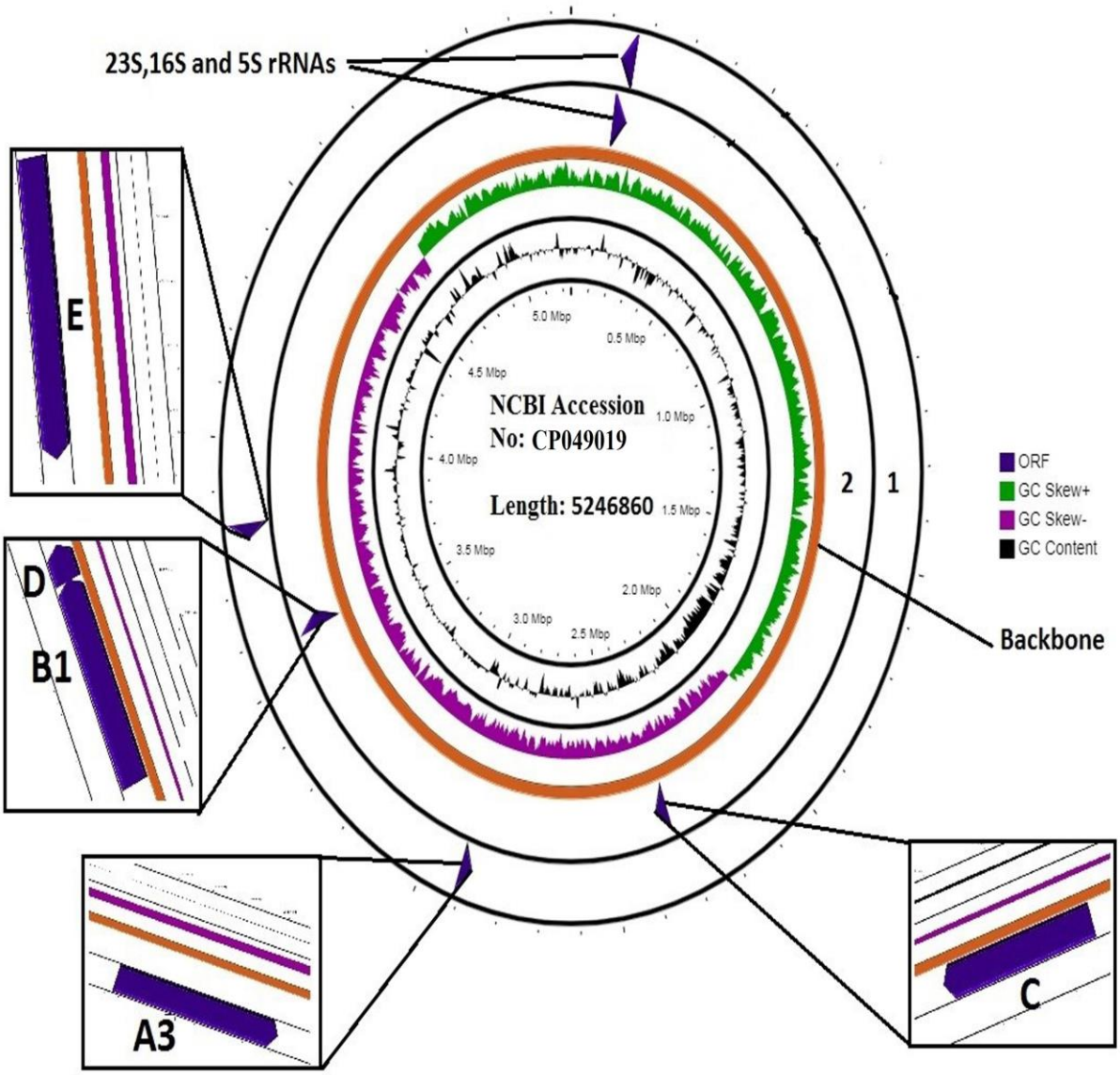


Figure 3: Customized graphical map of the *Bt*AOA chromosome showing the length of the chromosome (5,246,860), backbone of the chromosome, 23S-, 16S- and 5S rRNA regions, reverse open reading frame (1), forward open reading frame (2), GC content (black), GC skew+ (green), GC skew- (pink) and locations of the functioning PCP degradation genes. (A) 2,6-dichloro-p-hydroquinone 1,2-dioxygenases (CpsA), (B) Pentachlorophenol 4-monooxygenase CpsB, (C) Glutathione-dependent formaldehyde dehydrogenase CpsC, (D) Pterin-4-alpha-carbinolamine dehydratase (CpsD) and (E) Malate dehydrogenase (CpsE). The circular graphical map was drawn on CGView server (Grant and Stothard 2008) and annotation features were added by integrated program PROKKA 2.6 (Seemann 2014).

10.2.7 Regulation of the expression of PCP degradation genes

Gene neighbourhoods surrounding PCP catabolic genes in *Bt*AOA is presented in figure 3. The operon encoding CpsB and CpsD also contained transcription regulator (MerR) family, minor extracellular protease 'VpR' and a hypothetical protein. CpsC neighbourhood's genes are small acid-soluble spore protein from α/β family 'SASP_6', manganese catalase (EC 1.11.1.6), uncharacterized protein YvdQ and integral membrane protein. CpsA is surrounded by another uncharacterized protein, glyoxalase/bleomycin resistance protein/dioxygenase superfamily (another CpsA), carD-like transcriptional regulator, sporulation kinase A, carboxylesterase (EC 3.1.1.1), osmo-regulated sodium/proline symporter OpuE, CcdC protein, ferredoxin-dependent glutamate synthase (EC 1.4.7.1), 2,4-dienoyl-CoA reductase (NADPH) (EC 1.3.1.34), and formiminoglutamase (EC 3.5.3.8). While CpsE is surrounded by transcriptional regulatory protein (PhoP), phosphate regulon sensor protein PhoR (EC 2.7.13.3), MaoC like domain protein, isocitrate dehydrogenase (EC 1.1.1.42), citrate synthase (EC 2.3.3.1) and DNA polymerase (EC 2.7.7.7). From the neighbourhoods' genes surrounding the PCP catabolic genes, it was observed that all the catabolic genes except CpsC surrounded by probable regulatory genes. However, these regulatory genes have not been detected experimentally and need experimental data to ascertain whether these genes regulate the expression of the catabolic genes or involve in PCP degradation, in the current study.

10.2.8 Evolution of PCP catabolic genes in *Bt*AOA

The *Bt*AOA chromosome houses genes encoding chromosome and plasmid partitioning proteins ParA, ParB and ParB-2 which are involved in plasmid replication. It was also observed that seven transposable elements were distributed all over the chromosome of *Bt*AOA. These enzymes serve as tools for horizontal genes transmission (HGT), in agreement with reports that *de novo* protein and secondary chromosomes could evolve in an organism by intra-genomic transfer of essential genes to a plasmid (Slater et al. 2009; Huang et al. 2012). The presence of these transposable elements shows a trend of continuous assault by a transposable genetic element like other PCP degrading microbe *S. chlorophenolicum*. However, the PCP catabolic genes were located on the chromosome and not associated with these mobile elements which are also consistent with that of *S. chlorophenolicum* (Copley et al. 2012).

Horizontal gene transmission is common among microorganisms (Cohen et al. 2013; Rossoni et al. 2019) and it has long been recognized as a critical process in the evolution of prokaryotes (Wickell and Li 2020). The HGT events play major roles in the acquisition of virulence, resistance, persistence, tolerance, and most especially xenobiotic degradation, in response to environmental stressors. More importantly, HGT events have led to the transfer of functional nuclear genes with significant adaptive characteristics have been reported in some taxa (Wickell and Li 2020). Unlike the gradual process by which newly duplicated genes diverge, HGT can have immediate and significant effects for host adaptation, fitness, and tolerance.

Moreover, multiple HGT events significantly confer a selective advantage in diverse clusters (Dunning et al. 2019; Yang et al. 2019). Furthermore, genes that could aid microbe's adaptation, tolerance, survival, and growth in the presence of xenobiotics and other environmental stressors could also evolve via homologous recombination (Rossoni et al. 2019; Crickard and Greene 2019), gene duplication and divergences (Teufel et al. 2019; Wickell and Li 2020). However, unlike the gradual process by which newly duplicated genes diverge, HGT can have immediate and significant effects for host adaptation, fitness, and tolerance.

Extensive mining of *Bt*AOA genome revealed the locations of PCP degrading enzymes and provided insights into the evolution in *Bt*AOA PCP degradation pathway (Fig. 3). PCP metabolic genes (*cps*BDCAE) are all chromosomally encoded and are found at different sites. The *cps*B and *cps*D were located on the same operon as the pterin metabolism subsystem. Three variants of CpsA (with different annotations) were found in *Bt*AOA genome and were located at two different sites (which signifies gene duplication events) whereas *cps*C and *cps*E were seen together.

Recruitments of these genes into *Bt*AOA PCP degradation pathway can occur via different evolutionary events ranging from: (a) horizontal and plasmid transduction of xenobiotic degradation enzymes with promiscuous catalytic activities that became useful (in circumventing the toxicity of PCP) when *Bt*AOA is exposed to high concentration of PCP; (b) recruitment of an ancestral enzymes without gene duplication which require Shuffling of the metabolic enzymes between the original and new functions and; (c) recruitment of pre-existing metabolic enzymes

following gene duplication and homologous recombination of the original gene to provide a copy that is not optimized for its function in PCP degradation pathway (Copley et al. 2012).

Genome comparison has previously been used to determine which of the above evolutionary events led to the emergence the PCP degradation genes in *S. chlorophenolicum* (Copley et al. 2012). If a gene evolves via HGT or plasmid transduction, there would be no close homologs found in either *Bt*AOA or its closest ancestor (*B. cereus* ATCC 14579) but might be found in other variants (Copley et al. 2012). If an enzyme is recruited from the most recent common ancestor of *Bt*AOA and *B. cereus* ATCC 14579, then a close homolog is likely to be found in *B. cereus* ATCC 14579. If gene duplication and homologous DNA recombination have resulted in the specialization of an enzyme for its role in PCP degradation, a close homologue should be found in the genome of the organism (Copley et al. 2012). CpsB is located upstream to CpsD on *Bt*AOA chromosome, both are sense (+) strands and are surrounded by transcription regulator (MerR) family and minor extracellular protease *VpR* (Fig. 4) similar to that of *B. subtilis* str. *subtilis* 168 (Hirose et al. 2000)**.**

The arrangement of CpsB and CpsD is similar to a typical bacterial phenylalanine hydroxylating system (Litwack 2018; Mitchell and Steventon 2019) and that of *S. chlorophenolicum* PCP degradation pathway (Copley et al. 2012). CpsD is currently annotated as pterin-4-alpha-carbinolamine dehydratase PCD (EC 4.2.1.96) by RAST SEED Viewer. The KEGG metabolic analysis of CpsD showed that the enzyme is part of pterin carbinolamine dehydratase subsystem in *Bt*AOA and its role in the subsystem is PCD (EC 4.2.1.96). PCD plays a bifunctional (regulatory and catalytic) role in PheOHS (Naponelli et al. 2008; Wang et al. 2015) and is also involved in folate biosynthesis in bacterial system (Engevik et al. 2019; Kok et al. 2020).

Also, CpsB is currently assigned phenylalanine-4-hydroxylase 'Phe4MO' (EC 1.14.16.1)/N-terminal regulatory domain by RAST. Analysis of *Bt*AOA metabolic reconstruction (KEGG) model showed that CpsB is a part of aromatic amino acid degradation and pterin carbinolamine dehydratase subsystems and its role is Phe4MO (EC 1.14.16.1). Phe4MO is a member of an aromatic amino acid hydroxylase (Mitchell and Steventon 2019), required in the transformation of phenylalanine (Phe) to tyrosine (Greule et al. 2018), during dopamine (Mitchell and Steventon 2019), in the PheOHS (Wang et al. 2013).

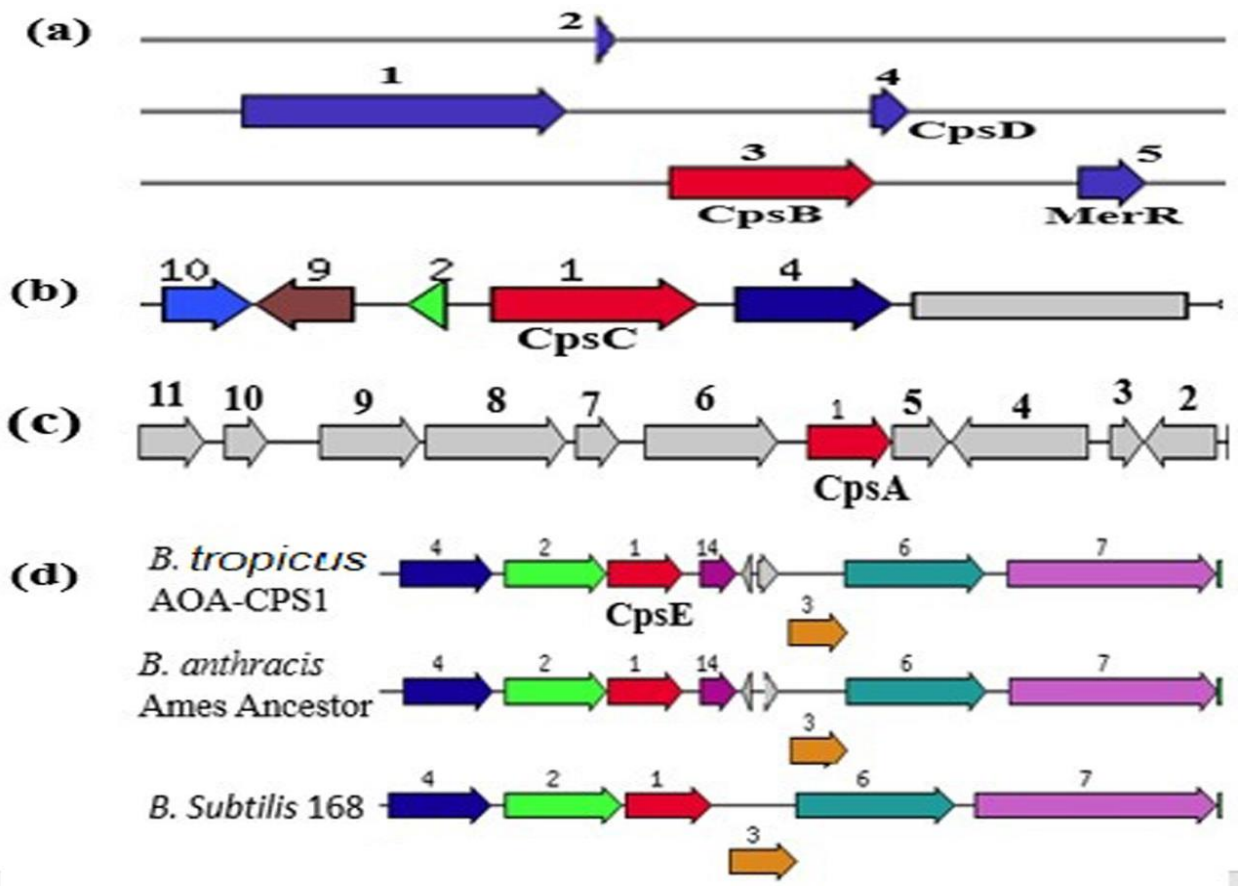


Figure 4: Gene's neighborhoods surrounding PCP degradation genes in *Bt*AOA. Predicted PCP degradation enzymes in *Bt*AOA PCP degradation pathway with neighborhoods proteins surrounding the PCP degradation enzymes and their probable regulatory protein. **(A)** Minor extracellular protease 'VpR' (1), hypothetical protein (2), phenylalanine-4-hydroxylase/N-terminal regulatory domain (3), pterin-4-alpha-carbinolamine dehydratase (4) and transcription regulator (MerR) family (5). **(B)** Predicted CpsC, glutathione-dependent formaldehyde dehydrogenase (1), small acid-soluble spore protein, α/β family 'SASP_6 (2), manganese catalase (EC 1.11.1.6) (4), Uncharacterized protein YvdQ (9) and integral membrane protein (10). **(C)** Predicted CpsA, ring-cleaving dioxygenase (1), uncharacterized protein (2), glyoxalase/bleomycin resistance protein/dioxygenase superfamily (3), sporulation kinase A (4), carboxylesterase (EC 3.1.1.1) (5), osmoregulated sodium/proline symporter OpuE (6), CcdC protein (7), ferredoxin-dependent glutamate synthase (EC 1.4.7.1) (8), 2,4-dienoyl-CoA reductase (NADPH) (EC 1.3.1.34) (9), carD-like transcriptional regulator (10) and formiminoglutamase (EC 3.5.3.8) (11). **(D)** Predicted CpsE and other strains with similar sequences homology CpsE as malate dehydrogenase (1), (14) MaoC like domain protein, (2) isocitrate dehydrogenase (EC 1.1.1.42), (4) citrate synthase (EC 2.3.3.1), (3) transcriptional regulatory protein (PhoP), (6) phosphate regulon sensor protein PhoR (SphS) (EC 2.7.13.3), and (7) DNA polymerase (EC 2.7.7.7).

Phe4MO is involved in secondary metabolites degradation and enzymes associated with intermediary metabolisms such as CpsB and CpsD may have some yet unrecognized alternative catalytic function(s) of protecting the organism from potential toxic compounds (Mitchell and Steventon 2019). Phe4MO also plays a critical role in xenobiotic degradation (Steventon and Mitchell 2009; Nguyen et al. 2012; Wang et al. 2013; Parthasarathy et al. 2018; Flydal et al. 2019; Hidalgo and Zamora 2019) and lipids metabolism (Wang et al. 2013). We explored the catalytic promiscuities of both CpsB and CpsD (individually) in enzymatic PCP biodegradation *in vitro* and found that CpsB is involved in the hydroxylation of PCP to Tet-CBQ (data not shown), similar to PCP 4-monooxygenase (PcpB) (Hlouchova et al. 2012), while CpsD catalysed the enzymatic reduction of Tet-CBQ to Tet-CHQ, which is also similar to Tet-CBQ reductase (PcpD) (Copley et al. 2012; Yadid et al. 2013). However, PcpBD and CpsBD are not closely related, and the catalytic efficiencies of CpsB and CpsD were far higher than those of PcpB and PcpD (Table 2).

Pterin metabolism subsystem in *Bt*AOA metabolic reconstruction model is made up of Phe4MO (EC 1.14.16.1), PCD (EC 4.2.1.96), PCD-like, and uncategorized protein COG3146. The synthase and the aldolase were also part of the enzymes around CpsB and CpsD in the whole genome of the isolate (Fig. 4). Based on the recent studies on CpsB and CpsD, pterin metabolism subsystem may be recruited into the PCP degradation pathway in *Bt*AOA via HGT events.

CpsC encoding gene is also encoded by chromosome of *Bt*AOA and surrounded by small acid-soluble spore protein, manganese catalase and integral membrane protein in *B. substilis* str. Ames ancestor genome. CpsC is currently annotated as a glutathione-dependent formaldehyde dehydrogenase (GSH-FDH) in *Bt*AOA genome and it is not yet part of a metabolic subsystem in the *Bt*AOA metabolic reconstruction model and xenobiotic degradation pathways detected in *Bt*AOA via KEGG metabolic analysis. GSH-FDH plays a prominent role in the metabolism of glutathione adducts such as S-(hydroxymethyl)glutathione and S-nitrosoglutathione in eukaryotes (Sanghani et al. 2002). It has also been reported that mutations which decrease or increase the expression of CpsC homologs changed the specific activity of GSH-FDH in *Rhodobacter sphaeroides* extracts (Barber et al. 1996).

Gene duplication has earlier been recognized as a major source of new genes and functions; however, genomes comparisons have suggested that most of the new genes emanated from gene duplication may have no new functions (Tanaka et al. 1995; Chebrou et al. 1999). In addition, paralogous gene pairs are often sub-functional and may have overlapping activities that allow both complementation and maintenance of the original functions and independent evolution of other biological functions (Roca et al. 2009). This was evident in the detoxification of nitroaromatic compounds in *Pseudomonas putida* and other organisms by flavoproteins (Van Den Hemel et al. 2006; Van Dillewijn et al. 2008). CpsC might have acquired the ability to transform Tet-CHQ via gene duplication, spontaneous mutations or homologous recombination events and recruited into *Bt*AOA PCP degradation via HGT.

The assignment of CpsC as GSH-FDH in *Bt*AOA may be erroneous, due to the fact that similar protein was annotated as glutathione-dependent formaldehyde dehydrogenase and alcohol dehydrogenase in *B. subtilis* strain R5 genome while the experimental evidence showed that the protein is an alcohol dehydrogenase (Raza et al. 2017). Furthermore, a gene encoding a GSH-FDH homolog has also been identified as part of an operon (adhI-cycI) that also encodes an isoform of cytochrome c2 family of electron transport proteins (Barber et al. 1996). Automatic pipelines can produce inaccurate genome annotation, thus necessitating manual curation (Richardson and Watson 2013). We have cloned and expressed CpsC heterologously (Fig. 1B), and experimental evidence showed that CpsC is a GSH dependent Tet-CHQ dehalogenase (Table 2) similar to glutathione-S-transferase (PcpC) in *S. chlorophenolicum* (Belchik and Xun 2012).

The putative CpsA/glyoxalase family protein was annotated as putative ring-cleaving dioxygenase MhqA by RASTtk and MhqO in other organisms. CpsA is not yet a part of a subsystem in the metabolic model of *Bt*AOA. The presence of MhqA around glyoxalase family protein (CpsA) which happen to be the ancestral protein showed that the gene has been duplicated. The MhqA and MhqO may be the variants of the putative glyoxalase family protein (CpsA). CpsA might have been recruited into *Bt*AOA PCP degradation pathway via HFT event.

In this study, CpsA has been cloned (Fig. 1A), overexpressed heterologous, purified to homogeneity (Fig. 1B) and characterized to belong to the putative glyoxalase family protein (Fig. 1A, 1B). The study also discovered the biological role of the putative CpsA and provided experimental shreds of evidence that CpsA is a ring-cleaving dioxygenase that catalyses the aromatic ring-cleavage of 2,6-dichloro-p-hydroquinone to 2-chloromaleylacetate in *Bt*AOA PCP degradation pathway (Table 2). The functional role of CpsA in PCP degradation is similar to that of PcpA (PDB: 4huz) from *S. chlorophenolicum* (Hayes et al. 2013).

Interestingly, the template (PDB: 1ZSW) used to construct the 3D structural model of PcpA from *S. chlorophenolicum* (Machonkin et al. 2010) is the same putative glyoxalase family protein from *B. cereus* ATCC 14579 which happened to be the most similar whole-genome neighbour of *Bt*AOA (Table 6). Therefore, we propose that CpsA be named 2,6-dichloro-p-hydroquinone 1,2-dioxygenase and should be added to the hydroquinone 1,2-dioxygenase of the vicinal oxygen chelate (VOC) superfamily as earlier proposed (Machonkin et al. 2010; Hayes et al. 2013).

CpsE is annotated as a homolog of malate dehydrogenase (EC 1.1.1.37), according to the enzyme commission. The neighbouring enzymes to cpsE (Fig. 4d) were alkaline phosphatase synthesis transcriptional regulatory protein (PhoP), MaoC like domain protein, isocitrate dehydrogenase (EC 1.1.1.42), citrate synthase (EC 2.3.3.1), phosphate regulon sensor protein PhoR (SphS) (EC 2.7.13.3) and DNA polymerase (EC 2.7.7.7). The arrangement of CpsE and its neighbouring enzymes is similar to those of *B subtilis* subsp. *subtilis* str. 168 and *B. anthracis str*. Ames Ancestor. The enzyme is annotated for the same functional role in the genomes of *B. subtilis* subsp. *Subtilis* str. 168 (Jin et al. 1996), *B. anthracis* str. Ames Ancestor, *B. cereus* ATCC 10987 and some other *Bacillus* species.

According to the RASTtk annotation and KEGG metabolic analysis, CpsE is a part of the subsystems of TCA Cycle, serine-glyoxylate cycle and glyoxylate bypass pathway in *Bt*AOA metabolic model and its annotated functional role in those pathways is malate dehydrogenase (EC 1.1.1.37). The enzyme is involved in several microbial metabolisms in diverse environments, including polycyclic aromatic hydrocarbon and chlorobenzene degradation. According to the *Bt*AOA metabolic reconstruction model, CpsE may be in clusters (2, 3 and 4 sizes) function with other proteins.

The homologs of CpsE which may be in cluster functions (2 cluster size) are malate dehydrogenase (EC 1.1.1.37) and isocitrate dehydrogenase [NADP] (EC 1.1.1.42). This two-size cluster is found in *B. halodurans* C-125, *Oceanobacillus iheyensis* HTE831, *Symbiobacterium thermophilum* IAM 14863, *Heliobacterium modesticaldum* Ice1, *Blastopirellula marina* DSM 3645, *Aurantimonas* sp. SI85-9A1, *Sinorhizobium meliloti* 1021, *Oceanicola batsensis* HTCC2597 and *Silicibacter pomeroyi* DSS-3. Homologs that may be in three size cluster with CpsE are isocitrate dehydrogenase (NADP) (EC 1.1.1.42) and Succinyl-CoA ligase (ADP-forming) β-chain (EC 6.2.1.5).

This 3 size cluster functions are present in the genomes of *Anaeromyxobacter* sp. K, *Anaeromyxobacter* sp. Fw109-5, *Agrobacterium tumefaciens* str. C58 and *Rhizobium leguminosarum* bv. *viciae* 3841. Whereas those that may form four cluster function with CpsE are 2-oxoglutarate dehydrogenase E1 component (EC 1.2.4.2), AFG1-like ATPase and succinyl-CoA ligase (ADP-forming) β-chain (EC 6.2.1.5) which was found *Azorhizobium caulinodans* ORS 571 genome (Hannenhalli and Russell 2000). CpsE might have been recruited from either the TCA cycle or serine-glyoxalase cycle into *Bt*AOA PCP degradation pathway via horizontal gene transfer.

*Bt*AOA seems to have assembled a metabolically efficient pathway for PCP degradation (Fig. 5). The first two steps involve hydroxylation of PCP to Tet-CBQ (catalysed by CpsB) and reduction of the Tet-CBQ to Tet-CHQ (catalysed by CpsD). These two steps are believed to probably take place in the pterin metabolism subsystem recruited into the pathway and the reduced product (Tet-CHQ) released from the system to CpsC. The futile cycle that limited the catalytic efficiencies of PcpB and PcpD in *S. chlorophenolicum* (Su, Yunyou. et al. 2008) may have been circumvented in the CpsBD subsystem, as a way of protecting the organism from the toxic metabolite (Tet-CBQ). If the reactions take place as proposed, there will be no substrate for the futile cycle/redox cycles to take place in the CpsBD subsystem. The next two steps in the pathway involve sequential dehalogenation of Tet-CHQ to Di-CHQ by glutathione-dependent-CpsC. Di-CHQ is converted to 2-CIMA by CpsA while the final state involves the reduction of 2-CMA to MA in a reaction catalysed by CpsE.

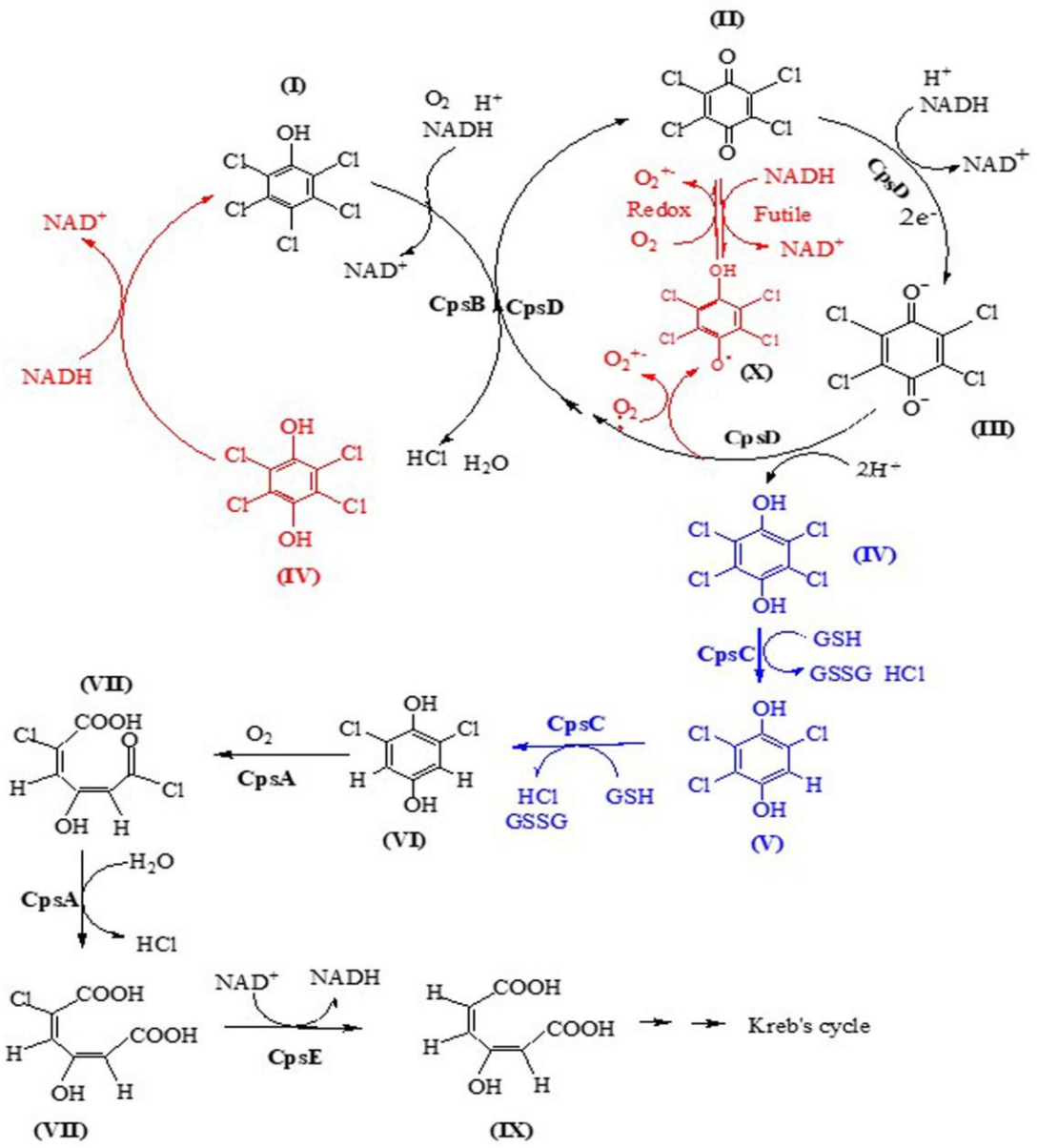


Figure 5: Proposed *Bt*AOA PCP degradation pathway. (I) pentachlorophenol PCP, (II) tetrachloro-1,4-benzoquinone Tet-CBQ, (III) deprotonated tetrachloro-1,4-benzoquinone Tet-CBQ, (IV) tetrachloro-1,4-hydroquinone Tet-CHQ, (V) 2,5,6-trichloro-1,4-hydroquinone Tri-CHQ, (VI) 2,6-dichloro-1,4-hydroquinone Di-CHQ, (VII) cleaved 2,6-dichloro-1,4-hydroquinone intermediate, (VIII) 2-chloromaleylacetate 2-CMA, (IX) maleylacetate MA, (X) tetrachloro-1,4-benzoquinone Tet-CBQ radical. The reactions shaded in red were the futile cycle identified in *S. chlorophenolicum* PCP pathway which may not take place in the *Bt*AOA PCP degradation pathway due to unavailability of the substrates for the reaction to take place.

10.2.9 Comparative analysis of metabolic reconstruction of the genome of *Bt*AOA with other related strains

Comparison of the metabolic reconstruction of the genome *Bt*AOA to that of *B. cereus* ATCC 14579 and other related species allowed for the comparison of the functioning parts of *Bt*AOA with that of *B. cereus* ATCC 14579 and other related strains, to determine the distribution of *Bt*AOA PCP catabolic enzymes in other strains. Metabolic reconstruction of *Bt*AOA annotated at RASTtk showed that CpsB is in the categories of amino acids and derivatives, sub-categories of aromatic amino acids and derivatives, in a subsystem aromatic amino acids degradation and pterin-carbinolamine dehydratase and its functioning role is phenylalanine-4-hydroxylase (EC 1. 14. 16.1). Comparison of the metabolic reconstruction of *Bt*AOA with that of *B. cereus* ATCC 14579 and other close strains showed that out only *Bt*AOA is involved in aromatic amino acid degradation subsystem with CpsB functioning as phenylalanine-4-hydroxylase. However, some of the closely related strains are part of the pterin-carbinolamine dehydratase subsystem where CpsB also function as Phe4MO. The analyses also showed that Phe4MO is not widely distributed among the isolates (Table 8), which means that Phe4MO is not domicile in the ancestor of Bacillaceae but results of ongoing HGT events. When *Bt*AOA was blasted against *P. mendocina* strain AOA-CPS5 isolated from the same sample where *Bt*AOA was obtained, CpsB was also found in AOA-CPS5, which further stressed that CpsB might have been acquired from the environment and distributed from species to species. The assignment of CpsB into aromatic amino acids and derivatives sub-categories, further buttress the promiscuity of the enzyme and its involvement in xenobiotics degradation including PCP (and its congeners) degradation as reported previously (Teufel et al. 2010; Zhang et al. 2011; Fuchs et al. 2011; Flydal et al. 2012; Flydal and Martinez 2013; Wang et al. 2013). Cluster analysis of CpsB with the closely related neighbourhood whole genomes showed that CpsB from *Bt*AOA is close relatedness with high percentages of amino residues conserved among them (Fig. S3). However, phylogenetic analysis of the protein sequences indicated that CpsB is evolutionarily related (82%) to CpsB from *B. cereus* G9241 (269801), *B. cereus* ATCC 10923 (222523) and *B. cereus* E33L (288681) (Fig. 6). CpsB from *Bt*AOA is also related (49%) to CpsB from *B. cereus* strain B4264 (405532), *B. cereus* ATCC 14579 (226900), *B. weihenstephanensis* KBAB4 (315730), *B. cereus* subsp. Cytotoxis NVH (315749) and that of *P. mendocina* AOA-CPS5 (482703pm).

Table 8: Comparison of the genome of *Bt*AOA with the genome of *B. cereus* ATCC 14579 and other closely related strains to determine the distribution of PCP catabolic enzymes.

Organisms	Similarity Score	*Bt*AOA PCP metabolic enzymes				
		CpsB	CpsD	CpsC	CpsA	CpsE
Bacillus tropicus strain AOA-CPS1		yes	yes	yes	yes	yes
B. cereus ATCC 14579	516	yes	yes	yes	yes	yes
B. cereus str. Ames	494	yes	yes	yes	yes	yes
B. cereus str. Ames Ancestor	477	yes	yes	yes	yes	yes
B. thuringiensis serovar konkukian str. 97-27	465	yes	yes	none	yes	yes
B. licheniformis ATCC 14580	284	none	none	yes	yes	yes
B. subtilis str. *subtilis* 168	250	none	none	yes	yes	yes
B. halodurans C-125	212	none	none	yes	yes	yes
B. clausii KSM-K16	170	none	none	yes	yes	yes
B. amyloliquefaciens FZB42	ND	none	none	yes	yes	yes
B. anthracis str. A1055	ND	yes	yes	yes	yes	yes
B. anthracis str. Australia 94	ND	yes	yes	yes	yes	yes
B. anthracis str. CNEVA-9066	ND	yes	yes	yes	yes	yes
B. anthracis str. Kruger B	ND	yes	yes	yes	yes	yes
B. anthracis str. Sterne	ND	yes	yes	yes	yes	yes
B. anthracis str. Vollum	ND	yes	yes	yes	yes	yes
Bacillus B-14905 (101031.3)	ND	none	none	yes	yes	yes
B. cereus ATCC 10987	ND	yes	yes	yes	yes	yes
B. cereus B4264	ND	yes	yes	yes	yes	yes
B. cereus E33L	ND	yes	yes	yes	yes	yes
B. cereus G9241	ND	yes	yes	yes	yes	yes
B. cereus subsp. cytotoxis NVH 391-98	ND	yes	yes	yes	yes	yes
B. pumilus SAFR-032	ND	none	none	yes	none	yes
B. thuringiensis str. Al Hakam	ND	yes	yes	yes	yes	yes
B. weihenstephanensis KBAB4	ND	yes	yes	yes	yes	yes
Pseudomonas mendocina AOA-CPS5	ND	yes	yes	yes	none	none

Catabolic genes associated with a subsystem in the respective organism (*); subsystem where the gene has been classified into was found to have an active variant in the organism (**). *B. cereus* ATCC 14579 (ATCC 14579), *B. anthracis* str. Ames (Ames), *Bacillus anthracis* str. 'Ames Ancestor (Ames Ancestor), *B. thuringiensis* serovar konkukian str. 97-27 (konkukian), *Bacillus licheniformis* ATCC 14580 (ATCC 14580).

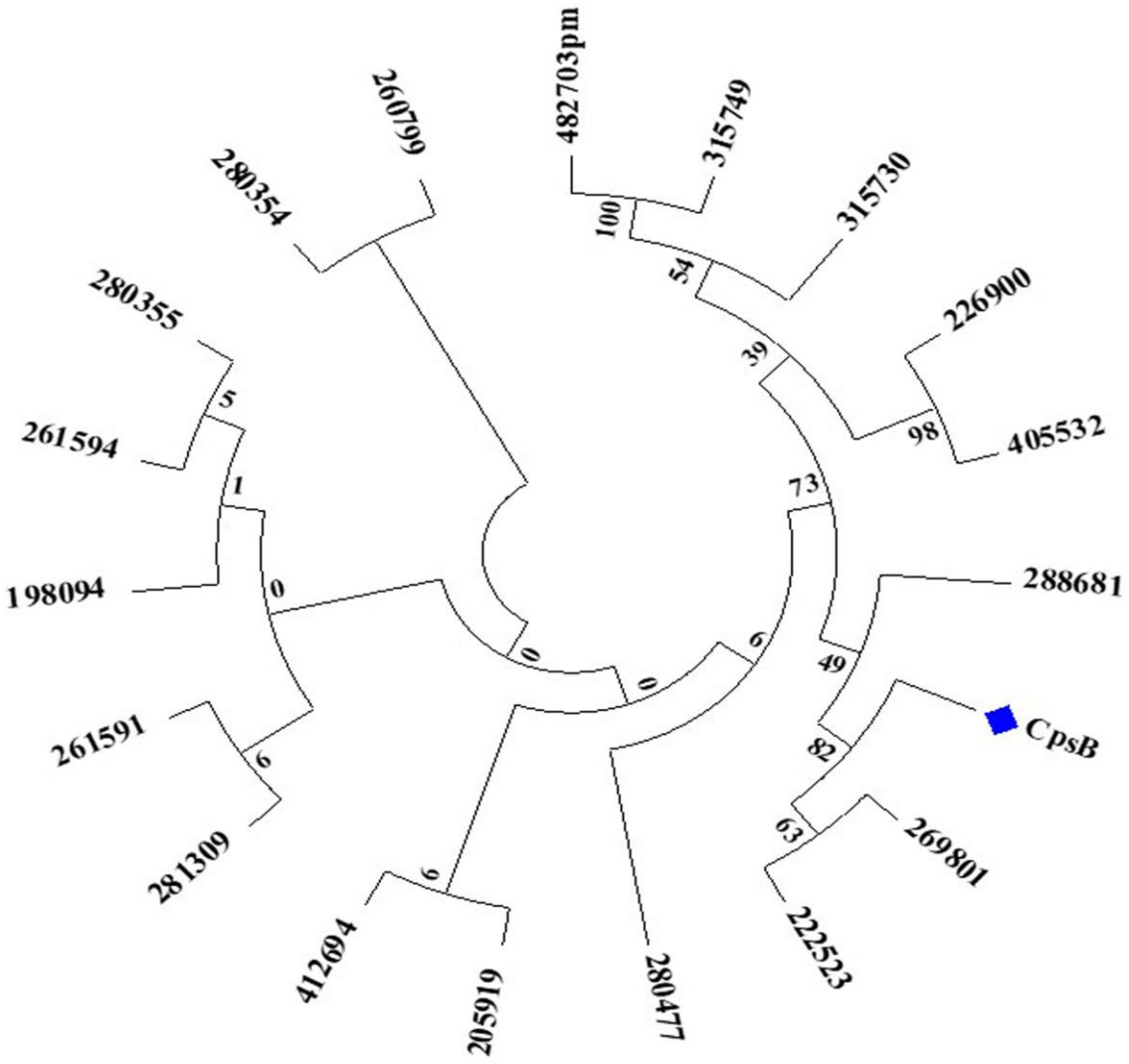


Figure 6: Evolutionary relationships of CpsB with closely related neighbour whole-genome neighbours. Phenylalanine-4-hydroxylase (EC 1.14.16.1) from *Bt*AOA (CpsB), *Pseudomonas mendocina* (482703), *B. anthracis* str. Ames Ancestor (261594), *B. cereus* ATCC 14579 (226900), *B. thuringiensis* serovar konkukian str. 97-27 (281309), *B. anthracis* str. Ames (198094), *B. anthracis* str. A1055 (280355), *B. anthracis* str. Australia 94 (280477), *B. anthracis* str. CNEVA-9066 (280354), *B. anthracis* str. Kruger B (205919), *B. anthracis* str. Sterne (260799), *B. anthracis* str. Vollum (261591), *B. cereus* G9241 (269801), *B. cereus* E33L (288681), *B. cereus* B4264 (405532), *B. cereus* ATCC 10987 (222523), *B. cereus* subsp. cytotoxis NVH 391-98 (315749), *B. thuringiensis* str. Al Hakam (412694) and *B. weihenstephanensis* KBAB4 (315730). The analysis involved 19 amino acid sequences. All positions containing gaps and missing data were eliminated. There was a total of 261 positions in the final dataset.

CpsD is in the categories of cofactors, vitamins, prosthetic groups and pigments, a sub-category of folate and pterines, and in a subsystem of pterin carbinolamine dehydratase (alongside with CpsB) and its assigned functioning role is pterin-4-alpha carbinolamine dehydratase (EC 4.2.1.96). Comparison of the metabolic construction of *Bt*AOA with other strains indicated that CpsD is also assigned the same role in other strains that has CpsB (Table 8). This also buttress our early claim that CpsBD function as a cluster subsystem in which either both enzymes are present, or none is present in the organism. Based on results from our previous experimental work (data not shown) and the arrangement of CpsB and CpsD in the whole genome of *Bt*AOA, it is strongly believed that the pterin-carbinolamine subsystem is recruited into the *Bt*AOA PCP degradation pathway via HGT.

Multiple sequences alignments of CpsD from *Bt*AOA and those from the whole genome of related strain showed that the sequences similarities were not as high as those of CpsD. Also rampant with the sequences are genes deletion overtime that led to the differences in the CpsD molecular weights among the test strains (Fig. S4). However, the residues (His57, His 58 and Pro 59) at the active sites of CpsD (detected in previous study, data not shown) were retained among members. Phylogenetic characterization of CpsD from *Bt*AOA and those from the neighbouring whole-genome indicated that CpsD is only related to CpsD from *B. cereus* ATCC 10987 and the percentage (4%) similarity between the two proteins is extremely low (Fig. 7).

CpsC is present in the genome of most of the organism analysed except *B. thuringiensis* serovar *Konkukian* str. 97-27. CpsC is believed to have been recruited from pre-existing enzyme whose catalytic function is not optimised for any role. Multiple sequences alignments of CpsC from *Bt*AOA and those from neighbourhood's whole-genome showed that the sequences are not similar, and that the percentage similarity is very low (Fig. S5). The fact that the protein is annotated as glutathione-dependent formaldehyde dehydrogenase, alcohol dehydrogenase, sorbitol dehydrogenase homologue, galactitol-1-phosphate 5-dehydrogenase, 2,3-butanediol dehydrogenase (R-alcohol forming), hypothetical proteins, threonine dehydrogenase and related zinc-dependent dehydrogenase in different strains of Bacillus already showed that the proteins are not similar.

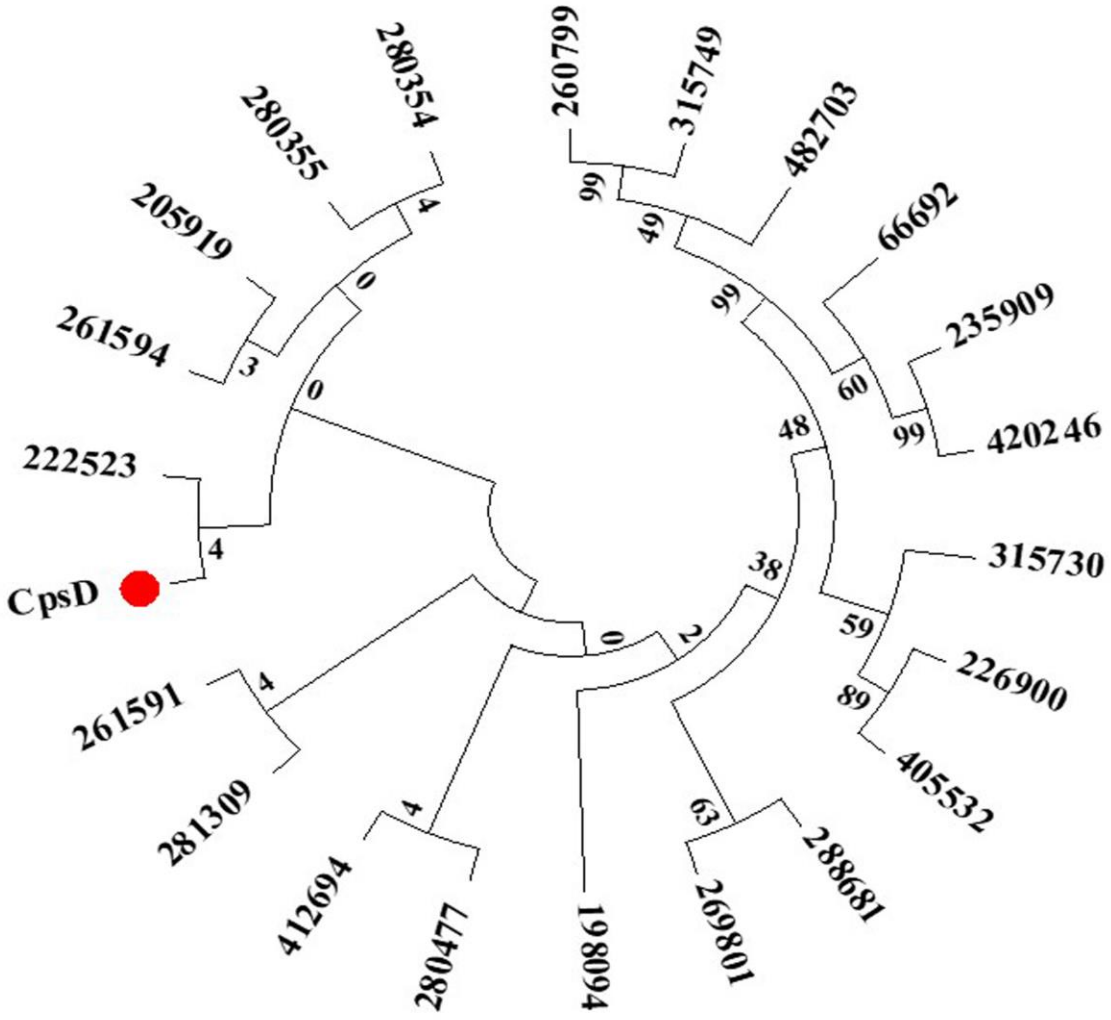


Figure 7: Evolutionary relationships of CpsD with other whole genome neighbours. Pterin-4-alpha-carbinolamine dehydratase (EC 4.2.1.96) from *Bt*AOA (CpsD), *Pseudomonas mendocina* (482703), *B. anthracis* str. Ames Ancestor (261594), *B. cereus* ATCC 14579 (226900), *B. thuringiensis* serovar konkukian str. 97-27 (281309), *B. anthracis* str. Ames (198094), *B. anthracis* str. A1055 (280355), *B. anthracis* str. Australia 94 (280477), *B. anthracis* str. CNEVA-9066 (280354), *B. anthracis* str. Kruger B (205919), *B. anthracis* str. Sterne (260799), *B. anthracis* str. Vollum (261591), *B. cereus* G9241 (269801), *B. cereus* E33L (288681), *B. cereus* B4264 (405532), *B. cereus* ATCC 10987 (222523), *B. cereus* subsp. cytotoxis NVH 391-98 (315749), *B. thuringiensis* str. Al Hakam (412694) and *B. weihenstephanensis* KBAB4 (315730).

Multiple genes deletion and bases substitution occurring at different points in the respective genome of each organism might have led to gain of function or the proteins acquiring the capacity to perform other functions while retaining the original function. Despite the fact that the proteins sequences are not similar, phylogenetic analysis showed that the proteins are all evolutionarily related (Fig. 8), CpsC from *Bt*AOA is closely related to glutathione-dependent formaldehyde dehydrogenase from ***B. cereus*** E33L (288681) and on the same node with those of *B. cereus* ATCC 14579 (226900) and *B. cereus* B4264 (405532).

CpsA is an ancestral family protein of *Bacillus* lineage, present in virtually all the strains except *B. pumilus* SAFR-032. CpsA is also a pre-existing enzyme with no specific function (Table 8) and it is not yet part of a subsystem in *Bt*AOA and other related organisms analysed. However, CpsA homolog is associated with serine glyoxylate cycle (GJO) subsystem in *B. subtilis* str. *subtilis* 168 and *B. anthracis* str. Ames Ancestor. This enzyme was also believed to have been recruited into the PCP pathway via HGT. Cluster analysis of all the CpsA proteins showed that the protein sequences are relatively similar (Fig. S6). Phylogenetic analysis of the cpsA proteins showed that CpsA from *Bt*AOA is most similar to that of *B. cereus* B4264 (405532) and share a node with CpsA from *B. cereus* strains ATCC 14579 (226900), ATCC 10987 (222523) and E33L (288681) respectively (Fig. 9).

Metabolic reconstruction of *Bt*AOA shows that CpsE is in carbohydrates metabolism category and is involved in central carbohydrate and one-carbon metabolism, and it is a part of glyoxylate bypass, TCA cycle and serine-glyoxylate cycle subsystem and its specific role in those metabolic subsystems is malate dehydrogenase (EC 1.1.1.37). The enzyme is present in all *Bacillus* species analysed and it is believed to have been recruited from the TCA cycle or glyoxylate cycle. Multiple sequences alignments of CpsE proteins from the whole-genome neighbourhoods indicated that the protein sequences from different species are not similar (Fig. S7) and the phylogenetic analysis (Fig. 10) also showed that CpsE from *Bt*AOA is only distantly related to other strains.

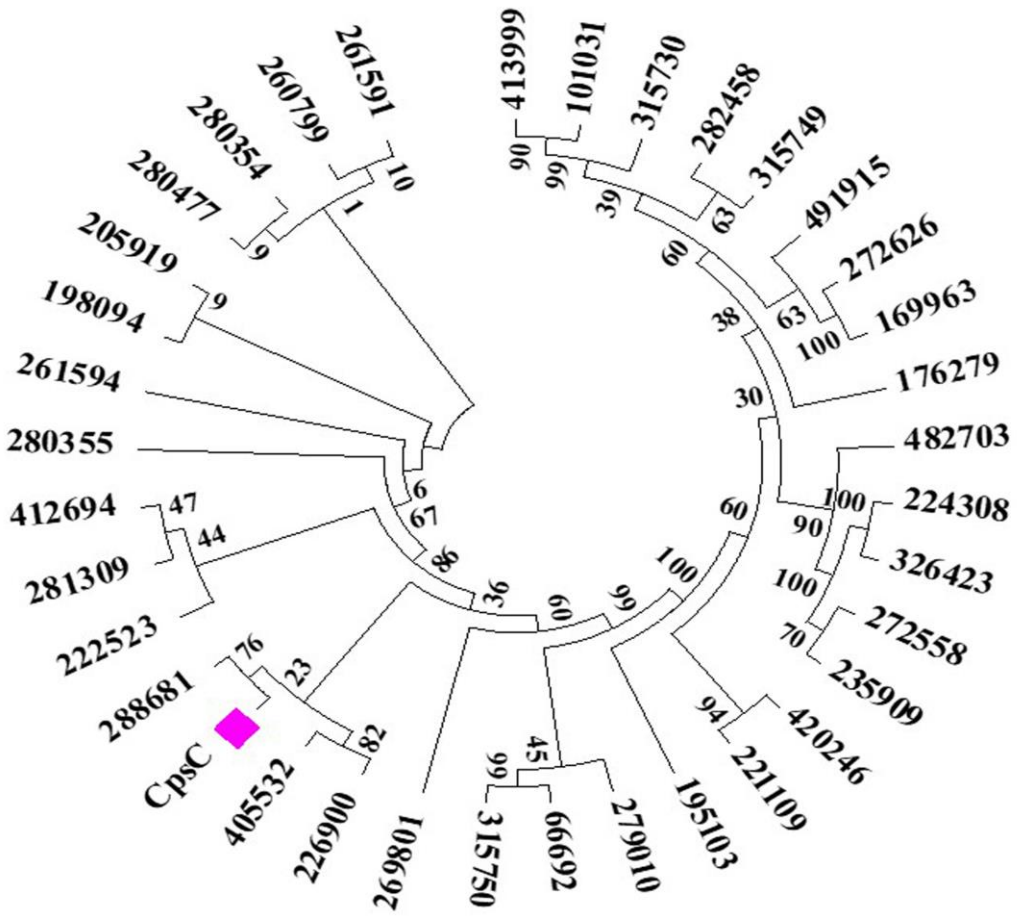


Figure 8: Evolutionary relationships of CpsC with CpsC homologs from closely related strains. L-lactate dehydrogenase (EC 1.1.1.27) from *Clostridium botulinum* A str. ATCC 3502 (413999), *C. perfringens* ATCC 13124 (195103), *Listeria innocua* Clip11262 (272626), *Listeria monocytogenes* EGD-e (169963), *Staphylococcus aureus* subsp. *aureus* MRSA252 (282458) and malate dehydrogenase (EC 1.1.1.37) from *S. epidermidis* RP62A (176279), *B. thuringiensis* serovar konkukian str. 97-27 (281309), *B. subtilis* subsp subtilis str. 168 (224308), *B. cereus* ATCC 14579 (226900), *B. tropicus* AOA-CPS1 (CpsE), *B.* anthracis str. Ames (198094), *B.clausii* KSM-K16 (66692), *B. halodurans* C-125 (272558), *B. licheniformis* ATCC 14580 (279010), *Anoxybacillus flavithermus* WK1 (491915), *Geobacillus kaustophilus* HTA426 (235909), *G. thermodenitrificans* NG80-2 (420246), *Oceanobacillus iheyensis* HTE831 (221109), *B. amyloliquefaciens* FZB42 (326423), *B. anthracis* str. A1055 (280355), *B. anthracis* str. Australia 94 (280477), *B. anthracis* str. CNEVA-9066 (280354), *B. anthracis* str. Kruger B (205919), *B. anthracis* str. Sterne (260799), *B. anthracis* Vollum (261591), *Bacillus* B-14905 (101031), *B. cereus* ATCC 10987 (222523), *B. cereu*s B4264 (405532), *B. cereus* e33L (288681), *B. cereus* G9241 (269801), *B. weihenstephanensis* KBAB4 (315730), *B. thuringiensis* str. Al Hakam (412694), *B. pumilus* SAFR-032 (315750), and *B. cereus* subsp. cytotoxis NVH 391-98 (315749).

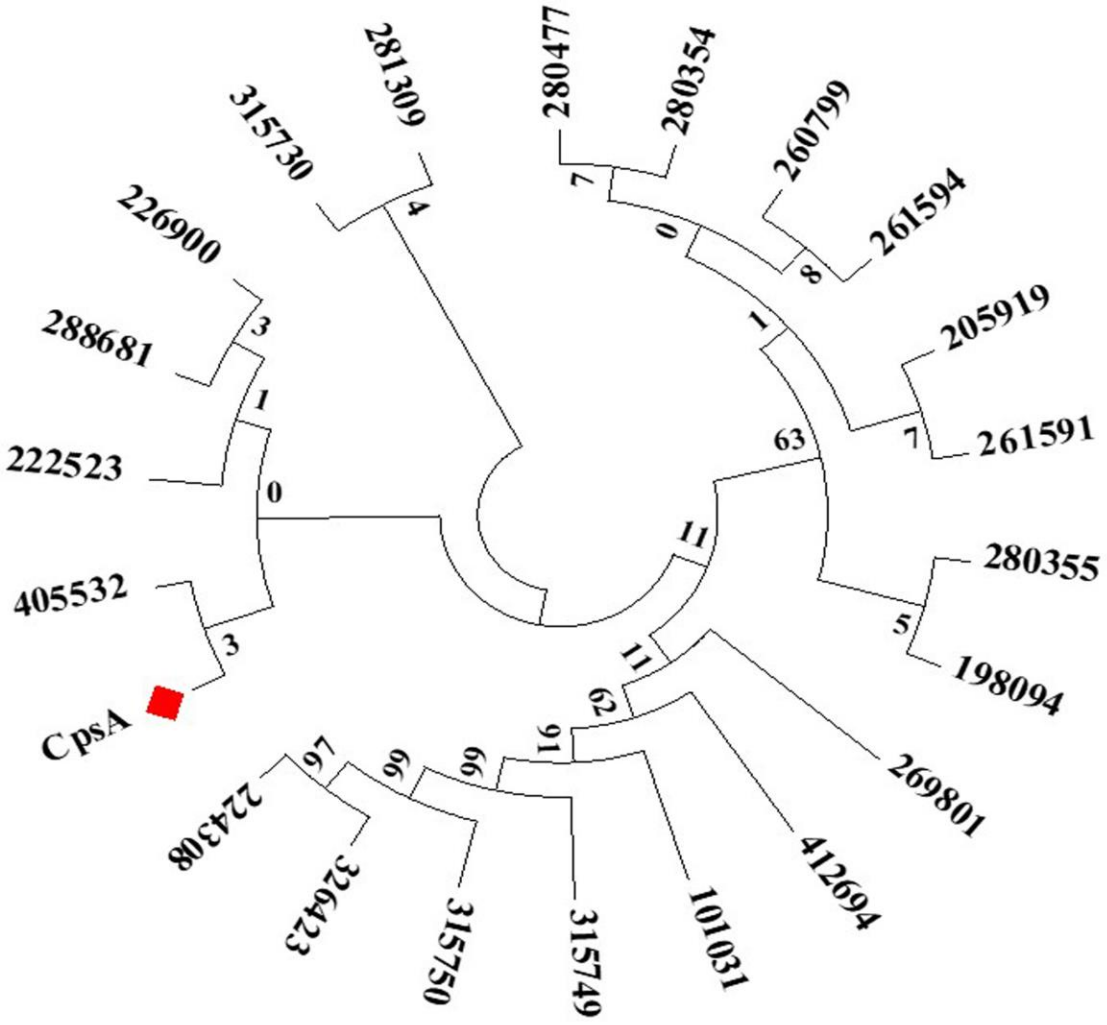


Figure 9: Evolutionary relationships of CpsA with glyoxylase family proteins from closely related strain. Glyoxalase family protein from *Bt*AOA (CpsA), *B. amyloliquefaciens* FZB42 (326423), *B. anthracis* str. A1055 (280355), *B. anthracis* str. Australia 94 (280477), *B. anthracis* str. CNEVA-9066 (280354), *B. anthracis* str. Kruger B (205919), *B. anthracis* str. Sterne (260799), *B. anthracis* str. Vollum (261591), *Bacillus* B-14905 (101031), *B. cereus* ATCC 10987 (222523), *B. cereus* B4264 (405532), *B. cereus* E33L (288681), *B. cereus* G9241 (269801), *B. cereus* subsp. cytotoxis NVH 391-98 (315749), *B. pumilus* SAFR-032 (315750), *B. thuringiensis* str. Al Hakam (412694), *B. weihenstephanensis* KBAB4 (315730), *B. subtilis* subsp. *subtilis* str. 168 (224308), *B. cereus* ATCC 14579 (226900), *B. anthracis* str. Ames (198094), *B. anthracis* str. 'Ames Ancestor (261594), *B. thuringiensis* serovar konkukian str. 97-27 (281309).

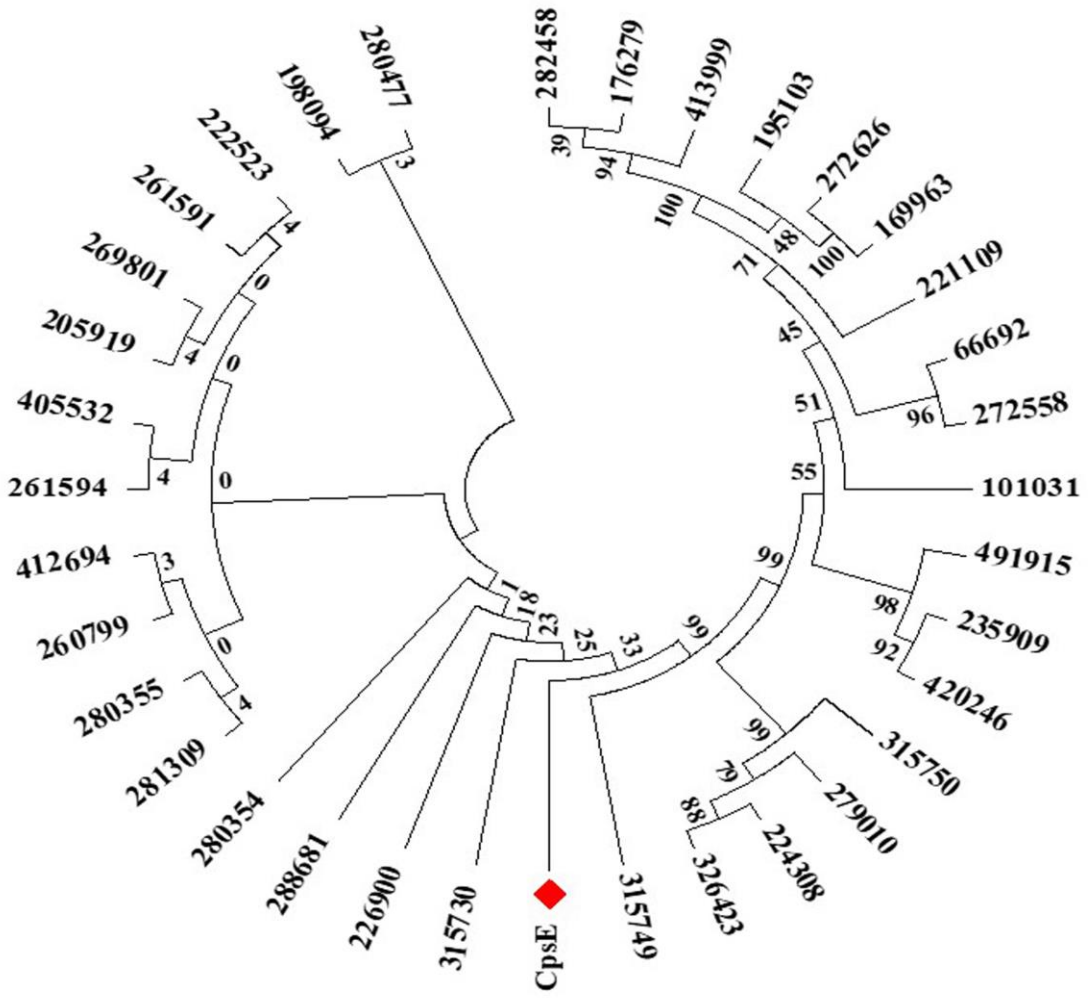


Figure 10: Evolutionary relationships of CpsE with CpsE from closely related neighbour. L-lactate dehydrogenase (EC 1.1.1.27) from *Clostridium botulinum* A str. ATCC 3502 (413999), *C. perfringens* ATCC 13124 (195103), *Listeria innocua* Clip11262 (272626), *Listeria monocytogenes* EGD-e (169963), *Staphylococcus aureus* subsp. *aureus* MRSA252 (282458)and malate dehydrogenase (EC 1.1.1.37) from *Staphylococcus epidermidis* RP62A (176279), *Bacillus thuringiensis* serovar konkukian str. 97-27 (281309), *B. subtilis* subsp subtilis str. 168 (224308), *B. cereus* ATCC 14579 (226900), *B. tropicus* AOA-CPS1 (CpsE), *B.* anthracis str. Ames (198094), *B.clausii* KSM-K16 (66692), *B. halodurans* C-125 (272558), *B. licheniformis* ATCC 14580 (279010), *Anoxybacillus flavithermus* WK1 (491915), *Geobacillus kaustophilus* HTA426 (235909), *G. thermodenitrificans* NG80-2 (420246), *Oceanobacillus iheyensis* HTE831 (221109), *B. amyloliquefaciens* FZB42 (326423), *B. anthracis* str. A1055 (280355), *B. anthracis* str. Australia 94 (280477), *B. anthracis* str. CNEVA-9066 (280354), *B. anthracis* str. Kruger B (205919), *B. anthracis* str. Sterne (260799), *B. anthracis* Vollum (261591), *Bacillus* B-14905 (101031), *B. cereus* ATCC 10987 (222523), *B. cereu*s B4264 (405532), *B. cereus* e33L (288681), *B. cereus* G9241 (269801), *B. weihenstephanensis* KBAB4 (315730), *B. thuringiensis* str. Al Hakam (412694), *B. pumilus* SAFR-032 (315750), and *B. cereus* subsp. cytotoxis NVH 391-98 (315749).

10.3 Conclusion

Characterization, whole-genome sequencing and extensive mining of the whole-genome data (annotated at NCBI, RASTtk and PROKKA) and KEGG metabolic reconstruction of *Bt*AOA showed that this strain harboured genes associated with biodegradation of many chlorophenolic compounds and other xenobiotics. The organism is fortified with thirty-eight stress response subsystem to mitigate the toxicity effects of these compounds. Metabolic reconstruction model of *Bt*AOA clearly showed that the organism has been exposed to different chlorophenolic compounds including the very recalcitrant DDT and atrazine. Hence, its ability to effectively degrade PCP is not surprising. Also, the isolate harboured phages, prophages, transposable elements and plasmids distributed into ten subsystems in the *Bt*AOA metabolic model, these genetic elements are believed to serve as a reservoir of new genes for the organism as the need arise. *Bt*AOA seems to have assembled metabolically efficient enzymes for PCP degradation subsystem via recruitment of existing enzymes and HGT events. The first two enzymes in the pathway (CpsB and CpsD) is an operon recruited into the PCP degradation pathway from existing pterin-carbinolamine metabolism subsystem. CpsC and CpsA were recruited from the pool of hypothetical proteins with no specific role reported previously but shown here to function as dehalogenase and ring-cleaving dioxygenase, respectively, while CpsE was recruited from pre-existing enzymes reported to be involved in the the TCA or serine-glyoxalase cycle via HGT events. Analysis of the metabolic reconstruction of *Bt*AOA showed that both the organisms and its catabolic enzymes have potential application in the bioremediation of various xenobiotics.

10.4 Acknowledgments:

Authors thank University of KwaZulu-Natal for the Ph.D. fellowship.

10.5 Author contributions:

O.A., A.O. and A.K. conceived and designed the project; O.A. and A.K. designed the experiments; O.A. performed the experiments; M.P. contributed reagents and materials; O.A., A.K., M.P. and A.O. wrote the manuscript; all the authors have read and approved the manuscript.

10.6 Funding:

National Research Foundation, South Africa (Grant No: 94036 and 92803).

10.7 Conflict of interest:

All the authors declare no conflict of interest.

10.8 Ethical statement:

This article does not contain any studies with human participants or animals performed by any of the authors.

10.9 References

Almpanis, A., Swain, M., Gatherer, D., & McEwan, N. (2018). Correlation between bacterial G+C content, genome size and the G+C content of associated plasmids and bacteriophages. *Microbial Genomics*, 4(4). https://doi.org/10.1099/mgen.0.000168

Ammeri RW, Mehri I, Badi S, et al (2017) Pentachlorophenol degradation by *Pseudomonas fluorescens*. *Water Quality Research Journal*, 52:99–108. https://doi.org/10.2166/wqrj.2017.003

Arnaldos M, Amerlinck Y, Rehman U, et al (2015) From the affinity constant to the half-saturation index: Understanding conventional modeling concepts in novel wastewater treatment processes. *Water Research*, 70:458–470. https://doi.org/10.1016/j.watres.2014.11.046

ATSDR. (2017). ATSDR (2017) Agency for toxic substances and disease registry. substance priority list (candidates for toxicological profiles). 190. URL: https://www.atsdr.cdc.gov/ (accessed 30-09-2020)

Barber, R. D., Rott, M. A., & Donohue, T. J. (1996). Characterization of a glutathione-dependent formaldehyde dehydrogenase from Rhodobacter sphaeroides. *Journal of Bacteriology*, 178(5): 1386–1393. https://doi.org/10.1128/jb.178.5.1386-1393.1996

Belchik, S. M., & Xun, L. (2012). S-Glutathionyl-(chloro)hydroquinone reductases: A new class of glutathione transferases functioning as oxidoreductases. *Drug Metabolism Reviews*, 43(2): 307–316. https://doi.org/10.3109/03602532.2011.552909

Biosciences, P. (2019). *Sequel II System: Delivering Highly Accurate Long Reads Gain: a Comprehensive View of Genomes, Transcriptomes, and Epigenomes*. URL: http://www.pacb.com/legal-and-trademarks/terms-and-conditions-of-sale/. (accessed 30-09-2020)

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., … Xia, F. (2015). RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Scientific Reports*, 5. https://doi.org/10.1038/srep08365

Cai M, Xun L (2002) Organization and regulation of pentachloro- phenol degrading genes in *Sphingobium chlorophenolicum* ATCC 39723. *Journal of Bacteriology* 184:4672–4680

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, 10:421. https://doi.org/10.1186/1471-2105-10-421

Chang, A. Y., Chau, V. W. Y., Landas, J. A., & Pang, Y. (2017). Preparation of calcium competent *Escherichia coli* and heat-shock transformation. *JEMI Methods*, 1:22–25.

Chebrou, H., Hurtubise, Y., Barriault, D., & Sylvestre, M. (1999). Heterologous expression and characterization of the purified Oxygenase component of *Rhodococcus globerulus* P6 biphenyl Dioxygenase and of chimeras derived from it. *Journal of Bacteriology*, 181(16): 4805–4811.

Ciufo, S., Kannan, S., Sharma, S., Badretdin, A., Clark, K., Turner, S., … DiCuccio, M. (2018). Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. *International Journal of Systematic and Evolutionary Microbiology*, *68*(7), 2386–2392. https://doi.org/10.1099/ijsem.0.002809

Cohen, N. R., Lobritz, M. A., & Collins, J. J. (2013). Microbial persistence and the road to drug resistance. *Cell Host and Microbe Review*. https://doi.org/10.1016/j.chom.2013.05.009

Copley, S. D. (2010). Evolution of Efficient Pathways for Degradation of Anthropogenic Chemicals. *PMC*, 5(8): 559–566. https://doi.org/10.1038/nchembio.197.

Copley, S. D., Rokicki, J., Turner, P., Daligault, H., Nolan, M., & Land, M. (2012). The whole genome sequence of *Sphingobium chlorophenolicum* L-1: Insights into the evolution of the pentachlorophenol degradation Pathway. *Genome Biology and Evolution*, 4(2): 184–198. https://doi.org/10.1093/gbe/evr137

Crickard, J. B., & Greene, E. C. (2019). Helicase mechanisms during homologous recombination in *Saccharomyces cerevisiae* . *Annual Review of Biophysics*, *48*(1): 255–273. https://doi.org/10.1146/annurev-biophys-052118-115418

Dai, M. H., & Copley, S. D. (2004). Genome shuffling improves degradation of the anthropogenic pesticide pentachlorophenol by *Sphingobium chlorophenolicum* ATCC 39723. *Applied and Environmental Microbiology,* 70(4): 2391–2397. https://doi.org/10.1128/AEM.70.4.2391-2397.2004

Dunning, L. T., Olofsson, J. K., Parisod, C., Choudhury, R. R., Moreno-Villena, J. J., Yang, Y., … Christin, P. A. (2019). Lateral transfers of large DNA fragments spread functional genes among grasses. *Proceedings of the National Academy of Sciences of the United States of America,* 116(10), 4416–4425. https://doi.org/10.1073/pnas.1810031116

Durruty I, Okada E, González JF, Murialdo SE (2011) Degradation of chlorophenol mixtures in a fed-batch system by two soil bacteria. *Water SA*, 37:547–552.

El-Bialy HA, Khalil OAA, Gomaa OM (2018) Bacterial-mediated biodegradation of pentachlorophenol via electron shuttling. *Environmental Technology,* 40:1–11. https://doi.org/10.1080/09593330.2018.1442501

El-naas, M. H., Mousa, H. A., & Gamal, M. El. (2017). *Microbe-Induced Degradation of Pesticides*. (S.N. Singh (ed.), Ed.), *Environmental Science and Engineering*. Springer International Publishing Switzerland. https://doi.org/10.1007/978-3-319-45156-5

Engevik, M. A., Morra, C. N., Röth, D., Engevik, K., Spinler, J. K., Devaraj, S., … Versalovic, J. (2019). Microbial metabolic capacity for intestinal folate production and modulation of host folate receptors. *Frontiers in Microbiology,* 10. https://doi.org/10.3389/fmicb.2019.02305

Federhen, S., Rossello-Mora, R., Klenk, H., Tindall, B. J., Konstantinidis, K.T., Whitman, W. B. et al. (2016). Meeting report: GenBank microbial genomic taxonomy workshop (12-13 May, 2015). *Standards in Genomic Sciences*, 11(15): 1–8. https://doi.org/10.1186/s40793-016-0134-1

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, 39: 783–791.

Flydal, M. I., & Martinez, A. (2013). Phenylalanine hydroxylase: Function, structure, and regulation. *IUBMB Life*, 65(4): 341–349. https://doi.org/10.1002/iub.1150

Flydal, M. I., Chatfield, C. H., Zheng, H., Gunderson, F. F., Aubi, O., Cianciotto, N. P., &

Martinez, A. (2012). Phenylalanine Hydroxylase from *Legionella pneumophila* is a thermostable enzyme with a major functional role in pyomelanin synthesis. *PLoS ONE*, 7(9): e46209. https://doi.org/10.1371/journal.pone.0046209

Flydal, M. I., Alcorlo-Pagés, M., Johannessen, F. G., Martínez-Caballero, S., Skjærven, L., Fernandez-Leiro, R., Martinez, A. & Hermoso, J. A. (2019). Structure of full-length human phenylalanine hydroxylase in complex with tetrahydrobiopterin. *Proceedings of the National Academy of Sciences of the United States of America*, 166(23): 11229–11234. https://doi.org/10.1073/pnas.1902639116

Fuchs, G., Boll, M., & Heider, J. (2011). Microbial degradation of aromatic compounds- From one strategy to four. *Nature Reviews Microbiology*, 9: 803–816. https://doi.org/10.1038/nrmicro2652

Grant, J. R., Stothard, P. (2008) The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Res 36:W181-4. https://doi.org/10.1093/nar/gkn179

Greule, A., Stok, J. E., De Voss, J. J., & Cryle, M. J. (2018). Unrivalled diversity: the many roles and reactions of bacterial cytochromes P450 in secondary metabolism. *Natural Product Reports*, 35: 757–791. https://doi.org/10.1039/c7np00063d

Haft, D. H., DiCuccio, M., Badretdin, A., Brover, V., Chetvernin, V., O'Neill, K. at al. (2018). RefSeq: An update on prokaryotic genome annotation and curation. *Nucleic Acids Research*, 46(D1): D851–D860. https://doi.org/10.1093/nar/gkx1068

Hannenhalli, S. S., & Russell, R. B. (2000). Analysis and prediction of functional sub-types from protein sequence alignments. *Journal of Molecular Biology*, 303(1): 61–76. https://doi.org/10.1006/jmbi.2000.4036

Hayes, R. P., Green, A. R., Nissen, M. S., Lewis, K. M., Xun, L., & Kang, C. (2013). Structural characterization of 2,6-dichloro-p-hydroquinone 1,2-dioxygenase (PcpA) from *Sphingobium chlorophenolicum*, a new type of aromatic ring-cleavage enzyme. *Molecular Microbiology*, 88(3): 523–536. https://doi.org/10.1111/mmi.12204

Hidalgo, F. J., & Zamora, R. (2019). Formation of phenylacetic acid and benzaldehyde by degradation of phenylalanine in the presence of lipid hydroperoxides: New routes in the amino acid degradation pathways initiated by lipid oxidation products. *Food Chemistry: X*, 2. https://doi.org/10.1016/j.fochx.2019.100037

Hirose, I., Sano, K., Shioda, I., Kumano, M., Nakamura, K., & Yamane, K. (2000). Proteome

analysis of *Bacillus subtilis* extracellular proteins: a two-dimensional protein electrophoretic study. *Microbiology*, 146(Pt 1),:5–75. https://doi.org/10.1099/00221287-146-1-65

Hlouchova, K., Rudolph, J., Pietari, J. M. ., Behlen, L. S., & Copley, S. D. (2012). Pentachlorophenol hydroxylase, a poorly functioning enzyme required for degradation of pentachlorophenol by *Sphingobium chlorophenolicum*. *Biochemistry*, 51(18): 3848–3860. https://doi.org/10.1021/bi300261p

Huang, W. M., DaGloria, J., Fox, H., Ruan, Q., Tillou, J., Shi, K., … Casjens, S. (2012). Linear chromosome-generating system of agrobacterium tumefaciens C58: Protelomerase generates and protects hairpin ends. *Journal of Biological Chemistry*, 287(30): 25551–25563. https://doi.org/10.1074/jbc.M112.369488

Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, 11. https://doi.org/10.1186/1471-2105-11-119

IARC. (2019a). International Agency for Research on Cancer monographs on the identification of carcinogenic hazards to humans. In: Agents Classified by the IARC Monographs, Volumes 1–127. URL: https://monographs.iarc.fr/agents-classified-by-the-iarc/ (accessed 30-09-2020).

IARC. (2019b). Pentachlorophenol and some related compounds. IARC monographs on the evaluation of carcinogenic risks to humans. URL: https://publications.iarc.fr/Book-And-Report-Series/Iarc-Monographs-On-TheIdentification-Of-Carcinogenic-Hazards-To-Humans/Pentachlorophenol-And-Some-Related-Compounds-2019 (accessed 30-09-2020)

Igbinosa, E. O., Odjadjare, E. E., Chigor, V. N., Igbinosa, I. H., Emoghene, A. O., Ekhaise, F. O., Igiehon, N. O. & Idemudia, O. G. (2013). Toxicological profile of chlorophenols and their derivatives in the environment: the public health perspective. *The Scientific World Journal*, 2013: 1–11. https://doi.org/10.1155/2013/460215

Jin, S., De Jesús-Berríos, M., & Sonenshein, A. L. (1996). A *Bacillus subtilis* Malate dehydrogenase gene. *Journal of Bacteriology*, 178(2): 560–563. https://doi.org/10.1128/jb.178.2.560-563.1996

Khessairi A, Fhoula I, Jaouani A, et al (2014) Pentachlorophenol degradation by *Janibacter* sp., a new *Actinobacterium* isolated from saline sediment of arid land. Biomed Research International 2014:1–9. https://doi.org/http://dx.doi.org/10.1155/2014/296472

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S. et al. (2019). PubChem 2019 update: improved access to chemical data. *Nucleic Acids Research,* 47(D1):D1102-1109. doi:10.1093/nar/gky1033

Kingan, S. B., Scientist, B., & Applications, P. (2018). Assembly and annotation of diploid and polyploid genomes with PacBio San Diego Botanical Garden. URL: https://pb-falcon.readthedocs.io/en/latest/_downloads/KinganPacBio_ToolsForPolyploidsPAG2018.pdf. (accessed 30-09-2020)

Kok, D. E., Steegenga, W. T., Smid, E. J., Zoetendal, E. G., Ulrich, C. M., & Kampman, E. (2020). Bacterial folate biosynthesis and colorectal cancer risk: more than just a gut feeling. *Critical Reviews in Food Science and Nutrition*, 60(2): 244–256. https://doi.org/10.1080/10408398.2018.1522499

Kumar, A., Khan, F.I. & Olaniran, A.O. (2018) Chloroacetaldehyde dehydrogenase from *Ancylobacter aquaticus* UV5: Cloning, expression, characterization and molecular modeling. *International Journal of Biological Macromolecules,* 114:1117–1126. https://doi.org/https://doi.org/10.1016/j.ijbiomac.2018.03.176

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, 33(7): 1870–1874. https://doi.org/10.1093/molbev/msw054

Li D, Park J, Oh JR (2001) Silyl derivatization of alkylphenols, chlorophenols, and bisphenol A for simultaneous GC/MS determination. *Analytical Chemistry,* 73:3089–3095.

Litwack, G. (2018). Metabolism of amino acids. In *Human Biochemistry* (pp. 359–394). Elsevier. https://doi.org/10.1016/b978-0-12-383864-3.00013-2

Lopez-Echartea, E., Macek, T., Demnerova, K., & Uhlik, O. (2016). Bacterial biotransformation of pentachlorophenol and micropollutants formed during its production process. *International Journal of Environmental Research and Public Health*, 13(1146): 1–21. https://doi.org/10.3390/ijerph13111146

Lv Y, Chen Y, Song W, Hu Y (2014) Enhanced selection of micro-aerobic pentachlorophenol degrading granular sludge. *Journal of Hazardus Materials,* 280:134–142. https://doi.org/10.1016/j.jhazmat.2014.07.067

Machonkin, T. E., Holland, P. L., Smith, K, N., Liberman, J. . S., Dinescu, A., Cundari, T. R., & Rocks, S. . S. (2010). Determination of the active site of *Sphingobium chlorophenolicum*

2,6-dichlorohydroquinone dioxygenase (PcpA). *Journal of Biological Inorganic Chemistry*, 15: 291–301.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N. et al. (2019). The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research*, 47(W1): W636–W641. https://doi.org/10.1093/nar/gkz268

Marchesi, J. R., Sato, T., Weightman, A. J., Martin, T. A., Fry, J. C., Hiom, S. J., & Wade, W. G. (1998). Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. *Applied and Environmental Microbiology*, 64:795–799.

Mitchell, S. C., & Steventon, G. B. (2019). Phenylalanine 4-monooxygenase: the "sulfoxidation polymorphism." *Xenobiotica*, 1–13. https://doi.org/10.1080/00498254.2019.1636419

Moine-Scientist, D. I., & Applications Support, F. (2019). *The Sequel II System An introduction to Single Molecule Real-Time (SMRT) Sequencing and its applications*.

Naponelli, V., Noiriel, A., Ziemak, M. J., Beverley, S. M., Lye, L. F., Plume, A. M. et al. (2008). Phylogenomic and functional analysis of pterin-4α-carbinolamine dehydratase family (COG2154) proteins in plants and microorganisms. *Plant Physiology*, 146(4): 1515–1527. https://doi.org/10.1104/pp.107.114090

Nguyen, K. T., Virus, C., Günnewich, N. et al. (2012) Changing the regioselectivity of a P450 from C15 to C11 hydroxylation of progesterone. *ChemBioChem*, 13:1161–1166. https://doi.org/10.1002/cbic.201100811

Niesler, M., & Surmacz-Górska, J. (2018). Pentachlorophenol degradation by activated sludge with phenol and glucose as growth substrates. *Archives of Environmental Protection*, 44(3):31–41. https://doi.org/10.24425/123360

Parthasarathy, A., Cross, P. J., Dobson, R. C. J., Adams, L. E., Savka, M. A., & Hudson, A. O. (2018). A three-ring circus: Metabolism of the three proteogenic aromatic amino acids and their role in the health of plants and animals. *Frontiers in Molecular Biosciences*. 5. https://doi.org/10.3389/fmolb.2018.00029

Patel, B., Kumar, A. (2016a) Multi-substrate biodegradation of chlorophenols by defined microbial consortium. *3 Biotech*, 6:191. https://doi.org/10.1007/s13205-016-0511-x

Patel, B., Kumar, A. (2016b) Optimization study for maximizing 2,4-dichlorophenol degradation by *Kocuria rhizophila* strain using response surface methodology and kinetic study. *Desalination and Water Treatment*, 57:18314–18325.

https://doi.org/10.1080/19443994.2015.1091988

Raza, A., Rashid, N., Basheer, S., Aziz, I., & Akhtar, M. (2017). Glutathione-dependent Formaldehyde dehydrogenase homolog from *Bacillus subtilis* Strain R5 is a propanol-preferring Alcohol dehydrogenase. *Biochemistry (Moscow)*, 82(1): 82–23. https://doi.org/10.1134/S0006297917010023

Richardson, E. J., & Watson, M. (2013). The automatic annotation of bacterial genomes. *Briefings in Bioinformatics*, 14(1): 1–12. https://doi.org/10.1093/bib/bbs007

Roca, A., Rodríguez-Herva, J. J., & Ramos, J. L. (2009). Redundancy of enzymes for formaldehyde detoxification in *Pseudomonas putida*. *Journal of Bacteriology*, 191(10: 3367–3374. https://doi.org/10.1128/JB.00076-09

Rossoni, A. W., Price, D. C., Seger, M., Lyska, D., Lammers, P., Bhattacharya, D., & Weber, A. P. M. (2019). The genomes of polyextremophilic cyanidiales contain 1% horizontally transferred genes with diverse adaptive functions. *ELife*, 8. https://doi.org/10.7554/eLife.45017

Saber, D. L., & Crawford, R. L. (1985). Isolation and characterization of *Flavobacterium* strains that degrade pentachlorophenol. *Applied and Environmental Microbiology*, 50(6): 1512–1518.

Zhang, W. & Sun, Z. (2008) Random local neighbor joining: A new method for reconstructing phylogenetic trees. *Molecular Phylogenetic Evolution*, 47(1):117-128. https://doi.org/10.1016/j.ympev.2008.01.019

Sambrook J, Russell DW, Laboratory. CSH (2012) Molecular cloning : a laboratory manual / Joseph Sambrook, David W. Russell, 4th. ed. Cold Spring Harbor Laboratory Cold Spring Harbor, N.Y

Sanghani, P. C., Bosron, W. F., & Hurley, T. D. (2002). Human glutathione-dependent Formaldehyde dehydrogenase. Structural changes associated with ternary complex formation. *Biochemistry*, 41(51): 15189–15194. https://doi.org/10.1021/bi026705q

Seemann, T. (2014) Prokka: Rapid prokaryotic genome annotation. *Bioinformatics*, 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Setlhare, B., Kumar, A., Mokoena, M.P. & Olaniran A.O. (2020) Phenol hydroxylase from *Pseudomonas* sp. KZNSA: Purification, characterization and prediction of three-dimensional structure. *Internationa Journal of Biological Macromolecules*, 146:1000–1008.

https://doi.org/10.1016/j.ijbiomac.2019.09.224

Sharma, A., Thakur, I.S., Dureja, P. (2009) Enrichment, isolation and characterization of pentachlorophenol degrading bacterium *Acinetobacter* sp. ISTPCP-3 from effluent discharge site. *Biodegradation*, 20:643–650. https://doi.org/10.1007/s10532-009-9251-5

Slater, S. C., Goldman, B. S., Goodner, B., et al. (2009). Genome sequences of three *Agrobacterium Biovars* help elucidate the evolution of multichromosome genomes in bacteria. *Journal of Bacteriology*, 191(8): 2501–2511. https://doi.org/10.1128/JB.01779-08

Smith, R. M. (2003) Before the injection--modern methods of sample preparation for separation techniques. J Chromatogr A 1000:3–27

Steventon, G. B., & Mitchell, S. C. (2009). Phenylalanine 4-monooxygenase and the role of endobiotic metabolism enzymes in xenobiotic biotransformation. *Expert Opinion on Drug Metabolism and Toxicology*, 5:1213–1221. https://doi.org/10.1517/17425250903179318

Stockholm convention, 2019. Stockholm convention on persistent organic pollutants. (UNEP/POPS/COP.9/INF/16). Geneva, Switzerland. URL: http://chm.pops.int/TheConvention/ConferenceoftheParties/Meetings/COP9/tabid/7521/ctl/Download/mid/20310/Default.aspx?id=5&ObjID=27106 (accessed 30-09-2020)

Su, Yunyou., Chen, Lifeng., Bandy, Brian., & Yang, Jian. (2008). The catalytic product of Pentachlorophenol 4-monooxygenase is tetra-chlorohydroquinone rather than tetrachlorobenzoquinone. *The Open Microbiology Journal*, 2:100. https://doi.org/10.2174/1874285800802010100

Tanaka, K., Kusano, S., Fujita, N., Ishihama, A., & Takahashi, H. (1995). Promoter determinants for *Escherichia coli* RNA polymerase holoenzyme containing σ38 (the rpoS gene product). *Nucleic Acids Research*, 23(5): 827–834. https://doi.org/10.1093/nar/23.5.827

Tatusova, Tatiana., Dicuccio, Michael., Badretdin, Azat., Chetvernin, Vyacheslav., Nawrocki, Eric P., Zaslavsky, Leonid., … Ostell, James. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Research*, 44(14): 6614–6624. https://doi.org/10.1093/nar/gkw569

Teufel, A. I., Johnson, M. M., Laurent, J. M., Kachroo, A. H., Marcotte, E. M., & Wilke, C. O. (2019). The many nuanced evolutionary consequences of duplicated genes. *Molecular Biology and Evolution*, 36(2): 304–314. https://doi.org/10.1093/molbev/msy210

Teufel, R., Mascaraque, V., Ismail, W., Voss, M., Perera, J., Eisenreich, W., … Fuchs, G. (2010). Bacterial phenylalanine and phenylacetate catabolic pathway revealed. *Proceedings*

of the National Academy of Sciences of the United States of America, 107(32): 14390–14395. https://doi.org/10.1073/pnas.1005399107

Uotila, J.S., Kitunen, V.H., Coote, T et al (1995) Metabolism of halohydroquinones in *Rhodococcus chlorophenolicus* PCP-1. Biodegradation 6:119–126. https://doi.org/10.1007/BF00695342

Van Den Hemel, D., Brigé, A., Savvides, S. N., & Van Beeumen, J. (2006). Ligand-induced conformational changes in the capping subdomain of a bacterial old yellow enzyme homologue and conserved sequence fingerprints provide new insights into substrate binding. *Journal of Biological Chemistry*, 281(38): 28152–28161. https://doi.org/10.1074/jbc.M603946200

Van Dillewijn, P., Wittich, R. M., Caballero, A., & Ramos, J. L. (2008). Type II hydride transferases from different microorganisms yield nitrite and diarylamines from polynitroaromatic compounds. *Applied and Environmental Microbiology*, 74(21): 6820–6823. https://doi.org/10.1128/AEM.00388-08

Verbrugge, L. A., Kahn, L., & Morton, J. M. (2018). Pentachlorophenol, polychlorinated dibenzo-p-dioxins and polychlorinated dibenzo furans in surface soil surrounding pentachlorophenol-treated utility poles on the Kenai National Wildlife Refuge, Alaska USA. *Environmental Science and Pollution Research International*, 25(19): 19187–19195. https://doi.org/10.1007/s11356-018-2269-7

Villemur R (2013) The pentachlorophenol-dehalogenating *Desulfitobacterium hafniense* strain PCP-1. *Philosophical Transactions B Royal Society London: Biological Sciences*, 368:20120319. https://doi.org/10.1098/rstb.2012.0319

Wang, D., Coco, M. W., & Rose, R. B. (2015). Interactions with the bifunctional interface of the transcriptional coactivator DCoH1 are kinetically regulated. *Journal of Biological Chemistry*, 290(7): 4319–4329. https://doi.org/10.1074/jbc.M114.616870

Wang, H., Chen, H., Hao, G., Yang, B., Feng, Y., Wang, Y., et al. (2013). Role of the phenylalanine-hydroxylating system in aromatic substance degradation and lipid metabolism in the oleaginous Fungus Mortierella alpina. *Applied and Environmental Microbiology*, 79(10): 3225–3233. https://doi.org/10.1128/AEM.00238-13

Wickell, D. A., & Li, F. (2020). On the evolutionary significance of horizontal gene transfers in plants. *New Phytologist*, 225(1): 113–117. https://doi.org/10.1111/nph.16022

Xu, Y., He, Y., Egidi, E., Franks, A. E., Tang, C., & Xu, J. (2019). Pentachlorophenol alters the acetate-assimilating microbial community and redox cycling in anoxic soils. *Soil Biology and Biochemistry*, 131:133–140. https://doi.org/10.1016/j.soilbio.2019.01.008

Yadid, I., Rudolph, J., Hlouchova, K., & Copley, S. D. (2013). Sequestration of a highly reactive intermediate in an evolving pathway for degradation of pentachlorophenol. *Proceedings of the National Academy of Sciences of the United States of America*, 110(24): E2182-90. https://doi.org/10.1073/pnas.1214052110

Yang, C.F., Lee, C.M., Wang, C.C. (2006) Isolation and physiological characterization of the pentachlorophenol degrading bacterium *Sphingomonas chlorophenolica*. Chemosphere 62:709–714. https://doi.org/10.1016/j.chemosphere.2005.05.012

Yang, Z., Wafula, E. K., Kim, G., Shahid, S., McNeal, J. R., Ralph, P. E. et al. (2019). Convergent horizontal gene transfer and cross-talk of mobile nucleic acids in parasitic plants. *Nature Plants*, 5(9): 991–1001. https://doi.org/10.1038/s41477-019-0458-0

Zhang, W., Ames, B. D., & Walsh, C. T. (2011). Identification of Phenylalanine 3-hydroxylase for *meta* -Tyrosine biosynthesis. *Biochemistry*, 50(24): 5401–5403. https://doi.org/10.1021/bi200733c

SUPPLEMENTARY MATERIALS

Table S1: Parameters used for HGAP4 analysis

Parameters	Value
Aggressive option	false
Genome Length	5000000
Seed coverage	30
Seed length cutoff	-1
Save Output for Unzip	false
Algorithm	best
Masking	true
Minimum confidence	40
Minimum coverage	5
Consolidate.bam	false
Number of .bam files	1
Filters to add to the Data Set	rq >= 0.7
Output unaligned.bam	false
Minimum sub-read length	0
Sample Name	
Process HQ Regions	false
Pick One Read per ZMW of Median Length	false
Minimum Concordance (%)	70
Minimum Length (bp)	50
Memory per thread for sorting	4G
Split by Sample	false
Strip Base Tags	false
Override Options	
Process ZMW Reads	false

Table S2: Retention time(s) and mass spectra of derivatized (TMS) and underivatized metabolites of pentachlorophenol biodegradation by *Bacillus tropicus* strain AOA-CPS1

S/N	Compounds	Ret. time(s) (min)	Mass spectrum *m/z* (relative intensity)
1	2,4-Dimethylbenzaldehyde	8.480	133.10 (100); 118.90 (0.027); 105.00 (63.92); 91.0 (18.72); 77.05 (49.31); 62.90 (7.05); 51.05 (8.66)
2	3,5-Dimethylbenzaldehyde	8.480	133.10 (100); 118.90 (0.027); 105.00 (63.92); 91.0 (18.72); 77.05 (49.31); 62.90 (7.05); 51.05 (8.66)
3	2,5-Dimethylbenzaldehyde	8.480	143.10 (89.63); 118.90 (0.027);105.00 (63.92); 91.0 (18.72); 77.05 (49.31); 62.90 (7.05); 51.05 (8.66)
4	4-Ethylbenzaldehyde	8.480	143.10 (89.63); 118.90 (0.027); 105.00 (63.92); 91.0 (18.72); 77.05 (49.31); 62.90 (7.05); 51.05 (8.66)
5	Benzene-1,4-bis(1,1-dimethylethyl)	9.145	190.10 (2.39); 175.10 (100); 159.90 (0.62); 145.05 (1.92); 128.05 (3.91); 105.10 (4.39); 91.05 (9.33); 77.00 (3.80); 57.10 (54.38); 51.05 (1.25)
6	1-tert-Butyl-3-isopropyl-5-methylbenzene	9.145	190.10 (2.39); 175.10 (100); 147.10 (8.42); 133.00 (1.74); 119.10 (4.39); 105.10 (4.39); 91.05 (9.33); 80.10 (4.55); 57.10 (54.38)
7	1,3-Dimethyl-4,6-diisopropylbenzene	9.145	190.10 (2.39); 175.10 (100); 147.10 (8.42); 133.00 (1.74); 119.10 (4.39); 105.10 (4.39); 91.05 (9.33); 77.95 (2.13); 65.00 (8.87)
8	Trimethylsilyl phenylacetate	10.225	208.10 (0.40); 192.95 (8.46); 177.10 (0.26); 164.05 (15.88); 137.10 (4.39); 105.10 (0.27); 90.95 (13.00); 73.05 (100.00); 65.00 (8.14)
9	1-Methoxy-5-trimetylsilyoxyhexane	10.225	117.05 (31.78); 106.90 (0.44); 89.05 (12.82); 73.05 (100); 58.95 (1.80)
10	Methyl 2-hydroxyl-3-methylbutanoate	10.225	90.00 (16.45); 73.05 (100.00); 58.95 (1.80)
11	Trimethylsilyl 2-butoxyacetate	10.225	161.10 (0.06); 145.10 (0.10); 117.05 (31.78); 103.00 (0.17); 89.05 (12.82); 73.05 (100.00); 57.05 (21.22)
12	2,4-Di-tert-butylphenol	15.310	206.05 (16.65); 191.10 (100); 163.05 (7.34); 135.10 (2.8); 107.05 (5); 91.05 (5.33); 74.05 (7.93); 57.10 (31.97); 50.90 (1.48)
13	2,6-Bis(tert-butyl) phenol	15.310	206.05 (16.65); 191.10 (100.00); 163.05 (7.34); 147.00 (3.40); 131.00 (1.47); 105.15 (2.71); 91.05 (5.33); 74.05 (7.93); 57.10 (31.97); 55.10 (1.97)
14	3,5-Di-t-butylphenol	15.310	206.05 (16.65); 191.10 (100); 163.05 (7.34); 135.10 (2.8); 107.05 (5.00); 91.05 (5.33); 74.05 (7.93); 57.10 (31.97); 55.10 (1.97)
15	2,5-bis(1,1-Dimethylethyl) phenol	15.310	206.05 (16.65); 191.10 (100); 163.05 (7.34); 135.10 (2.790); 107.05 (5); 88.05 (3.59); 73.05 (6.04); 57.10 (31.97); 55.10 (1.97)
16	Pentachlorophenol	19.815	265.75 (100.00); 263.75 (57.84); 229.80 (18.83); 201.80 (15.62); 168.90 (8.84); 164.90 (30.24); 129.85 (15.10); 115.00 (3.17); 95.00 (21.19); 83.10 (5.63); 60.00 (9.59)
16	2,6-Di-tert-butylbenzoquinone	21.808	220 (100); 205 (15.08); 192 (1.64); 177 (34.07); 163 (0.95); 149 (3.17); 135 (1.05); 121 (1.24): 107 (0.09); 95 .00 (1.42); 77. 05 (2.15); 67.01 (1.04)

TMS: trimethylsilated

Table S3: Sequence analysis matric of the draft genome

Analysis Metric	**Value**
Polished Contigs	02
Maximum Contig Length	5,246,860
N50 Contig Length	5,246,860
Sum of Contig Lengths	5,305,309
E-size (sum of squares / sum)	5,189,699

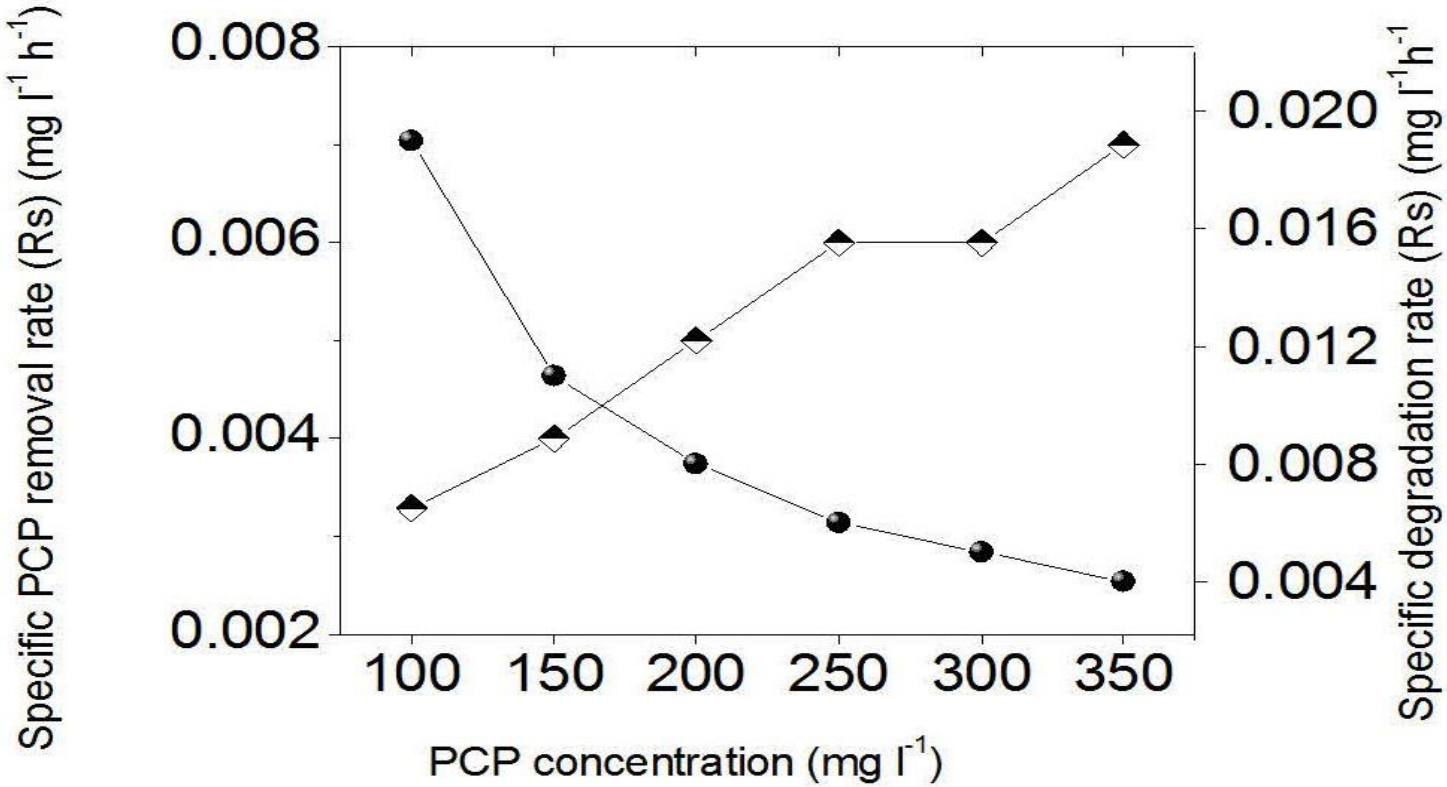


Figure S1: Reciprocal plot of transformation rate (1/R_s) vs PCP concentration (1/S).

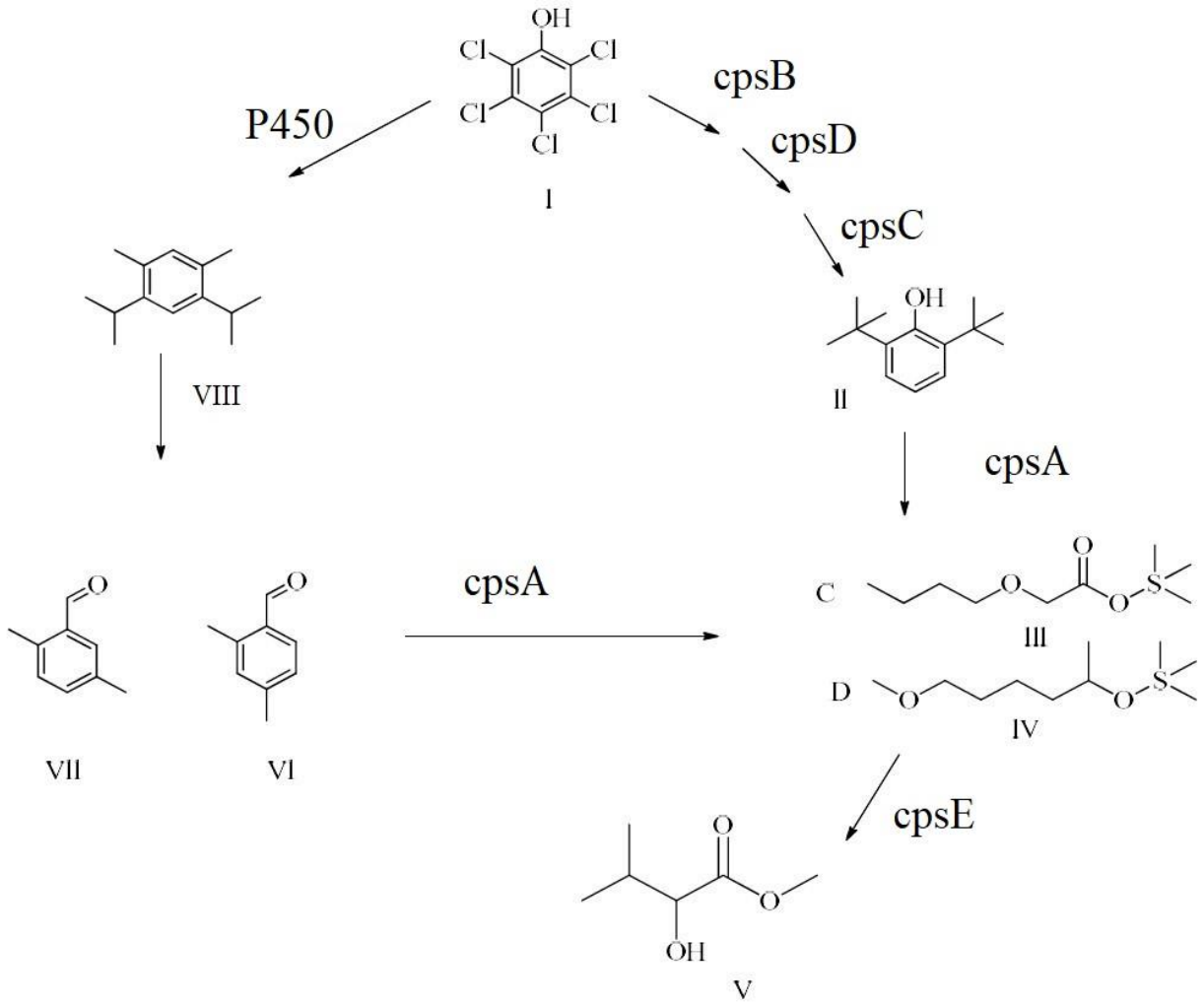


Figure S2: Proposed PCP degradation pathways in *Bacillus tropicus*AOA-CPS-1.

PCP (I); 2,6-bis(1,1-dimethylethyl) phenol (II); trimethylsilyl 2-butoxyacetate (III); 1-methoxy-5-trimetylsilyoxyhexane (IV); Methyl 2-hydroxyl-3-methylbutanoate (V); 2,4-Dimethylbenzenecarboxaldehyde (VI); 2,5-Dimethylbenzaldehyde (VII); 1,3-Dimethyl-4,6-diisopropylbenzene (VIII).

```
CpsB        MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN    60
315749      -MKKTEIPAHLKPFVSKQHYDQYTPINHAVWRYIMRQNHNFLKDVAHPAYVNGLKSSGIN    59
315730      MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPSYVNGLQSSGIN    60
226900      MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN    60
405532      MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN    60
412694      MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN    60
261594      MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN    60
198094      MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN    60
280355      MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN    60
280477      MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN    60
280354      MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN    60
205919      MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN    60
260799      MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN    60
261591      MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN    60
281309      MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN    60
222523      MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN    60
269801      MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN    60
288681      MTKKTEIPSHLKPFVSTQHYDQYTPVNHAVWRYIMRQNHSFLKDVAHPAYVNGLQSSGIN    60
              ******:*******.********:*************.********:*****:*****

CpsB        IDAIPKVEEMNECLAPSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP    120
315749      IDAIPKVEEMNECLAPSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP    119
315730      IDAIPKVEEMNDCLAPSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP    120
226900      IDAIPKVEEMNECLAPSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP    120
405532      IDAIPKVEEMNECLAPSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP    120
412694      IEAIPKVEEMNECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP    120
261594      IEAIPKVEEMNECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP    120
198094      IEAIPKVEEMNECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP    120
280355      IEAIPKVEEMNECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP    120
280477      IEAIPKVEEMNECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP    120
280354      IEAIPKVEEMNECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP    120
205919      IEAIPKVEEMNECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP    120
260799      IEAIPKVEEMNECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP    120
261591      IEAIPKVEEMNECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP    120
281309      IEAIPKVEEMNECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP    120
222523      IDAIPKVEEMNECLAPSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP    120
269801      IDAIPKVEEMNECLAPSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP    120
288681      IEAIPKVEEMNECLASSGWGAVTIDGLIPGVAFFDFQGHGLLPIATDIRKVENIEYTPAP    120
            *:*********:*** *******************************************

CpsB        DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD    180
315749      DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPE    179
315730      DIVHEAAGHAPILLDSTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPE    180
226900      DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD    180
405532      DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD    180
412694      DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD    180
261594      DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD    180
198094      DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD    180
280355      DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD    180
280477      DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD    180
280354      DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD    180
205919      DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD    180
260799      DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD    180
261591      DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD    180
281309      DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD    180
222523      DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD    180
269801      DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD    180
288681      DIVHEAAGHAPILLDPTYAKYVKRFGQIGAKAFSTKEEHDAFEAVRTLTIVKESPTSTPD    180
            *************** *******************************************:
```

```
CpsB        EVTAAENNVLEKQKLVSGLSEAEQISRLFWWTVEYGLIGNIDTPKIYGAGLLSSVGESKH    240
315749      EIEAAEKEVIEKQKLVSGVSEAEQISRLFWWTVEYGLIGDLDNPKIYGAGLLSSVGESKY    239
315730      EVAVAENTVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGNIDDPKIYGAGLLSSVGESKH    240
226900      EVKAAENAVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGNIDDPKIYGAGLLSSVGESKH    240
405532      EVKAAENAVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGNIDDPKIYGAGLLSSVGESKH    240
412694      EVTAAENNVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGDIDNPKIYGAGLLSSVGESKH    240
261594      EVTAAENNVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGDIDNPKIYGAGLLSSVGESKH    240
198094      EVTAAENNVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGDIDNPKIYGAGLLSSVGESKH    240
280355      EVTAAENNVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGDIDNPKIYGAGLLSSVGESKH    240
280477      EVTAAENNVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGDIDNPKIYGAGLLSSVGESKH    240
280354      EVTAAENNVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGDIDNPKIYGAGLLSSVGESKH    240
205919      EVTAAENNVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGDIDNPKIYGAGLLSSVGESKH    240
260799      EVTAAENNVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGDIDNPKIYGAGLLSSVGESKH    240
261591      EVTAAENNVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGDIDNPKIYGAGLLSSVGESKH    240
281309      EVTAAENNVIEKQNLVSGLSEAEQISRLFWWTVEYGLIGDIDNPKIYGAGLLSSVGESKH    240
222523      EVTAAENNVLEKQKLVSGLSEAEQISRLFWWTVEYGLIGNIDAPKIYGAGLLSSVGESKH    240
269801      EVTAAEHNVLEKQKLVSGLSEAEQISRLFWWTVEYGLIGNIDAPKIYGAGLLSSVGESKH    240
288681      EVTAAENNVLEKQKLVSGLSEAEQISRLFWWTVEYGLIGNIDAPKIYGAGLLSSVGESKH    240
            *: .**: *:***:****:********************::* ****************:

CpsB        CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCESFEELTEALEKFSETMAFKTGGKEGL    300
315749      CLTDAVEKVPFSLEACIKTTYDVTKMQPQLFVCQSFEELIEALEAFSKTMAFQTGGAEGL    299
315730      CLTDAVEKVPFSIEACTGTTYDVTKMQPQLFVCESFEELTEELEKFSETMAFKTGGTEGL    300
226900      CLTDAVEKVPFSIEACIGTTYDVTKMQPQLFVCESFEELTDALETFSKTMAFKTGGKEGL    300
405532      CLTDAVEKVPFSIEACIGTTYDVTKMQPQLFVCESFEELTDALETFSKTMAFKTGGKEGL    300
412694      CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCKSFEELTEALEKFAETMAFKTGGKEGL    300
261594      CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCKSFEELTEALEKFAETMAFKTGGKEGL    300
198094      CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCKSFEELTEALEKFAETMAFKTGGKEGL    300
280355      CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCKSFEELTEALEKFAETMAFKTGGKEGL    300
280477      CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCKSFEELTEALEKFAETMAFKTGGKEGL    300
280354      CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCKSFEELTEALEKFAETMAFKTGGKEGL    300
205919      CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCKSFEELTEALEKFAETMAFKTGGKEGL    300
260799      CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCKSFEELTEALEKFAETMAFKTGGKEGL    300
261591      CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCKSFEELTEALEKFAETMAFKTGGKEGL    300
281309      CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCKSFEELTEALEKFAETMAFKTGGKEGL    300
222523      CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCESFEELTEALEKFSETMAFKTGGKEGL    300
269801      CLTDAVEKVPFSIEACTSTTYDVTKMQPQLFVCESFEELTEALEKFSETMAFKTGGKEGL    300
288681      CLTDAVEKVPFSIETCTSTTYDVTKMQPQLFVCESFEELTEALEKFSETMAFKTGGKEGL    300
            ************:*:*  ***************:***** : ** *::****:*** ***

CpsB        EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKQLANHSTSV    360
315749      EKAIRSENIATAELSSGLQITGTFSEMVQNEVGEVIYLKTNTPTALAFNHKQLPHHSTAI    359
315730      EKAIRSENHATTELSSGLQITGTFTETIKNDTGEVIYMRTNTPTALAINHKQLVNHSTSV    360
226900      EKAIRSENYATAELNSGLQITGTFSETIENDAGELIYMRTNSPTALALHNKQLANHSTSV    360
405532      EKAIRSENYATAELNSGLQITGTFSETIENDAGELIYMRTNSPTALALHNKQLANHSTSV    360
412694      EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKQLANHSTAV    360
261594      EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKELANHSTAV    360
198094      EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKELANHSTAV    360
280355      EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKELANHSTAV    360
280477      EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKELANHSTAV    360
280354      EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKELANHSTAV    360
205919      EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKELANHSTAV    360
260799      EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKELANHSTAV    360
261591      EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKELANHSTAV    360
281309      EKAIRSENNATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKELANHSTAV    360
222523      EKAIRSENHATAELNSGLQITGTFTETIENDTGELIYMRTSSPTALAIHNKQLANHSTSV    360
269801      EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKQLANHSTSV    360
288681      EKAIRSENHATAELNSGLQITGTFTETIENDAGELIYMRTSSPTALAIHNKQLANHSTSV    360
            ******** **:**.*********:* ::*:.**:**::*.:*****:::*:* :***::
```

```
CpsB        HSDGFGTPIGLLTGNIALENCTDEQLQSLGITIGNKAAFTFASGIHVKGTVTDIVKNDKK    420
315749      HEDGFGTPIGLLQNNIALEDCTEESLQSLGILIGNNTDLSFASGVHVKGTVTDIIKQDEK    419
315730      HKDGFGTPVGLLDGNIALEDCTEEKLQSLGIIIGNLVELSFTSGVHVKGTVTDIVKNDKK    420
226900      HSDGFGTPIGLLTENIALENCTDEQLQSLGITIGTIAEFTFASGIHVKGTVTDIVKNDKK    420
405532      HSDGFGTPIGLLTENIALENCTDEQLQSLGITIGTIAEFTFASGIHVKGTVTDIVKNDKK    420
412694      HSDGFGTPIGLLTENIALENCTDEQLQALGITIGNIAEFTFESGIHVKGTVIDIVKNHNK    420
261594      HSDGFGTPIGLLTENIALENCTDEQLQALGITIGNIAEFTFESDIHVKGTVTDIVKNDNK    420
198094      HSDGFGTPIGLLTENIALENCTDEQLQALGITIGNIAEFTFESDIHVKGTVTDIVKNDNK    420
280355      HSDGFGTPIGLLTENIALENCTDEQLQALGITIGNIAEFTFESDIHVKGTVTDIVKNDNK    420
280477      HSDGFGTPIGLLTENIALENCTDEQLQALGITIGNIAEFTFESDIHVKGTVTDIVKNDNK    420
280354      HSDGFGTPIGLLTENIALENCTDEQLQALGITIGNIAEFTFESDIHVKGTVTDIVKNDNK    420
205919      HSDGFGTPIGLLTENIALENCTDEQLQALGITIGNIAEFTFESDIHVKGTVTDIVKNDNK    420
260799      HSDGFGTPIGLLTENIALENCTDEQLQALGITIGNIAEFTFESDIHVKGTVTDIVKNDNK    420
261591      HSDGFGTPIGLLTENIALENCTDEQLQALGITIGNIAEFTFESDIHVKGTVTDIVKNDNK    420
281309      HSDGFGTPIGLLTENIALENCTDEQLQALGITIGNIAEFTFESGIHVKGTVTDIVKNDNK    420
222523      HSDGFGTPIGLLTENIALENCTDEQLQSLGITIGNVAEFTFASGIHVKGTVTDIVKNDKK    420
269801      HSDGFGTPIGLLTENIALENCTDEQLQSLGITIGNKAAFTFASGIHVKGTVTDIVKNDKK    420
288681      HSDGFGTPIGLLTENIALENCTDEQLQSLGITIGNKAAFTFASGIHVKGTVTDIVKNDKK    420
            *.******:***  *****:**:*.**:*** **. . ::* *.:****** **:*:.:*

CpsB        IALISFINCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAASFFPMEEEIQEIPAP    480
315749      VVLISFTNCTVVYKDRLLFDASWGTFDMAVGSNITSVFPGAADAASFFPMDEEIEKTPAP    479
315730      IVLISFAGCTVTHKDCLLFDPSWGTFDMAVGSTITSVFPGAADAAAFFPMDEEVHETPAP    480
226900      IALISFIDCTVTYNARVLFDASWGAFDMAVGSQITSVFPGAADAAAFFPMDEEVHEIPAP    480
405532      IALISFIDCTVTYNARVLFDASWGAFDMAVGSQITSVFPGAADAAAFFPMDEEVHEIPAP    480
412694      IALISFINCTVTYKDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP    480
261594      IALISFINCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP    480
198094      IALISFINCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP    480
280355      IALISFINCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP    480
280477      IALISFINCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP    480
280354      IALISFINCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP    480
205919      IALISFINCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP    480
260799      IALISFINCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP    480
261591      IALISFINCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP    480
281309      IALISFINCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP    480
222523      IALISFISCTVTYNDRVLFDASWGAFDMAVGSTITSVFPGAADAAAFFPTDEEVQKIPAP    480
269801      IALISFINCTVTYNDRVLFDASWGSFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP    480
288681      IALISFINCTVTYNDRVLFDASWGSFDMAVGSTITSVFPGAADAAAFFPMDEEIQEIPAP    480
            :.**** .***.::  :*** ***:******* ************:*** :**:.: ***


CpsB        LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYTKEWLLRLEILELLLEHNK    540
315749      LSLSELDRMYQMVRDIRNKGELQDSDVAQLVAIHEVLNQFYKKEWLLRLEILELLVEHNK    539
315730      LVLTEVERMYQTVRNIRNQGTLQDAHVDQLIAIQEVLNTFYPKEWLLRLEILELLLQHNK    540
226900      LVLNELERMYQTVRDIRSEGILHDAHIDQLIAIQEVLNKFYAKEWLLRLEVLELLLEHNK    540
405532      LVLNELERMYQTVRDIRSEGILHDAHIDQLVAIQEVLNKFYAKEWLLRLEVLELLLEHNK    540
412694      LVLNELERMYQTVRDIRNEGILHDAHIDQLVAIQEVLNKFYAKEWLLRLEILELLLEHNK    540
261594      LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYTKEWLLRLEILELLLEHNK    540
198094      LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYTKEWLLRLEILELLLEHNK    540
280355      LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYTKEWLLRLEILELLLEHNK    540
280477      LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYTKEWLLRLEILELLLEHNK    540
280354      LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYTKEWLLRLEILELLLEHNK    540
205919      LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYTKEWLLRLEILELLLEHNK    540
260799      LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYTKEWLLRLEILELLLEHNK    540
261591      LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYTKEWLLRLEILELLLEHNK    540
281309      LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYAKEWLLRLEILELLLEHNK    540
222523      LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKLYPKEWLLRLEILELLLEHNK    540
269801      LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYTKEWLLRLEILELLLEHNK    540
288681      LVLNELERMYQTVRDIRNEGILHDAHIEQLVAIQEVLNKFYTKEWLLRLEILELLLEHNK    540
            * *.*::**** **:**.:* *:*:.: **:**:**** :* ********:****::***
```

```
CpsB        GHETSAALLQQLSTFATDEAVTRLINNGLTLLPVKDVKNDATIN--  584
315749      DQKTASFLLQQLSTFTENESVQRLIHNGLALLPIKDVKNNATINRS  585
315730      GHETSTVLLKQLSTFTTDKAVTRLITNGLALLPIKDVKNDATIS--  584
226900      GHETSAALLHQLSTFTTDEAVTRLINNGLALLPVKDVKNDAKIN--  584
405532      GHETSAALLHQLSTFTTDEAVTRLINNGLALLPVKDVKNDAKIN--  584
412694      GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKDVKNDATIN--  584
261594      GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKGVKNDATIN--  584
198094      GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKGVKNDATIN--  584
280355      GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKGVKNDATIN--  584
280477      GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKGVKNDATIN--  584
280354      GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKGVKNDATIN--  584
205919      GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKGVKNDATIN--  584
260799      GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKGVKNDATIN--  584
261591      GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKGVKNDATIN--  584
281309      GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKDVKNDATIN--  584
222523      GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKDVKNDATIN--  584
269801      GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKDVKNDATIN--  584
288681      GHETSAALLQQLSTFTTDEAVTRLINNGLTLLPVKDVKNDATIN--  584
            .::*:: **:*****: :::* *** ***:***:*.***:*.*.
```

Fig. S3: Multiple sequences alignments between of CpsB from *Bacillus tropicus* AOA-CPS1 with closely related neighbourhoods' whole-genome. Where the sequences are similar were indicated with asterisk (*). Pterin-4-alpha-carbinolamine dehydratase (EC 4.2.1.96) from *Bacillus tropicus* AOA-CPS1 (CpsB), *Pseudomonas mendocina* (482703), *B. anthracis* str. Ames Ancestor (261594), *B. cereus* ATCC 14579 (226900), *B. thuringiensis* serovar konkukian str. 97-27 (281309), *B. anthracis* str. Ames (198094), *B. anthracis* str. A1055 (280355), *B. anthracis* str. Australia 94 (280477), *B. anthracis* str. CNEVA-9066 (280354), *B. anthracis* str. Kruger B (205919), *B. anthracis* str. Sterne (260799), *B. anthracis* str. Vollum (261591), *B. cereus* G9241 (269801), *B. cereus* E33L (288681), *B. cereus* B4264 (405532), *B. cereus* ATCC 10987 (222523), *B. cereus* subsp. cytotoxis NVH 391-98 (315749), *B. thuringiensis* str. Al Hakam (412694) and *B. weihenstephanensis* KBAB4 (315730).

```
                                  ::         * :    ..  : : : *..: :.: *
CpsD        --------------MMLRLTEEEVQEELLKLDKWVVK---DEKWIERKYMFSDYLKGVEF    43
66692       --------------------MSQIKKTAKALPGWTVE---GEA-LWRTFTFDHFLKGIDF    36
315749      ---------------MQRLTDHEVQEELKKLDKWTVK---DEKWIERKYMFSDYLKGVEF    42
315730      --------------MMLRLADEEVHEELAKLDKWTVK---DEKWIERKYMFSDYLKGVEF    43
226900      --------------MMLRLTEEEVHEELLKVDKWMVK---DEKWIERKYMFSDYLKGVEF    43
405532      --------------MMLRLTEEEVHEELLKVDKWMVK---DEKWIERKYMFSDYLKGVEF    43
288681      --------------MMLRLTEEEVQEELLRLDKWVVK---DEKWIERKYMFSDYLKGVEF    43
261594      --------------MMLRLTEEEVQEELLKLDKWVVK---DEKWIERKYMFSDYLKGVEF    43
198094      --------------MMLRLTEEEVQEELLKLDKWVVK---DEKWIERKYMFSDYLKGVEF    43
280355      --------------MMLRLTEEEVQEELLKLDKWVVK---DEKWIERKYMFSDYLKGVEF    43
280477      --------------MMLRLTEEEVQEELLKLDKWVVK---DEKWIERKYMFSDYLKGVEF    43
280354      --------------MMLRLTEEEVQEELLKLDKWVVK---DEKWIERKYMFSDYLKGVEF    43
261591      --------------MMLRLTEEEVQEELLKLDKWVVK---DEKWIERKYMFSDYLKGVEF    43
260799      ---------------MLRLTEEEVQEELLKLDKWVVK---DEKWIERKYMFSDYLKGVEF    42
205919      --------------MMLRLTEEEVQEELLKLDKWVVK---DEKWIERKYMFSDYLKGVEF    43
412694      --------------MMLRLTEEEVQEELLKLDKWVVK---DEKWIERKYMFSDYLKGVEF    43
269801      --------------MMLRLTEEEVQEELLRLDKWVVK---DEKWIERKYMFSDYLKGVEF    43
281309      --------------MMLRLTEEEVQEELLKLDKWVVK---DEKWIERKYMFSDYLKGVEF    43
222523      --------------MMLRLTEEEVQEELLKLDKWVVK---DEKWIERKYMFSDYLKGVEF    43
235909      ----------------MRLTEEEVQALLEKADGWKLA---DERWIVKKYRFQDYLQGIEF    41
420246      ----------------MRLPETEVQALLAKVDGWKLT---DERWIVKKYRFQDYLQGIEF    41
482703      MTSLAQAQCEACRADAPKVSDEELAELIREIPDWNIEVRGDHMELERVYLFKNFRHALAF    60
                                  ::         * :    ..  : : : *..: :.: *


            .   .  :*   *** :  :   : :.  :     * : *:  *   : .  .
CpsD        VSEAAKLSEEHNHHPFILIQYKAVIITLSSWNAKGLTKLDFELAKQFDELFIQNEKAIIRK   104
66692       VQNVAAYSEGAQHHPVITIDHTAVSIKWTTVDEGKLTQKDIDAAKACNQFFEN--------   89
315749      VSEAAKLSEEHNHHPFILIQYKAVILTLSSWNAKGLTKLDFDLAKQFDDLFLQNEKAIIKK   103
315730      VSEAAKLSEEHNHHPFILIQYKAVIITLSSWNAKGLTKLDFDLAKQFDDLFLQNEKAIIRK   104
226900      VSEAAKLSEEHNHHPFILIQYKAVIITLSSWNAKGLTKLDFELAKQFDELFLQNEKAIIRK   104
405532      VSEAAKLSEEHNHHPFILIQYKAVIITLSSWNAKGLTKLDFELAKQFDELFLQNEKAIIRK   104
288681      VSEAAKLSEEHNHHPFILIQYKAVIITLSSWNAKGLTKLDFELAKQFDELFVQNEKAVIRK   104
261594      VSEAAKLSEEHNHHPFILIQYKAVIITLSSWNAKGLTKLDFELAKQFDELFVQNEKAVIRK   104
198094      VSEAAKLSEEHNHHPFILIQYKAVIITLSSWNAKGLTKLDFELAKQFDELFVQNEKAVIRK   104
280355      VSEAAKLSEEHNHHPFILIQYKAVIITLSSWNAKGLTKLDFELAKQFDELFVQNEKAVIRK   104
280477      VSEAAKLSEEHNHHPFILIQYKAVIITLSSWNAKGLTKLDFELAKQFDELFVQNEKAVIRK   104
280354      VSEAAKLSEEHNHHPFILIQYKAVIITLSSWNAKGLTKLDFELAKQFDELFVQNEKAVIRK   104
261591      VSEAAKLSEEHNHHPFILIQYKAVIITLSSWNAKGLTKLDFELAKQFDELFVQNEKAVIRK   104
260799      VSEAAKLSEEHNHHPFILIQYKAVIITLSSWNAKGLTKLDFELAKQFDELFVQNEKAVIRK   103
205919      VSEAAKLSEEHNHHPFILIQYKAVIITLSSWNAKGLTKLDFELAKQFDELFVQNEKAVIRK   104
412694      VSEAAKLSEEHNHHPFILIQYKAVIITLSSWNAKGLTKLDFELAKQFDELFVQNEKAVIRK   104
269801      VSEAAKLSEEHNHHPFILIQYKAVIITLSSWNAKGLTKLDFELAKQFDELFIQNEKAIIRK   104
281309      VSEAAKLSEEHNHHPFILIQYKAVIITLSSWNAKGLTKLDFELAKQFDELFVQNENAIIRK   104
222523      VSEAAKLSEEHNHHPFILIQYKAVIITLSSWNAKGLTKLDFELAKQFDELFVQNEKAIIRK   104
235909      VRRIAAISENANHHPFISIDYKLITVKLSSWRAKGLTKLDFDLAKQYDEVYNQMKQGEGEK   102
420246      VRQIALISENANHHPFISIDYKLITVKLSSWRAKGLTKLDFDLAQQYDDVYKQIKQGDGEK   102
482703      TNAVGAIAEEVGHHPALLTEWGKVTVTWWSHEMRGLHRNDFIMAARTDVLAASAEG---RK   118
            .   .  :*   *** :  :   : :.  :     * : *:  *   : .  .
```

Fig. S4: Multiple sequences alignments between of CpsD from *Bacillus tropicus* AOA-CPS1 with closely related neighbourhoods' whole-genome. Where the sequences are similar were indicated with asterisk (*) shaded in green. Pterin-4-alpha-carbinolamine dehydratase (EC 4.2.1.96) from *Bacillus tropicus* AOA-CPS1 (CpsB), *Pseudomonas mendocina* (482703), *B. anthracis* str. Ames Ancestor (261594), *B. cereus* ATCC 14579 (226900), *B. thuringiensis*

serovar konkukian str. 97-27 (281309), *B. anthracis* str. Ames (198094), *B. anthracis* str. A1055 (280355), *B. anthracis* str. Australia 94 (280477), *B. anthracis* str. CNEVA-9066 (280354), *B. anthracis* str. Kruger B (205919), *B. anthracis* str. Sterne (260799), *B. anthracis* str. Vollum (261591), *B. cereus* G9241 (269801), *B. cereus* E33L (288681), *B. cereus* B4264 (405532), *B. cereus* ATCC 10987 (222523), *B. cereus* subsp. cytotoxis NVH 391-98 (315749), *B. thuringiensis* str. Al Hakam (412694) and *B. weihenstephanensis* KBAB4 (315730).

```
                 * .                            :::    :* :*
176279      ----MKAYEYL--KPGHAQLTDKEKPTITSSTDAIIRIVKTTICGTDLHIIKGDTP----     50
482703      ----MQALTYH--GSHDVRVEQVPDPVIEQPDDIILRVTATAICGSDLHLYRGKIP----     50
279010      ----MKAVTFQ--GPQHVEVSQVEDAKIEKSDDIVVRITSTAICGSDLHLYQGNFP----     50
281309      ----MKAVTYQ--GPNKVQVKQVDDAKLEKKDDIIVKITSTAICGSDLHLYQGNMP----     50
412694      ----MKAVTYQ--GPNQVQVKQVDDAKLEKKDDIIVKITSTAICGSDLHLYQGNMP----     50
222523      ----MKAVTYQ--GPNQVQVKQVDDAKLEKKDDIIVKITSTAICGSDLHLYQGNMP----     50
261594      ----MKAVTYQ--GPNQVQVKQVDDAKLEKKDDIIVKITSTAICGSDLHLYQGNMP----     50
198094      ----MKAVTYQ--GPNQVQVKQVDDAKLEKKDDIIVKITSTAICGSDLHLYQGNMP----     50
280477      ----MKAVTYQ--GPNQVQVKQVDDAKLEKKDDIIVKITSTAICGSDLHLYQGNMP----     50
280354      ----MKAVTYQ--GPNQVQVKQVDDAKLEKKDDIIVKITSTAICGSDLHLYQGNMP----     50
205919      ----MKAVTYQ--GPNQVQVKQVDDAKLEKKDDIIVKITSTAICGSDLHLYQGNMP----     50
260799      ----MKAVTYQ--GPNQVQVKQVDDAKLEKKDDIIVKITSTAICGSDLHLYQGNMP----     50
261591      ----MKAVTYQ--GPNQVQVKQVDDAKLEKKDDIIVKITSTAICGSDLHLYQGNMP----     50
280355      ----MKAVTYQ--GPNQVQVKQVDDAKLEKKDDIIVKITSTAICGSDLHLYQGNMP----     50
269801      ----MKAVTYQ--GPNKVQVKQVDDAKLEKKDDIIIKITSTAICGSDLHLYQGNMP----     50
CpsC        ----MKAVTYQ--GPNQVQVKQVDDAKLEKKDDIIVKITSTAICGSDLHLYQGNMP----     50
288681      ----MKAVTYQ--GPNQVQVKQVDDAKLEKKDDIIVKITSTAICGSDLHLYQGNMP----     50
226900      ----MKAVTYQ--GPNQVQVKQVDDAKLEKKDDIIVKITSTAICGSDLHLYQGNMP----     50
405532      ----MKAVTYQ--GPNQVQVKQVDDAKLEKKDDIIVKITSTAICGSDLHLYQGNMP----     50
66692       ----MKAVTYQ--GMNHVEVSTIKAPEIQDPEDVIVRITSTAICGSDLHLYQGNMY----     50
315750      ----MRAVTYQ--GKNSIAVKKVDAPSIQDREDVIIRITSTAICGSDLHLYQGNFP----     50
224308      ----MKAVTYQ--GIKNVVVKDVPDPKIEKSDDMIIKVTSTAICGSDLHLIHGFIP----     50
326423      ----MKAVTYQ--GVKNVVVKDVPDPKIEKHDDMIIKVTSTAICGSDLHLIHGLIP----     50
272558      ----MKAVTYQ--GIKDVKVKEVPDPKLEKKDDVLVRITSTAICGSDLHLVHGMIP----     50
235909      ----MKAVTYQ--GIKDVAVKQMPDPKILKDDDIIVKITSTAICGSDLHLVHGMIP----     50
420246      ----MLALYVTKPNELELRHL--DGIRSPVSDEVKIKLIYGGICGSDLGVFKGKLP----     50
491915      ----MKAAVVE-QFKEPLQIKDVEKPTIS-YGEVLVRIKACGVCHTDLHAAHGDWPVK--     52
282458      ----MKALVKTREGHGNLELLDKEVATPL-DDKVKIKVHYAGICGTDIHTYEGHYK----     51
272626      ----MKAVVKTNPGYDQMELRDVEEPQVY-GDKVKIKVAFTGICGSDIHTFKGEYK----     51
169963      ----MKAVVKTNPGYDQMELKDVEEPQVY-GDKVKIKVAFTGICGSDIHTFKGEYK----     51
221109      ----MEALNLY--GIEDLRYEDTPKPMIEKDDDVIIKVKSVGICGSDTSRYKKLGP----     50
195103      MEGKMKVAVMN--GIGKMDLIERDIPIVK-ENEVLVKLDYVGICGSDLHYYENGRIGDY-     56
315749      ----MKAAVWY--GEKDIRIEEREVKELQ-PNDVKVKVAWTGICGSDLHAYLHPDS-VP-     51
413999      ----MKAALWY--EKKDVRVEEIEEPKVV-EGSVKIKVKWCGICGSDLHEYLGGPIFIPV     53
101031      ----MKAARWY--KAKDIRVETIEEPVIA-PGKVKIKVHWTGICGSDLHEYAAGPIFIPV     53
315730      ----MKALLWH--NQRDVRVEEVPEPTVR-PGAVKIKVKWCGICGTDLHEYLAGPIFIPT     53
                 * .                            :::    :* :*
```

```
                          ***  * : * *      . *  *         . .
176279      -------EVKSHTTLGHEGIGIIEEIGDNVNNFKVGDKVIISCI-SSCGKCYYCKKGIYA    102
482703      -------QVKDGDIFGHEFMGVVEEVGPQVTAVSKGDRVIVPFV-IACGDCFFCQMDLHA    102
279010      --------LPKGFQLGHEPMGIVEETGPDVTKVKKGDRVVIPFT-VACGHCFYCENKLES    101
281309      --------LPPGYIIGHEPMGIVEEVGPDVTKVKKGDRVVIPFN-VACGHCFYCQHEMES    101
412694      --------LPQGYIIGHEPMGIVEEVGPDVTKVKKGDRVVIPFN-VACGHCFYCQHEMES    101
222523      --------LPQGYIIGHEPMGIVEEVGPDVTKVKKGDRVVIPFN-VACGHCFYCQHEMES    101
261594      --------LPQGYIIGHEPMGIVEEVGPDVTKVKKGDRVVIPFN-VACGHCFYCQHEMES    101
198094      --------LPQGYIIGHEPMGIVEEVGPDVTKVKKGDRVVIPFN-VACGHCFYCQHEMES    101
280477      --------LPQGYIIGHEPMGIVEEVGPDVTKVKKGDRVVIPFN-VACGHCFYCQHEMES    101
280354      --------LPQGYIIGHEPMGIVEEVGPDVTKVKKGDRVVIPFN-VACGHCFYCQHEMES    101
205919      --------LPQGYIIGHEPMGIVEEVGPDVTKVKKGDRVVIPFN-VACGHCFYCQHEMES    101
260799      --------LPQGYIIGHEPMGIVEEVGPDVTKVKKGDRVVIPFN-VACGHCFYCQHEMES    101
261591      --------LPQGYIIGHEPMGIVEEVGPDVTKVKKGDRVVIPFN-VACGHCFYCQHEMES    101
280355      --------LPQGYIIGHEPMGIVEEVGPDVTKVKKGDRVVIPFN-VACGHCFYCQHEMES    101
269801      --------LPPGYIIGHEPMGIVEEVGPDVTKVKKGDRVVIPFN-VACGHCFYCQHEMES    101
CpsC        --------LPPGYIIGHEPMGIVEEVGPDVTKVKKGDRVVIPFN-VACGHCFYCQHEMES    101
288681      --------LPPGYIIGHEPMGIVEEVGPDVTKVKKGDRVVIPFN-VACGHCFYCQHEMES    101
226900      --------LPQGYIIGHEPMGIVEEVGPDVTKVKKGDRVVIPFN-VSCGHCFYCQHEMES    101
405532      --------LPQGYIIGHEPMGIVEEVGPDVTKVKKGDRVVIPFN-VSCGHCFYCQHEMES    101
66692       --------TPKGFVIGHEPMGIVEETGPEVTKVKKGDRVVIPFT-VACGHCPYCQNHQES    101
315750      --------LPIGYVIGHEPMGIVEEVGPDVTAVKKGDRVVIPFT-VACGQCQYCHHHLES    101
224308      -------NMQEDYVTGHEPMGIVEEVGSGVTKLKKGDRVIIPFN-IACGECFFCKNQLES    102
326423      -------NLQEDYIIGHEPMGIVEETGPGVTKVKKGDRVIIPFN-IACGECVYCKNHLES    102
272558      -------NFPKGYIIGHEPMGIVEEVGPDVTKVKKGDRVIVPFN-IACGKCYYCTHDLES    102
235909      -------NMPEDFIIGHEPMGIVEEVGPAVTKVKKGDRVIVPFT-IACGQCWYCQHGLES    102
420246      -------HANYPVRAGHELVGVVVEKGEQ-AAYDIGTRVVVLPN-TYCGTCDLCQKGYTN    101
491915      --------PKLPLIPGHEGVGVIEEVGPGVTHLKVGDRVGIPWLYSACGHCDYCLSGQET    104
282458      -------V-NFPVTLGHEFSGEIVEVGADVKDFKVGDRVTSETTFYVCNECEYCESKDYN    103
272626      -------NPTTPVTLGHEFSGVVVEVGPDVTSIKVGDRVTSETTFETCGECIYCKERDYN    104
169963      -------NPTTPVTLGHEFSGVVVEVGPDVTSIKVGDRVTSETTFETCGECIYCKEHDYN    104
221109      --------YVEGMTFGHEFAGEVIEVGKSVTNFKSGDRVVACPT-FSCGNCHYCREGHPT    101
195103      -------IVEPPFVLGHEPGGVVVEVGNKVKHLNIGDRVALEPG-KTCGHCEFCKTGRYN    108
315749      --------MNRNMVLGHEFSGEIVEVGSHVTKFKEGDRVCIYPM-MLKDPSNAEIERFI-    101
413999      GQPHALSGTTAPVVLGHEFSGEIVELGQGVTKFNIGDRVIVEPI-VACGKCPACMEGKYN    112
101031      EQPHYVSKDIAPIVMGHEFSGEVVEIGDAVTSVKVGDPVVVEPI-LSCGECAACKKGKYN    112
315730      EEH-PLTHVKAPVILGHEFSGEVVEIGEGVTSHKVGDRVVVEPI-YSCGKCEACKHGHYN    111
                          ***  * : * *      . *  *         . .
```

```
                                                  *   .                .
176279      HCENGGGWIL---------------G-----HLVNGTQAEYVKVPFADNSLYHAPSN--L     140
482703      ACETTNPGRGAILNKKQIPPGAALFGYSHLYGGVPGGQAELVRVPKANAGPFKVPDV--L     160
279010      QCDNANPHYD----------SGGYFGYSEKFGNHPGGQAEYLKVPFGNFTPFVIPESCEM     151
281309      QCDNSNPHYD----------SGGYFGYTEKFGNHPGGQAEYLKVPFGNFTPFVIPESCEL     151
412694      QCDNSNPHYD----------SGGYFGYTEKFGNHPGGQAEYLKVPFGNFTPFVIPESCEL     151
222523      QCDNSNPHYD----------SGGYFGYTEKFGNHPGGQAEYLKVPFGNFTPFVIPESCEL     151
261594      QCDNSNPHYD----------SGGYFGYTEKFGNHPGGQVEYLKVPFGNFTPFVIPESCEL     151
198094      QCDNSNPHYD----------SGGYFGYTEKFGNHPGGQVEYLKVPFGNFTPFVIPESCEL     151
280477      QCDNSNPHYD----------SGGYFGYTEKFGNHPGGQVEYLKVPFGNFTPFVIPESCEL     151
280354      QCDNSNPHYD----------SGGYFGYTEKFGNHPGGQVEYLKVPFGNFTPFVIPESCEL     151
205919      QCDNSNPHYD----------SGGYFGYTEKFGNHPGGQVEYLKVPFGNFTPFVIPESCEL     151
260799      QCDNSNPHYD----------SGGYFGYTEKFGNHPGGQVEYLKVPFGNFTPFVIPESCEL     151
261591      QCDNSNPHYD----------SGGYFGYTEKFGNHPGGQVEYLKVPFGNFTPFVIPESCEL     151
280355      QCDNSNPHYD----------SGGYFGYTEKFGNHPGGQAEYLKVPFGNFTPFVIPESCEL     151
269801      QCDNSNPHYD----------SGGYFGYTEKFGNHPGGQAEYLKVPFGNFTPFVIPESCEL     151
CpsC        QCDNSNPHYD----------SGGYFGYTEKFGNHPGGQAEYLKVPFGNFTPFVIPESCEL     151
288681      QCDNSNPHYD----------SGGYFGYTEKFGNHPGGQAEYLKVPFGNFTPFVIPESCEL     151
226900      QCDNSNPHYD----------SGGYFGYTEKFGNHPGGQAEYLKVPFGNFTPFVIPESCEL     151
405532      QCDNSNPHYD----------SGGYFGYTEKFGNHPGGQAEYLKVPFGNFTPFVIPESCEL     151
66692       QCDRSNPHYD----------SGGLFGYSEKFGDYPGGQAEYLRVPFGNYTPFRLPDDCEL     151
315750      QCDNSNPHYD----------SGGLFGYSEKYGNYPGGQAEYLRVPFGNYTPFKIPDDCEL     151
224308      QCDQSNDNGE----------MGAYFGYSGQTGGYPGGQAEYLRVPFANFTHFKIPESCEE     152
326423      QCDQSNENGD----------IGAYFGYSGNAGGYPGGQAEYLRVPFANFTHFKVPETCEE     152
272558      QCDNSNPNGQ----------AGAYFGYSGTFGGYPGGQAELLRVPYGNFTPFVVPEDAEM     152
235909      QCDASNPNGE----------SGGYFGYSETFGGYPGGQAEYLRVPFANFTPFVVPENCEL     152
420246      ICRHKQSLG----------------------INVDGGFSHEFVITSK--YVLSIPDDLPD     137
491915      LCEHQQNAG----------------------YSVDGGYAEYCRAAAD--YVVKIPDNLSF     140
282458      LCNHRKGIG----------------------TQVDGAFTNYVIAREE--SLHHIPDEVSY     139
272626      LCSNRRGIG----------------------TQANGSFAEFVLSREE--SCHVLDERISL     140
169963      LCSNRRGIG----------------------TQANGSFAEFVLSREE--SCHVLDERISL     140
221109      RCVELTVIG----------------------ARNPGAYAEYTKLPEG--HVKILPDSIDD     137
195103      LCPDVIFFAT---------------------PPVDGVFQEYVAHEAD--LCFKLPENVST     145
315749      -----TLDAV---------------G-----AQIDGGFAEYAILPQK--TIFKIPDTLPL     134
413999      LCSSLGFHGL---------------------CGSGGGLAEYTVFPEE--FVHKIPDEMSY     149
101031      ICKHLGFHGL---------------------SGGGGGFSEYTMVDEK--LVHKMPEGLSY     149
315730      VCEQLVFHGL---------------------GGEGGGFSEYTVVPED--MVHHIPDEMTY     148
                                                  *   .                .
```

```
176279      KEDALVMLSDILPTGYEIGVLKGKVKPGCTVAIVGAGPVGLAALLTAQFYSPSKIIM-ID    199
482703      DDEQVLFLTDILPTGYQ-AAVNAGVRPGSRVAIYGAGPVGLMSAACARMLGAEQIFM-VD    218
279010      EDESLLFLSDVLPTAYW-SVLHAGVKKGDTVIVLGCGPVGLMAQKFAWMEGAKRVIA-VD    209
281309      EDESLLFLSDVLPTAYW-SVINAGVKRGDTVIVLGCGPVGLMTQKFAWMQGAKRVIA-VD    209
412694      EDESLLFLSDVLPTAYW-SVINAGVKRGDTVIVLGCGPVGLMTQKFAWMQGAKRVIA-VD    209
222523      EDESLLFLSDVLPTAYW-SVINAGVKRGDTVIVLGCGPVGLMTQKFAWMQGAKRVIA-VD    209
261594      EDESLLFLSDVLPTAYW-SVINAGVKPGDTVIVLGCGPVGLMTQKFAWMHGAKRVIA-VD    209
198094      EDESLLFLSDVLPTAYW-SVINAGVKPGDTVIVLGCGPVGLMTQKFAWMHGAKRVIA-VD    209
280477      EDESLLFLSDVLPTAYW-SVINAGVKPGDTVIVLGCGPVGLMTQKFAWMHGAKRVIA-VD    209
280354      EDESLLFLSDVLPTAYW-SVINAGVKPGDTVIVLGCGPVGLMTQKFAWMHGAKRVIA-VD    209
205919      EDESLLFLSDVLPTAYW-SVINAGVKPGDTVIVLGCGPVGLMTQKFAWMHGAKRVIA-VD    209
260799      EDESLLFLSDVLPTAYW-SVINAGVKPGDTVIVLGCGPVGLMTQKFAWMHGAKRVIA-VD    209
261591      EDESLLFLSDVLPTAYW-SVINAGVKPGDTVIVLGCGPVGLMTQKFAWMHGAKRVIA-VD    209
280355      EDESLLFLSDVLPTAYW-SVINAGVKPGDTVIVLGCGPVGLMTQKFAWMHGAKRVIA-VD    209
269801      EDESLLFLSDVLPTAYW-SVINAGVKPGDTVIVLGCGPVGLMTQKFAWMQGAKRVIA-VD    209
CpsC        EDESLLFLSDVLPTAYW-SVINAGVKPGDTVIVLGCGPVGLMTQKFAWMQGAKRVIA-VD    209
288681      EDESLLFLSDVLPTAYW-SVINAGVKPGDTVIVLGCGPVGLMTQKFAWMQGAKRVIA-VD    209
226900      EDESLLFLSDVLPTAYW-SVINAGVRPGDTVIVLGCGPVGLMTQKFAWMQGAKRVIA-VD    209
405532      EDESLLFLSDVLPTAYW-SVINAGVKPGDTVIVLGCGPVGLMTQKFAWMQGAKRVIA-VD    209
66692       EDEQVLFLSDVLPTAYW-SVVHAGVKKGDTVIVLGCGPVGLMTQAFAWQQGAERVIA-VD    209
315750      EDEQLLFLSDVLPTAYW-SVEHAGVKKGDTVIVLGCGPVGLMAQQFAWQKGAERVIA-VD    209
224308      PDEKLSVIADAMTTGFW-SVDNAGVKKGDTVIVLGCGPVGLFAQKFCWLKGAKRVIA-VD    210
326423      PDEKLSVIADAMTTGFW-SVDNAGVKEGDTVIVLGCGPVGLFAQKFCWLKGAKRVIA-VD    210
272558      EDEKLLFLSDIIPTAYW-GVEQAEVKKGDTVVVLGCGPVGLLAQKFAWLKGAERVIA-VD    210
235909      EDEKLLFLSDIIPTAFW-GVDEAGVKDGDTVVVLGCGPVGLLTQKFAWMKGAKRVIA-VD    210
420246      EKAVLVEPFAV---VVHA-LQKVKIKQGMKVAIIGCGNEGMMAAVLARYLGADVTAS--D    191
491915      QEAAPIFCAGV---TTYKALKVTGAKPGEWVAIYGIGGLGHVAVQYAKAMGLNVVA--VD    195
282458      QSAAMTEPLAC---AHHG-VSKIQVNSGDVAVVMGPGPIGLLVAQVLKSKGATVVVTGLD    195
272626      EAAALTEPLAC---CVHSALEKTTICPDDTVLVFGPGPIGLLLAQVVKAQGATVIMAGIT    197
169963      EAAALTEPLAC---CVHSALEKTTIRPDDTVLVFGPGPIGLLLAQVVKAQGATVIMAGIT    197
221109      DTAALVEPSAV---VAHG-FYRASITPGSSVAIMGVGSIGLLAVQWAKIFGASKVIA-ID    192
195103      LEGALIEPLAV---GFHA-AIQGGARIGQTAVVMGAGCIGLVSMMALKAMGVSNVYI-VD    200
315749      ELAAMVEPAAV---SFQS-VKDSNVKEGDTIVVYGAGPIGLFAVLGAKAAGVSNIIV-VD    189
413999      EDAALVEPMAV---ALHS-ARIANFNTGDTALVLGAGPIGLATIECLKAAGARLIVV-LQ    204
101031      EQGALVEPAAV---ALHA-VRQSKLKAGDKAAVFGTGPIGLLVIEALRAAGASEIYA-VE    204
315730      EQGALVEPAAV---AVHA-VRQSKLKEGEAVAVFGCGPIGLLVIQAAKAAGATPVIA-VE    203
                                  .    : * *  *          .
```

```
176279     LDDNRLETAKELGATHLINSKETETAIKKVKSLNP--RGVDVAIEAVGIPQTFDL-----    252
482703     HHDYRLTYAQQTYGVIPINFDVIDDPASAIIEQTPGHRGVDAVIDAVGFEAKGSLTETVL    278
279010     YLDYRLKQAEALNRVEVFDFTKYPDMGEHLKEITK--GGADVVIDCVGMDGKKSPLEYLE    267
281309     YLDYRIHYAKKINNVEVFEFTKFPDMGEHLKEITH--GGADVVIDCVGMDGKKSPLEFLE    267
412694     YLDYRIHYAKKINNVEVFEFTKFPDMGEHLKEITH--GGADVVIDCVGMDGKKSPLEFLE    267
222523     YLDYRINYAKKINNVEVFEFTKFPDMGEHLKEITH--GGADVVIDCVGMDGKKSPLEFLE    267
261594     YLDYRINYAKKINNVEVFEFTKFPDMGEHLKEITH--GGADVVIDCVGMDGKKSPLEFLE    267
198094     YLDYRINYAKKINNVEVFEFTKFPDMGEHLKEITH--GGADVVIDCVGMDGKKSPLEFLE    267
280477     YLDYRINYAKKINNVEVFEFTKFPDMGEHLKEITH--GGADVVIDCVGMDGKKSPLEFLE    267
280354     YLDYRINYAKKINNVEVFEFTKFPDMGEHLKEITH--GGADVVIDCVGMDGKKSPLEFLE    267
205919     YLDYRINYAKKINNVEVFEFTKFPDMGEHLKEITH--GGADVVIDCVGMDGKKSPLEFLE    267
260799     YLDYRINYAKKINNVEVFEFTKFPDMGEHLKEITH--GGADVVIDCVGMDGKKSPLEFLE    267
261591     YLDYRINYAKKINNVEVFEFTKFPDMGEHLKEITH--GGADVVIDCVGMDGKKSPLEFLE    267
280355     YLDYRINYAKKINNVEVFEFTKFPDMGEHLKEITH--GGADVVIDCVGMDGKKSPLEFLE    267
269801     YLDYRMNYAKKINNVEIFEFTKFPDMGEHLKEITH--GGADVVIDCVGMDGKKSPLEFLE    267
CpsC       YLDYRMNYAKKINNVEVFEFTKFPDMGEHLKEITH--GGADVVIDCVGMDGKKSPLEFLE    267
288681     YLDYRMNYAKKINNVEVFEFTKFPDMGEHLKEITH--GGADVVIDCVGMDGKKSPLEFLE    267
226900     YLDYRMNYAKKINNVEVFEFTKFPDMGEHLKEITH--GGADVVIDCVGMDGKKSPLEFLE    267
405532     YLDYRMNYAKKINNVEVFEFTKFPDMGEHLKEITH--GGADVVIDCVGMDGKKSPLEFLE    267
66692      YFDYRLEKARQMNHTETFDFTKDKDTGETLKELTK--GGADVVIDCVGMDGKKSPLEMVE    267
315750     YIDYRLRHAKQMSGVEVFDFTEDPDMGETLKELTK--GGADVVIDCVGMDGKKSPLEKIE    267
224308     YVNYRLQHAKRTNKVEIVNFEDHENTGNYLKEITK--GGADVVIDAVGMDGKMSDLEFLA    268
326423     YVYYRLQHAKRTNKVEIVNFEDHENTGSYLKEITK--GGADVVIDAVGMDGKMTDLEFLA    268
272558     YVQYRLTHAKQTNQVEVLDFSTIDDVGTHIKEITK--GGADSVIDCVGMDGKKSAVEMVE    268
235909     YIDYRIEHAKKTNKVETLNFTEYENVGEYIKEMTG--GGADVVIDCVGLDGKMTPLELIG    268
420246     INPLKLETVKTMADIRTL-----------APSALGN-ESFDVVIEAAGTREA--------    231
491915     LGDEKLELAKQLGADLIVNPKH-EDAAQWMKEKVGG-VHAA--VVTAVSKTA--------    243
282458     NDKVRLDKAEALHMDYVVNLQQ-TDLKTYINGITDG-YGADVVVECSGAVPA--------    245
272626     KDSDRLRLAKELGMDRIVDTLK-EDLAEVVLGMTDG-YGAERVFDCSGAVPA--------    247
169963     KDSDRLRLAKELGMDRIVDTLK-EDLAEVVLGMTGG-YGAERVFDCSGAVPA--------    247
221109     IDDHKLKIAQELGADIVVNPMK-ENAEGIVKQHTDE-LGVDLAVESAGSPIT--------    242
195103     IMEKRLEKALELGATGIINAKE-KNAIEEVMKITNN-NGCDLVIETAGTEIT--------    250
315749     LLDSRLDKATELGATHVFNARE-VNPVEEIRKLFPD--GADVVFEAAGVEST--------    238
413999     RKSIRQKYAKRAGADVVLDPNE-VNIPEEVKKLTDG-LGVDVAFETTGAKIG--------    254
101031     LSAERAAKALEIGATAVINPKD-EDAVARLHELTNG--GVDVAFEVTGVPVV--------    253
315730     LSKERQELAKLAGADYVLNPAT-QDVLAEIRNLTNG-LGVNVSFEVTGVEVV--------    253
               :   .        .                          .
```

```
176279     ---------------CQNLIGVDGTIANVGVHGL-P-VQLDI-DKLWIKNINVTTGL---    291
482703     TTLKLEASSGVALRQCIAAVRRGGTVSVPGVYAGFI-HGFLF-GDAFDKGLTFKMGQ---    333
279010     QKLKLQGGTLGPIQISTKAVRKCGTVQITGVYGSNY-NMFPL-GAFFSRNVTLKMGQ---    322
281309     QKLKLQGGTLGPIQIATKAVRKYGTVQMTGVYGGNY-NAFPL-GAFWVRNINLKMGQ---    322
412694     QKLKLQGGTLGPIQIATKAVRKYGTVQMTGVYGGNY-NAFPL-GAFWVRNINLKMGQ---    322
222523     QKLKLQGGTLGPIQIATKAVRKYGTVQMTGVYGGNY-NAFPL-GAFWVRNINLKMGQ---    322
261594     QKLKLQGGTLGPIQIATKAVRKYGTVQMTGVYGGNY-NAFPL-GAFWVRNINLKMGQ---    322
198094     QKLKLQGGTLGPIQIATKAVRKYGTVQMTGVYGGNY-NAFPL-GAFWVRNINLKMGQ---    322
280477     QKLKLQGGTLGPIQIATKAVRKYGTVQMTGVYGGNY-NAFPL-GAFWVRNINLKMGQ---    322
280354     QKLKLQGGTLGPIQIATKAVRKYGTVQMTGVYGGNY-NAFPL-GAFWVRNINLKMGQ---    322
205919     QKLKLQGGTLGPIQIATKAVRKYGTVQMTGVYGGNY-NAFPL-GAFWVRNINLKMGQ---    322
260799     QKLKLQGGTLGPIQIATKAVRKYGTVQMTGVYGGNY-NAFPL-GAFWVRNINLKMGQ---    322
261591     QKLKLQGGTLGPIQIATKAVRKYGTVQMTGVYGGNY-NAFPL-GAFWVRNINLKMGQ---    322
280355     QKLKLQGGTLGPIQIATKAVRKYGTVQMTGVYGGNY-NAFPL-GAFWVRNINLKMGQ---    322
269801     QKLKLQGGTLGPIQIATKAVRKYGTVQMTGVYGGNY-NAFPL-GAFWVRNINLKMGQ---    322
CpsC       QKLKLQGGTLGPIQIATKAVRKYGTVQMTGVYGGNY-NAFPL-GAFWVRNINLKMGQ---    322
288681     QKLKLQGGTLGPIQIATKAVRKYGTVQMTGVYGGNY-NAFPL-GAFWVRNINLKMGQ---    322
226900     QKLKLQGGTLGPIQIATKAVRKYGTVQMTGVYGGNY-NAFPL-GAFWVRNINLKMGQ---    322
405532     QKLKLQGGTLGPIQIATKAVRKYGTVQMTGVYGGNY-NAFPL-GAFWVRNINLKMGQ---    322
66692      QKLKLQGGTLGPIQIATKAVKKCGIVQLTGVYGGNY-NLFPL-GAFFSRNVTLKMGQ---    322
315750     QKLKLQGGTIGPIQIATKAVRKCGTVQMTGVYGGLY-NMFPL-GAFFARNVTLKMGQ---    322
224308     SGLKLHGGTMSALVIASQAVRKGGTIQITGVYGGRY-NGFPL-GDIMQRNVNIRSGQ---    323
326423     SGLKLQGGAMSALVIASQAVRKGGTIQITGVYGGRY-NGFPL-GDIMQRNVNIRSGQ---    323
272558     TALKLQGGTLGPINIASQAVRKGGTVSIVGVYGTRY-NAFPL-GDFFARNITLKMGQ---    323
235909     SALKLQGGAMGAIVIASQAVRKGGTIQLVGVYGARY-NQFPL-GDLFSRNITLKMGQ---    323
420246     ------------VEQAVELVAPGGHLVLIGFAN-EA--TFPV-VRLVRNEVTVHGSV---    272
491915     ------------FESAYKAIRRGGACVLVGLPPEE--MPVPIFDTVLN-GVKIIGSI---    285
282458     ------------ARQGLDILRKKGFYSQIGIFK-DAEITFDM-EKVIQKEITVVGSR---    288
272626     ------------VNQGLPLTKKKGDFIQVGLFA-EKKNAIDE-ESIIQREIAYIGSR---    290
169963     ------------VNQGLPLTKKKGDFVQVGLFA-EKKNAIDE-ESIIQREIAYIGSR---    290
221109     ------------SEQVLALPKKGGEVVFLGIPYGDITLKRYYFEKIVRNELRILGSWNAL    290
195103     ------------TVQAIHMAKKGSNIVLVG-YSKSGEMTLPM-SLVLDKELTFKTVF---    293
315749     ------------FNQAIQSTKVRGTMMVISFHTQDIQFNAPS-SLLFS-GVKLMGSV---    281
413999     ------------FDTGIESLKFEGTMVITSIWEKDTSFN-PN-VLVFT-EKKIVGTL---    296
101031     ------------LQQAIDSTTFEGETIIVSIWETDASIL-PN-NIVLT-ERSVKGII---    295
315730     ------------LRQAIESTSFEGQTVIVSVWEKDATIT-PN-NLVLK-EKEVIGIL---    295
```

. .

```
176279     --VSG--NTTEELLEALKSKIIQPEQLVTHYSKLSEIESAYDLFRNAT-DHKAIKLIIEN    346
482703     --THV-HPLLPTLLEHIQRGDLNPEIIISHRMPLAEAAEGYRLFDSRR-EQ-CRKVILRP    388
279010     --APV-IHLMPEIYKKIEENQFDPKEIITHQLPLEEAGRAYHLFNDHE-DD-CIKVILKP    377
281309     --APV-IHFMPELFKKITNKEFDPKEIITHKIPLEEASYGYQIFNNRE-DD-CIKVILKP    377
412694     --APV-IHFMPELFKKITNKEFDPKEIITHKIPLEEASYGYQIFNNRE-DD-CIKVILKP    377
222523     --APV-IHFMPELFEKITNKEFDPKEIITHKIPLEEASYGYQIFNNRE-DD-CIKVILKP    377
261594     --APV-IHFMPELFEKITNKEFDPKEIITHKIPLEEASYGYQIFNNRE-DD-CIKVILKP    377
198094     --APV-IHFMPELFEKITNKEFDPKEIITHKIPLEEASYGYQIFNNRE-DD-CIKVILKP    377
280477     --APV-IHFMPELFEKITNKEFDPKEIITHKIPLEEASYGYQIFNNRE-DD-CIKVILKP    377
280354     --APV-IHFMPELFEKITNKEFDPKEIITHKIPLEEASYGYQIFNNRE-DD-CIKVILKP    377
205919     --APV-IHFMPELFEKITNKEFDPKEIITHKIPLEEASYGYQIFNNRE-DD-CIKVILKP    377
260799     --APV-IHFMPELFEKITNKEFDPKEIITHKIPLEEASYGYQIFNNRE-DD-CIKVILKP    377
261591     --APV-IHFMPELFEKITNKEFDPKEIITHKIPLEEASYGYQIFNNRE-DD-CIKVILKP    377
280355     --APV-IHFMPELFEKITNKEFDPKEIITHKIPLEEASYGYQIFNNRE-DD-CIKVILKP    377
269801     --APV-IHFMPELFEKITNKEFDPKEIITHKIPLEEASYGYQIFNNRE-DD-CIKVILKP    377
CpsC       --APV-IHFMPELFEKITNKEFDPKEIITHKIPLEEASYGYQIFNNRE-DD-CIKVILKP    377
288681     --APV-IHFMPELFEKITNKEFDPKEIITHKIPLEEASYGYQIFNNRE-DD-CIKVILKP    377
226900     --APV-IHFMPELFEKITNKEFDPKEIITHKIPLEEASYGYQIFNNRE-DD-CIKVILKP    377
405532     --APV-IHFMPELFEKITNKEFDPKEIITHKIPLEEASYGYQIFNNRE-DD-CIKVILKP    377
66692      --APA-RAYMETLYEQISKGLINPTSIITHKLALSDAPKGYDLFNNKK-DG-CIKVVLKP    377
315750     --APA-RGYMSKLYQKVTTGEIDPRAIITHQLPLDDAAHAYHIFNDKK-DD-CIKVILKP    377
224308     --APV-IHYMPYMFELVSTGKIDPGDVVSHVLPLSEAKHGYDIFDSKM-DD-CIKVVLKP    378
326423     --APV-IHYMPYMFELVSTGKIDPGDVVSHVIPLSEAKHGYEMFDTKT-DD-CIKVVLKP    378
272558     --API-IHYMPTLYQMIKDEVFDPTDLITHRLPLELAEHGYDIFDAKQ-DD-CIKVVLKP    378
235909     --APV-IHYIPTLYEWIVEGKFDPTDIITHRLPLDEAQYAYEIFDEKK-DG-CIKVVLKP    378
420246     --IYRFPDDYLQAIHYLRTMPYPIERVISCIFPVRDYQQAYELASSG----NMCKVVLSF    326
491915     --VGT-RKDLQEALQFAAEGKVKTIVEV---QPLENINDVFDRMLKGQ--IN-GRVVLKV    336
282458     --SQK-PADWEPSLQLMADGLVNAEALVTKIYDISKWDEAYQHLKSG----EGIKALLKP    341
272626     --SQK-PSSWILALDLLANGKINTDKMITKVYGLDDWREAFEAVMAG----NEIKVLVKS    343
169963     --SQK-PSSWILALDLLANGKIDTDKMITKVYGLDDWREAFEAVMAG----NEIKVLVKS    343
221109     SAPFP-GKEWNATLHYMSTGQLNIKPMISHRLGLQAGPEIFHQITNK--LSDAVKVIFHP    347
195103     --RYR-H-IYNMAIEAVASGKVNLKGIITNEFDLDDVQKAMDYSVNN--KADIVKAVIKI    347
315749     --GYS-NETYNEVIELLANGRLPAQSIITSKVDIDNIAEQGFEALIH--DKSQAKILVKL    336
413999     --AYR-H-EFPATMALMKDGRIKTDGYITKKIALDDIVEEGFGALTGPEKKKHVKIIVTP    352
101031     --AYR-D-IFPAVMELMKQGYFPADKLVTKRIALEEVVTEGFEALMK--EKNHIKILVNS    349
315730     --GYR-H-IFPAVIKLISSGQIQAEKLITKKITVDQVVEEGFEALVK--DKTQVKILVSP    349
                                     :     :                    : :.
```

```
CpsC        ----------   377
176279      DITI------   350
482703      ----------   388
279010      ----------   377
281309      ----------   377
412694      ----------   377
222523      ----------   377
261594      ----------   377
198094      ----------   377
280477      ----------   377
280354      ----------   377
205919      ----------   377
260799      ----------   377
261591      ----------   377
280355      ----------   377
269801      ----------   377
288681      ----------   377
226900      ----------   377
405532      ----------   377
66692       ----------   377
315750      ----------   377
224308      ----------   378
326423      ----------   378
272558      ----------   378
235909      ----------   378
420246      KEEEER----   332
491915      D---------   337
282458      LDLDENEGEN   351
272626      ----------   343
169963      ----------   343
221109      GK--------   349
195103      R---------   348
315749      SGAH------   340
413999      DKSLL-----   357
101031      QA--------   351
315730      K---------   350
```

Fig. S5: Multiple sequences alignments between of CpsC from *Bacillus tropicus* AOA-CPS1 with closely related neighbourhoods' whole-genome. Where the sequences are similar were shaded in pink and indicated with asterisk (*). L-lactate dehydrogenase (EC 1.1.1.27) from *Clostridium botulinum* A str. ATCC 3502 (413999), *C. perfringens* ATCC 13124 (195103), *Listeria innocua* Clip11262 (272626), *Listeria monocytogenes* EGD-e (169963), *Staphylococcus aureus* subsp. *aureus* MRSA252 (282458) and malate dehydrogenase (EC 1.1.1.37) from *S. epidermidis* RP62A (176279), *B. thuringiensis* serovar konkukian str. 97-27 (281309), *B. subtilis* subsp subtilis str. 168 (224308), *B. cereus* ATCC 14579 (226900), *B. tropicus* AOA-CPS1 (CpsE), *B.* anthracis str. Ames (198094), *B.clausii* KSM-K16 (66692), *B. halodurans* C-125 (272558), *B. licheniformis* ATCC 14580 (279010), *Anoxybacillus flavithermus* WK1 (491915), *Geobacillus kaustophilus* HTA426 (235909), *G. thermodenitrificans* NG80-2 (420246), *Oceanobacillus iheyensis* HTE831 (221109), *B. amyloliquefaciens* FZB42 (326423), *B. anthracis* str. A1055 (280355), *B. anthracis* str. Australia 94 (280477), *B. anthracis* str. CNEVA-9066 (280354), *B. anthracis* str. Kruger B (205919), *B. anthracis* str. Sterne (260799), *B. anthracis* Vollum (261591), *Bacillus* B-14905 (101031), *B. cereus* ATCC 10987 (222523), *B. cereus* B4264 (405532), *B. cereus* e33L (288681), *B. cereus* G9241 (269801), *B. weihenstephanensis* KBAB4 (315730), *B. thuringiensis* str. Al Hakam (412694), *B. pumilus* SAFR-032 (315750), and *B. cereus* subsp. cytotoxis NVH 391-98 (315749).

```
             :::*:***:*:*:.*::* .*:  :**::****:*** :   ****:.*: .. *:::
CpsA         MNQLKGIHHVTAITSSAEKNYEFFTHVLGMRLVKKTVNQDDIQTYHLFFADDKGSAGTDM     60
101031       MNHLKGVHHVTAITSSAEENYKFFTYVLGMRLVKKTVNQDDIQTYHLFFADDKGSAGTDM     60
315730       MNQLKGIHHVTAITSSAEKNYEFFTHVLGMRLVKKTVNQDDIQTYHLFFADDKGSAGTDM     60
226900       MNQLKGIHHVTAITSSAEKNYEFFTHVLGMRLVKKTVNQDDIQTYHLFFADDKGSAGTDM     60
269801       MNQLKGIHHVTAITSSAEKNYEFFTHVLGIRLVKKTVNQADIQTYHLFFADDKGSAGTDM     60
405532       MNQLKGIHHVTAITSSAEKNYEFFTHVLGMRLVKKTVNQDDIQTYHLFFADDKGSAGTDM     60
412694       MNHLKGIHHVTAITSSAEKNYEFFTHVLGMRLVKKTVNQDDIQTYHLFFADDKGSAGTDM     60
280355       MNQLKGIHHVTAITSSAEKNYEFFTHILGMRLVKKTVNQDDIQTYHLFFADDKGSAGTDM     60
280477       MNQLKGIHHVTAITSSAEKNYEFFTHILGMRLVKKTVNQDDIQTYHLFFADDKGSAGTDM     60
280354       MNQLKGIHHVTAITSSAEKNYEFFTHILGMRLVKKTVNQDDIQTYHLFFADDKGSAGTDM     60
205919       MNQLKGIHHVTAITSSAEKNYEFFTHILGMRLVKKTVNQDDIQTYHLFFADDKGSAGTDM     60
260799       MNQLKGIHHVTAITSSAEKNYEFFTHILGMRLVKKTVNQDDIQTYHLFFADDKGSAGTDM     60
261591       MNQLKGIHHVTAITSSAEKNYEFFTHILGMRLVKKTVNQDDIQTYHLFFADDKGSAGTDM     60
198094       MNQLKGIHHVTAITSSAEKNYEFFTHILGMRLVKKTVNQDDIQTYHLFFADDKGSAGTDM     60
261594       MNQLKGIHHVTAITSSAEKNYEFFTHILGMRLVKKTVNQDDIQTYHLFFADDKGSAGTDM     60
288681       MNQLKGIHHVTAITSSAEKNYEFFTHVLGMRLVKKTVNQDDIQTYHLFFADDKGSAGTDM     60
222523       MNQLKGIHHVTAITSSAEKNYEFFTHVLGMRLVKKTVNQDDIQTYHLFFADDKGSAGTDM     60
281309       MNQLKGIHHVTAITSSAEKNYEFFTHVLGMRLVKKTVNQDDIQTYHLFFADDKGSAGTDM     60
315750       -MKINGIHHVSALTANAQKNVDFYRQILGLKLVKKTVNQDDPSMYHLFYGDEVASPGTEL     59
326423       -MKVNGIHHVSALTADAQKNLDFYRNILGLKLVKKSVNQDEPTMYHLFYGDETANPGSEL     59
224308       -MKVNGIHHVSALTADAQKNLDFYKKVLGLKLVKKSVNQDEPTMYHLFYGDEVANPGTEL     59
             :::*:***:*:*:.*::* .*:  :**::****:*** :   ****:.*: .. *:::


             ***::* :     *** *:. .: *    :* :* :**:.  * :       *:  * *
CpsA         TFFDFPGIPKGVHGTNEISKTSFRVPTDAALAYWVKRFDRLEVEHTGIKEQFGKKTLSFV    120
101031       TFFDFPNIPKGVHGTNEISKTSFRVPTDASLDYWEKRFDRLHVKHTGIKEQFGKKTLSFV    120
315730       TFFDFPGVPKGVHGTNEISKTSFRVPTDASLAYWVNRFDRLEVEHTGIKEQFGKKTLSFV    120
226900       TFFDFPGVPKGVHGTNEISKTSFRVPTDAALAYWVKRFDRLEVEHTGIKEQFGKKTLSFV    120
269801       TFFDFPGIPKGVHGTNEISKTSFRVPTDAALAYWVKRFDRLEVEHTGIKEQFGKKTLSFV    120
405532       TFFDFPGVPKGVHGTNEISKTSFRVPTDAALAYWVKRFDRLEVEHTGIKEQFGKKILSFV    120
412694       TFFDFPGIPKGVHGTNEISKTSFRVPTDAALAYWVKRFDRLEVQHTGIKEQFGKKTLSFV    120
280355       TFFDFPGIPKGVHGTNEISKTSFRVPTDAALAYWVKRFDRLEVEHTGIKEQFGKKTLSFV    120
280477       TFFDFPGIPKGVHGTNEISKTSFRVPTDAALAYWVKRFDRLEVEHTGIKEQFGKKTLSFV    120
280354       TFFDFPGIPKGVHGTNEISKTSFRVPTDAALAYWVKRFDRLEVEHTGIKEQFGKKTLSFV    120
205919       TFFDFPGIPKGVHGTNEISKTSFRVPTDAALAYWVKRFDRLEVEHTGIKEQFGKKTLSFV    120
260799       TFFDFPGIPKGVHGTNEISKTSFRVPTDAALAYWVKRFDRLEVEHTGIKEQFGKKTLSFV    120
261591       TFFDFPGIPKGVHGTNEISKTSFRVPTDAALAYWVKRFDRLEVEHTGIKEQFGKKTLSFV    120
198094       TFFDFPGIPKGVHGTNEISKTSFRVPTDAALAYWVKRFDRLEVEHTGIKEQFGKKTLSFV    120
261594       TFFDFPGIPKGVHGTNEISKTSFRVPTDAALAYWVKRFDRLEVEHTGIKEQFGKKTLSFV    120
288681       TFFDFPGIPKGVHGTNEISKTSFRVPTDAALEYWVKRFDRLEVEHTGIKEQFGKKTLSFV    120
222523       TFFDFPGIPKGVHGTNEISKTSFRVPTDAALEYWVKRFDRLEVEHTGIKEQFGKKTLSFV    120
281309       TFFDFPGIPKGVHGTNEISKTSFRVPTDAALEYWVKRFDRLEVEHTGIKEQFGKKTLSFV    120
315750       TFFEIPMLARRLKGTNAITSTGLLVSSEEALSFWEKRFEETGVQQETASLRGGRPALRFQ    119
326423       TFFEIPRIAPFHAGTNSISSIGLRVPDAQALQYWKKRFEEHQVAHGDIMTIAGRETLAFW    119
224308       TFFEIPRIAPFHAGTNSISSIGLRVPGTEALHYWKERFEEQQVTHSGISKRAGRDILAFQ    119
             ***::* :     *** *:. .: *    :* :* :**:.  * :       *:  * *
```

```
           * :.*:  *  :    .   *    .. :* *::: ***** : :   .    ** .::
CpsA       DFDDQHYQLISDELDKGIESGTPWQKGPVPLEYAITGLGPVFIRIANFSFFKEVLEKVLL    180
101031     DFDDQQYQLISDEHDTGIASGTPWQDGPIPLEFAITGLGPVFVRVADFNYFKDVLEKVML    180
315730     DFDDQHYQLISDELDNGIESGTPWQNGPVPLEYAITGLGPVFIRIANFSFFKEVLEKVLL    180
226900     DFDDQHYQLISDELDKGIESGTPWQNGPVPLEYAITGLGPVFIRIANFSFFKEVLEKVLL    180
269801     DFDDQHYQLISDELDKGIESGTPWQNGPVPLEYAITGLGPVFIRIANFSFFKEVLEKVLL    180
405532     DFDDQHYQLISDELDKGIESGTPWQNGPVPLEYAITGLGPVFIRIANFSFFKEVLEKVLL    180
412694     DFDDQHYQLISDELDKGIESGTPWQNGPVPLEYAITGLGPVFIRIANFSFFKEVLEKVLL    180
280355     DFDDQHYQLISDELDKGIESGTPWQKGPVPLEYAITGLGPVFIRIANFSFFKEVLEKVLL    180
280477     DFDDQHYQLISDELDKGIESGTPWQKGPVPLEYAITGLGPVFIRIANFSFFKEVLEKVLL    180
280354     DFDDQHYQLISDELDKGIESGTPWQKGPVPLEYAITGLGPVFIRIANFSFFKEVLEKVLL    180
205919     DFDDQHYQLISDELDKGIESGTPWQKGPVPLEYAITGLGPVFIRIANFSFFKEVLEKVLL    180
260799     DFDDQHYQLISDELDKGIESGTPWQKGPVPLEYAITGLGPVFIRIANFSFFKEVLEKVLL    180
261591     DFDDQHYQLISDELDKGIESGTPWQKGPVPLEYAITGLGPVFIRIANFSFFKEVLEKVLL    180
198094     DFDDQHYQLISDELDKGIESGTPWQKGPVPLEYAITGLGPVFIRIANFSFFKEVLEKVLL    180
261594     DFDDQHYQLISDELDKGIESGTPWQKGPVPLEYAITGLGPVFIRIANFSFFKEVLEKVLL    180
288681     DFDDQHYQLISDELDKGIESGTPWQNGPVPLEYAITGLGPVFIRIANFSFFKEVLEKVLL    180
222523     DFDDQHYQLISDELDKGIESGTPWQKGPVPLEYAITGLGPVFIRIANFSFFKEVLEKVLL    180
281309     DFDDQHYQLISDELDKGIESGTPWQNGPVPLEYAITGLGPVFIRIANFSFFKEVLEKVLL    180
315750     DPEGQQLFLTVEPA--AQEAGSPPVHDDIPAEFSIRGLGPVELTVAEADKTIRVLTDILG    177
326423     DHEGQRLVLTADEK--GKDIGEAVKQSGIPEEYSFRGLGPVELTVPYAEPTLRVLTDVLG    177
224308     DHEGQRLVLTADEE--GKGYGLPVKQSGIPEEFSFRGLGPVELTVPYAEPTLHVLTNILG    177
           * :.*:  *  :    .   *    .. :* *::: ***** : :   .    ** .::


           *.* ..:         ::* .:**  :.: : ...  *  : * *:***.****.:   *
CpsA       FKEVAQE-----GEFYLFEVNEGGNGASVIVEHNTVLPEARQGFGTVHHAAFRVEDRAVL    235
101031     FTEIGED-----GHFHLFEVGEGGNGAQVIVEHNTVLPVARQGFGTVHHAAFRVEDRAVL    235
315730     FKEIAQE-----GEFYLFEVNEGGNGASVIVEHNTVLPEAQQGFGTVHHAAFRVEDRAVL    235
226900     FKEIAQE-----GEFYLFEVNKGGNGASVIVEHNTVLPEAQQGFGTVHHAAFRVEDRAVL    235
269801     FKEVAQE-----GEFYLFEVNKGGNGASVIVEHNTVLPEARQGFGTVHHAAFRVENRAVL    235
405532     FKEVAQE-----GEFYLFEVNKGGNGASVIVEHNTVLPEARQGFGTVHHAAFRVEDRAVL    235
412694     FKEVAQE-----GEFYLFEVNEGGNGASVIVEHNTVLPEARQGFGTVHHAAFRVEDRAVL    235
280355     FKEVAQE-----GEFYLFEVNEGGNGASVIVEHNTVLPEARQGFGTVHHAAFRVEDRAVL    235
280477     FKEVAQE-----GEFYLFEVNEGGNGASVIVEHNTVLPEARQGFGTVHHAAFRVEDRAVL    235
280354     FKEVAQE-----GEFYLFEVNEGGNGASVIVEHNTVLPEARQGFGTVHHAAFRVEDRAVL    235
205919     FKEVAQE-----GEFYLFEVNEGGNGASVIVEHNTVLPEARQGFGTVHHAAFRVEDRAVL    235
260799     FKEVAQE-----GEFYLFEVNEGGNGASVIVEHNTVLPEARQGFGTVHHAAFRVEDRAVL    235
261591     FKEVAQE-----GEFYLFEVNEGGNGASVIVEHNTVLPEARQGFGTVHHAAFRVEDRAVL    235
198094     FKEVAQE-----GEFYLFEVNEGGNGASVIVEHNTVLPEARQGFGTVHHAAFRVEDRAVL    235
261594     FKEVAQE-----GEFYLFEVNEGGNGASVIVEHNTVLPEARQGFGTVHHAAFRVEDRAVL    235
288681     FKEVAQE-----GEFYLFEVNEGGNGASVIVEHNTVLPEARQGFGTVHHAAFRVEDRAVL    235
222523     FKEVAQE-----GEFYLFEVNEGGNGASVIVEHNTVLPEARQGFGTVHHAAFRVENRAVL    235
281309     FKEVAQE-----GEFYLFEVNEGGNGASVIVEHNTVLPEARQGFGTVHHAAFRVENRAVL    235
315750     FTERAREQSNDGEIVTIFESGNGGAGTEIHVIEKTEGPSERSGRGSVHHVAFRVENAEEL    237
326423     FTEIGKETHEDRGTVTVLESGNGGARTEVHLIERNDLPRERPGKGSVHHVAFRVQNEEEL    237
224308     FTEISREPVEGQGTAVILESGEGGAATEIHLIERNDLPRERQGKGSVHHVAFRVRDEEEL    237
           *.* ..:         ::* .:**  :.: : ...  *  : * *:***.****.:   *
```

```
             *   :   *: .** *:*.:*::** *    ****::******  **  : **. ::*
CpsA         EEWIERISSVGLPSSGYVDRHFFESLYARVAPQILFEFATDGPGFMGDEPYETLGEKLSL    295
101031       DEWTARFESFGFHTSGYVNRHFFESLYARVAPQILFELATDGPGFMGDEPYETLGEKLSL    295
315730       EEWIERLSRVGLPSSGYVNRHFFESLYARVAPQILFEFATDGPGFMGDEPYETLGEKLSL    295
226900       EEWIERISSVGLPSSGYVNRHFFESLYARVAPQILFEFATDGPGFMGDESYETLGEKLSL    295
269801       EEWIERISSVGLPSSGYVDRHFFESLYARVAPQILFEFATDGPGFMGDEPYETLGEKLSL    295
405532       EEWIERISSVGLPSSGYVDRHFFESLYARVAPQILFEFATDGPGFMGDEPYETLGEKLSL    295
412694       EEWIERISSVGLPSSGYVDRHFFESLYARVAPQILFEFATDGPGFMGDEPYETLGEKLSL    295
280355       EEWIERISSVGLPSSGYVDRHFFESLYARVAPQILFEFATDGPGFMGDEPYETLGEKLSL    295
280477       EEWIERISSVGLPSSGYVDRHFFESLYARVAPQILFEFATDGPGFMGDEPYETLGEKLSL    295
280354       EEWIERISSVGLPSSGYVDRHFFESLYARVAPQILFEFATDGPGFMGDEPYETLGEKLSL    295
205919       EEWIERISSVGLPSSGYVDRHFFESLYARVAPQILFEFATDGPGFMGDEPYETLGEKLSL    295
260799       EEWIERISSVGLPSSGYVDRHFFESLYARVAPQILFEFATDGPGFMGDEPYETLGEKLSL    295
261591       EEWIERISSVGLPSSGYVDRHFFESLYARVAPQILFEFATDGPGFMGDEPYETLGEKLSL    295
198094       EEWIERISSVGLPSSGYVDRHFFESLYARVAPQILFEFATDGPGFMGDEPYETLGEKLSL    295
261594       EEWIERISSVGLPSSGYVDRHFFESLYARVAPQILFEFATDGPGFMGDEPYETLGEKLSL    295
288681       EEWIERISSVGLPSSGYVDRHFFESLYARVAPQILFEFATDGPGFMGDEPYETLGEKLSL    295
222523       EEWIERISSVGLPSSGYVDRHFFESLYARVAPQILFEFATDGPGFMGDEPYETLGEKLSL    295
281309       EEWIERISSVGLPSSGYVDRHFFESLYARVAPQILFEFATDGPGFMGDEPYETLGEKLSL    295
315750       EKWYDKISKAGFTNSGVVERFYFKALYFREPNGILFEISTDGPGFTVDEGVDELGDQLAL    297
326423       ALWHRIITKEGFSNSGIVERYYFKALYFREPNGILFELSTDGPGFMVDEKPEELGKTVAL    297
224308       AGWHRIISREGFSNSGIVERYYFKALYFREPNGILFELSTDGPGFMVDENLDELGQTIAL    297
             *   :   *: .** *:*.:*::** *    ****::******  **  : **. ::*


             **:** :* :**  : **:
CpsA         PPFLEPKREEIEKLVRPIDTVRSTKEFIKE--       325
101031       PPFLEPKRTQIEGLVRPIDTVRSSITFDKEYE       327
315730       PPFLEPKREEIEKLVRPIDTVRSTKEFIKE--       325
226900       PPFLEPKREEIEKLVRPIDTVRSTKEFIKE--       325
269801       PPFLEPKREEIEKLVRPIDTVRSTKEFIKE--       325
405532       PPFLEPKREEIEKLVRPIDTVRSTKEFIKE--       325
412694       PPFLEPKREEIEKLVRPIDTVRSTKEFIKE--       325
280355       PPFLEPKREEIEKLVRPIDTVRSTKEFIKE--       325
280477       PPFLEPKREEIEKLVRPIDTVRSTKEFIKE--       325
280354       PPFLEPKREEIEKLVRPIDTVRSTKEFIKE--       325
205919       PPFLEPKREEIEKLVRPIDTVRSTKEFIKE--       325
260799       PPFLEPKREEIEKLVRPIDTVRSTKEFIKE--       325
261591       PPFLEPKREEIEKLVRPIDTVRSTKEFIKE--       325
198094       PPFLEPKREEIEKLVRPIDTVRSTKEFIKE--       325
261594       PPFLEPKREEIEKLVRPIDTVRSTKEFIKE--       325
288681       PPFLEPKREEIEKLVRPIDTVRSTKEFIKE--       325
222523       PPFLEPKREEIEKLVRPIDTVRSTKEFIKE--       325
281309       PPFLEPKREEIEKLVRPIDTVRSTKEFIKE--       325
315750       PPFLEHKRAEIEQRLTPIQSS-----------       318
326423       PPYLEHRRSEIEARLKPIQ-------------       316
224308       PPYLEHRRAEIEAKLKPIQ-------------       316
             **:** :* :**  : **:
```

Fig. S6: Multiple sequences alignments between of CpsA from *Bacillus tropicus* AOA-CPS1 with closely related neighbourhoods' whole-genome. Where the sequences are similar were shaded in pink and indicated with asterisk (*). Glyoxalase family protein from *Bacillus tropicus* (CpsA), *B. amyloliquefaciens* FZB42 (326423), *B. anthracis* str. A1055 (280355), *B. anthracis* str. Australia 94 (280477), *B. anthracis* str. CNEVA-9066 (280354), *B. anthracis* str. Kruger B

(205919), *B. anthracis* str. Sterne (260799), *B. anthracis* str. Vollum (261591), *Bacillus* B-14905 (101031), *B. cereus* ATCC 10987 (222523), *B. cereus* B4264 (405532), *B. cereus* E33L (288681), *B. cereus* G9241 (269801), *B. cereus* subsp. cytotoxis NVH 391-98 (315749), *B. pumilus* SAFR-032 (315750), *B. thuringiensis* str. Al Hakam (412694), *B. weihenstephanensis* KBAB4 (315730), *B. subtilis* subsp. *subtilis* str. 168 (224308), *B. cereus* ATCC 14579 (226900), *B. anthracis* str. Ames (198094), *B. anthracis* str. 'Ames Ancestor (261594), *B. thuringiensis* serovar konkukian str. 97-27 (281309).

```
                 *: ::* * .*.  *       :   . :::       . ...:*: ..
CpsE        --MTIKRKKVSVIGAGFTGATTAFLLAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
176279      ---MVNRRKISIIGAGHTGGTLAFILAQKELGD-IVLIERQQSEGMAKGKALDILESGPI    56
101031      --MSLKRKKLSVIGGGFTGATAAFLAAQKELGD-VVLVDIPQAENPTKGKALDMWEAAPI    57
66692       --MAIKRRKISVIGSGFTGATTALMVAQKELGD-VVLLDIPNMEGPTKGKALDMLESTPV    57
272558      --MAIKRRKVSVIGAGFTGATTALMVAQKELGD-VVLVDIPQMEGPTKGKALDMLESTPV    57
221109      --MGLKRKKISVIGSGFTGATTALMVAQKELGD-VVLVDIPDMEDPTKGKALDMAEAAPV    57
315750      --MANKRKKVSVIGAGFTGATTAFLTAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
279010      --MGKKRNKVSVIGAGFTGATTAFLTAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
224308      --MGNTRKKVSVIGAGFTGATTAFLIAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
326423      --MGKTRSKVSVIGAGFTGATTAFLTAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
491915      --MGMKRKKISIIGAGFTGATTAFILAQKELGD-IVLVDIPQLENPTKGKALDMLESSPV    57
235909      --MAMKRKKISVIGAGFTGATTAFLLAQKELGD-VVLVDIPQLENPTKGKALDMLEASPV    57
420246      --MAMKRKKISVIGAGFTGATTAFLLAQKELGD-IVLVDIPQLENPTKGKALDMLESSPV    57
315749      --MTIKRKKVSVIGAGFTGATTAFLLAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
315730      --MTIKRKKVSVIGAGFTGATTAFLLAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
288681      --MTIKRKKVSVIGAGFTGATTAFLLAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
281309      --MTIKRKKVSVIGAGFTGATTAFLLAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
261594      --MTIKRKKVSVIGAGFTGATTAFLLAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
198094      --MTIKRKKVSVIGAGFTGATTAFLLAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
280355      --MTIKRKKVSVIGAGFTGATTAFLLAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
280477      --MTIKRKKVSVIGAGFTGATTAFLLAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
280354      --MTIKRKKVSVIGAGFTGATTAFLLAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
205919      --MTIKRKKVSVIGAGFTGATTAFLLAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
260799      --MTIKRKKVSVIGAGFTGATTAFLLAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
261591      --MTIKRKKVSVIGAGFTGATTAFLLAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
222523      --MTIKRKKVSVIGAGFTGATTAFLLAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
405532      --MTIKRKKVSVIGAGFTGATTAFLLAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
269801      --MTIKRKKVSVIGAGFTGATTAFLLAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
412694      --MTIKRKKVSVIGAGFTGATTAFLLAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
226900      --MTIKRKKVSVIGAGFTGATTAFLLAQKELAD-VVLVDIPQLENPTKGKALDMLEASPV    57
413999      MIKKRNTTKISVIGAGSVGATTAYALMLSGVATEIVLVDVN--KSKTEGEAMDLSHGADF    58
195103      -MIREKTNKISIIGAGFVGSTTAFALMQDGLASEIVIVDIN--KDKAHAEAMDLAQGAAF    57
282458      -MNKFKGNKVVLIGNGAVGSIYAFSLVNQSIVDELVIIDLD--AEKVRGDVMDLKHATPY    57
272626      ---MKDHQKIILVGDGAVGSSYAFACVNLSIGQEFGIIDID--KDRTIGDAMDLSHAVPF    55
169963      ---MKDHQKIILVGDGAVGSSYAFACVNLSIGQEFGIIDID--KDRTIGDAMDLSHAVPF    55
                 *: ::* * .*.  *       :   . :::       . ...:*: ..
```

```
                 : .     .:*:*:: **   * * :* :::  *  :.:    ::     .
CpsE        QGFDANIIGTSDYEDTADSDVVVITAGIARKPGMSRDDLVATNSKIMKSITRDIAKHSPN    117
176279      WGFDTSVHGSVNIEDIKDSDIVVMTAGIPRKSGMTREELVQTNEQIVRETALQIATYAPH    116
101031      QGFDAYVKGTSNYADTANSDVVIITAGVARKPGMSRDDLVQINQGVMKTVSKEIAAHSPN    117
66692       QGVDSTITGTSSYEDTKDSDVVVITAGIARKPGMSRDDLVATNAKIMKSVTKEVVKYSPN    117
272558      QGVDVNITGTSSYEYTKDSDVVVITAGIARKPGMSRDDLVSTNAGIMKAVTKEVVKHSPN    117
221109      QGFDAKITGTSNYADTEGSDLVIITAGIARKPGMSRDDLVNTNANIMKSVTKEIVHYSPN    117
315750      QGFDANITGTSNYEDTAGSDVVVITAGIARKPGMSRDDLVSTNEKIMRSVTREIVKYSPE    117
279010      QGFDANITGTANYEDTAGSDIVVITAGIARKPGMSRDDLVATNEKIMRSVTKEVVKYSPD    117
224308      QGFDAKITGTSNYEDTAGSDIVVITAGIARKPGMSRDDLVSTNEKIMRSVTQEIVKYSPD    117
326423      QGFDAKITGTSNYEDTAGSDIVVITAGIARKPGMSRDDLVSTNEKIMRSVTREIVKYSPD    117
491915      LGFDANIIGTSDYADTADSDVVVITAGIARKPGMSRDDLVTTNQGIMKAVTKEVVKYSPN    117
235909      LGFDANIIGTSDYADTADSDIVVITAGIARKPGMSRDDLVTTNQKIMKQVTKEVVKYSPN    117
420246      LGFDANIVGTSDYADTADSDIVVITAGIARKPGMSRDDLVTTNQKIMKQVTKEVVKYSPN    117
315749      QGFDANIIGTSDYADTADSDVVIITAGIARKPGMSRDDLVATNSKIMKSVTKEIAKHSPD    117
315730      QGFDANIIGTSDYADTADSDVVVITAGIARKPGMSRDDLVATNSKIMKSITRDIAKHSPN    117
288681      QGFDANIIGTSDYADTADSDVVVITAGIARKPGMSRDDLVATNSKIMKSITRDIAKHSPN    117
281309      QGFDANIIGTSDYADTADSDVVVITAGIARKPGMSRDDLVATNSKIMKSITRDIAKHSPN    117
261594      QGFDANIIGTSDYADTADSDVVVITAGIARKPGMSRDDLVATNSKIMKSITRDIAKHSPN    117
198094      QGFDANIIGTSDYADTADSDVVVITAGIARKPGMSRDDLVATNSKIMKSITRDIAKHSPN    117
280355      QGFDANIIGTSDYADTADSDVVVITAGIARKPGMSRDDLVATNSKIMKSITRDIAKHSPN    117
280477      QGFDANIIGTSDYADTADSDVVVITAGIARKPGMSRDDLVATNSKIMKSITRDIAKHSPN    117
280354      QGFDANIIGTSDYADTADSDVVVITAGIARKPGMSRDDLVATNSKIMKSITRDIAKHSPN    117
205919      QGFDANIIGTSDYADTADSDVVVITAGIARKPGMSRDDLVATNSKIMKSITRDIAKHSPN    117
260799      QGFDANIIGTSDYADTADSDVVVITAGIARKPGMSRDDLVATNSKIMKSITRDIAKHSPN    117
261591      QGFDANIIGTSDYADTADSDVVVITAGIARKPGMSRDDLVATNSKIMKSITRDIAKHSPN    117
222523      QGFDANIIGTSDYADTADSDVVVITAGIARKPGMSRDDLVATNSKIMKSITRDIAKHSPN    117
405532      QGFDANIIGTSDYADTADSDVVVITAGIARKPGMSRDDLVATNSKIMKSITRDIAKHSPN    117
269801      QGFDANIIGTSDYADTADSDVVVITAGIARKPGMSRDDLVATNSKIMKSITRDIAKHSPN    117
412694      QGFDANIIGTSDYADTADSDVVVITAGIARKPGMSRDDLVATNSKIMKSITRDIAKHSPN    117
226900      QGFDANIIGTSDYADTADSDVVVITAGIARKPGMSRDDLVATNSKIMKSITRDIAKHSPN    117
413999      VKPV-N-ILSGDYKDTEGSDIVVITAGAAQKVGETRLQLINKNININIFKSIIPQVVKYNKD    116
195103      VKSV-D-IKSGDYADTKDSDIVIITAGVGPKPGETRLDIINKNLKIFQSIVPEVVKYSPN    115
282458      SPTTVR-VKAGEYSDCHDADLVVICAGAAQKPGETRLDLVSKNLKIFKSIVGEVMASKFD    116
272626      STPK-K-IYSANYSDCHDADLVVVTAGTAQKPGETRLDLVNRNIKIMKGIVDEVMASGFD    113
169963      STPK-K-IYSANYSDCHDADLVVVTAGTAQKPGETRLDLVNRNIKIMKGIVDEVMASGFD    113
                 : .     .:*:*:: **   * * :* :::  *  :.:    ::     .
```

```
                 ::: :**.* ::* .:: :*:**:::**..  **::*    :..  ..  ..:   ::
CpsE             AIIVVLTNPVDAMTYSVFKEAGFPKERVIGQSGVLDTARFRTFIAQELNLSVKDITGFVL    177
176279           SIIIVLTNPVDVMTYTAFKASGFPKERIIGQSGILDAARYRTFIAQELNVSVKDVNGFVL    176
101031           ATIIVLTNPVDAMTYTVYKETGFPKNRVIGQSGVLDTARFCAFVAEELNVSVKDITGFVL    177
66692            SYIIVLTNPADAMTYTVYKESGFPKNRVIGQSGVLDTARFRTFVAQELNVSVEDVTGFVL    177
272558           AYIIVLTNPADAMTYTVYKESGFPKNRVIGQSGVLDTARFRTFVAQELNLSVEDITGFVL    177
221109           TTIVVLTNPVDAMTYTVFKESGLPKERVIGQSGILDTARFRTFVAEELNLSVKDVTGFVL    177
315750           AIIVVLTNPVDAMTYAVYKESGLPKEKVIGQSGILDTARFRTFVAQELNLSVKDVTGFVL    177
279010           CIIIVLTNPVDAMTYAVYKESGFPKERVIGQSGILDTARFRTFVAQELNLSVKDITGFVL    177
224308           SIIVVLTNPVDAMTYAVYKESGFPKERVIGQSGVLDTARFRTFVAEELNLSVKDVTGFVL    177
326423           CIIVVLTNPVDAMTYAVYKESGFPKERVIGQSGVLDTARFRTFVAEELNLSVKDVTGFVL    177
491915           CFIIVLTNPVDAMTYTVFKESGFPKNRVIGQSGVLDTARFRTFVAQELNLSVKDITGFVL    177
235909           CYIIVLTNPVDAMTYTVFQESGFPKNRVIGQSGVLDTARFRTFVAEELNISVKDVTGFVL    177
420246           CYIIVLTNPVDAMTYTVFKESGFPKNRVIGQSGVLDTARFRTFVAQELNISVKDVTGFVL    177
315749           TIIIVLTNPVDAMTYSVFKEAGFPKERVIGQSGVLDTARFRTFIAQELNLSVKDITGFVL    177
315730           AIILVLTNPVDAMTYSVFKEAGFPKERVIGQSGVLDTARFRTFISQELNLSVKDITGFVL    177
288681           AIIVVLTNPVDAMTYSVFKEAGFPKERVIGQSGVLDTARFRTFIAQELNFSVKDITGFVL    177
281309           AIIVVLTNPVDAMTYSVFKEAGFPKERVIGQSGVLDTARFRTFIAQELNLSVKDITGFVL    177
261594           AIIVVLTNPVDAMTYSVFKEAGFPKERVIGQSGVLDTARFRTFIAQELNLSVKDITGFVL    177
198094           AIIVVLTNPVDAMTYSVFKEAGFPKERVIGQSGVLDTARFRTFIAQELNLSVKDITGFVL    177
280355           AIIVVLTNPVDAMTYSVFKEAGFPKERVIGQSGVLDTARFRTFIAQELNLSVKDITGFVL    177
280477           AIIVVLTNPVDAMTYSVFKEAGFPKERVIGQSGVLDTARFRTFIAQELNLSVKDITGFVL    177
280354           AIIVVLTNPVDAMTYSVFKEAGFPKERVIGQSGVLDTARFRTFIAQELNLSVKDITGFVL    177
205919           AIIVVLTNPVDAMTYSVFKEAGFPKERVIGQSGVLDTARFRTFIAQELNLSVKDITGFVL    177
260799           AIIVVLTNPVDAMTYSVFKEAGFPKERVIGQSGVLDTARFRTFIAQELNLSVKDITGFVL    177
261591           AIIVVLTNPVDAMTYSVFKEAGFPKERVIGQSGVLDTARFRTFIAQELNLSVKDITGFVL    177
222523           AIIVVLTNPVDAMTYSVFKEAGFPKERVIGQSGVLDTARFRTFIAQELNLSVKDITGFVL    177
405532           AIIVVLTNPVDAMTYSVFKEAGFPKERVIGQSGVLDTARFRTFIAQELNLSVKDITGFVL    177
269801           AIIVVLTNPVDAMTYSVFKEAGFPKERVIGQSGVLDTARFRTFIAQELNLSVKDITGFVL    177
412694           AIIVVLTNPVDAMTYSVFKEAGFPKERVIGQSGVLDTARFRTFIAQELNLSVKDITGFVL    177
226900           AIIVVLTNPVDAMTYSVFKEAGFPKERVIGQSGVLDTARFRTFIAQELNLSVKDITGFVL    177
413999           AILLVVSNPVDVLSYVTYKLSGFPKERVIGSGTVLDTSRLKHEIGKRYKIDPRNVNTYIM    176
195103           SILLVVSNPVDILTYITYKLSGFPKERVIGSGTVLDTSRLKYMLSEHFDIDARNVHTYII    175
282458           GIFLVATNPVDILAYATWKFSGLPKERVIGSGTILDSARFRLLLSEAFDVAPRSVDAQII    176
272626           GIFLIASNPVDILTYATWKFSGLPKERVIGSGTSLDTARFRMSIADYLKVDARNVHGYIL    173
169963           GIFLIASNPVDILTYATWKFSGLPKERVIGSGTSLDTARFRMSIADYLKVDARNVHGYIL    173
                 ::: :**.* ::* .:: :*:**:::**..  **::*    :..  ..  ..:   ::
```

```
           * ***  .     :   .  :                . :   .: .. :::    :* :
CpsE       GGHGDDMVPLVRYSYAGGIPLETLIPKER-------LEAIVERTRKGGGEIVGLLGNGSA    230
176279     GGHGDTMLPLINNTHINGIPVKHLISEEK-------IDQIVERTRKGGAEIVALLGQGSA    229
101031     GGHGDTMVPLTRYSFAGGIPLETLIPAER-------LEAIVDRTRNGGAEIVNLLGNGSA    230
66692      GGHGDDMVPLIRYSAAGGVPLTKLIAPER-------IEEIVERTRKGGGEIVGLLGNGSA    230
272558     GGHGDDMVPLIRYSYAGGIPLEKLLPQER-------IDAIVERTRKGGGEIVGLLGNGSA    230
221109     GGHGDDMVPLIRYSYAGGIPLEKLIPQER-------LDAIVQRTRTGGGEIVNLLGNGSA    230
315750     GGHGDDMVPLVRYSYAGGIPLETLIPKER-------IDAIVERTRKGGGEIVNLLGNGSA    230
279010     GGHGDDMVPLVRYSYAGGIPLETLLPKDR-------IDAIVERTRKGGGEIVNLLGNGSA    230
224308     GGHGDDMVPLVRYSYAGGIPLETLIPKER-------IDAIVERTRKGGGEIVNLLGNGSA    230
326423     GGHGDDMVPLVRYSYAGGIPLETLIPKDR-------LDAIVERTRKGGGEIVNLLGNGSA    230
491915     GGHGDDMVPLVRYSYAGGIPLETLIPKER-------LDAIVERTRKGGGEIVNLLGNGSA    230
235909     GGHGDDMVPLVRYSYAGGIPLEKLIPKDR-------LDAIVERTRKGGGEIVNLLGNGSA    230
420246     GGHGDDMVPLVRYSYAGGIPLEKLIPKDR-------LDAIVERTRKGGGEIVNLLGNGSA    230
315749     GGHGDDMVPLVRYSYAGGIPLETLISKER-------LDAIVERTRKGGGEIVNLLGNGSA    230
315730     GGHGDDMVPLVRYSYAGGIPLETLIPKER-------LEAIVERTRKGGGEIVGLLGNGSA    230
288681     GGHGDDMVPLVRYSYAGGIPLETLIPKER-------LEAIVERTRKGGGEIVGLLGNGSA    230
281309     GGHGDDMVPLVRYSYAGGIPLETLIPKER-------LEAIVERTRKGGGEIVGLLGNGSA    230
261594     GGHGDDMVPLVRYSYAGGIPLETLIPKER-------LEAIVERTRKGGGEIVGLLGNGSA    230
198094     GGHGDDMVPLVRYSYAGGIPLETLIPKER-------LEAIVERTRKGGGEIVGLLGNGSA    230
280355     GGHGDDMVPLVRYSYAGGIPLETLIPKER-------LEAIVERTRKGGGEIVGLLGNGSA    230
280477     GGHGDDMVPLVRYSYAGGIPLETLIPKER-------LEAIVERTRKGGGEIVGLLGNGSA    230
280354     GGHGDDMVPLVRYSYAGGIPLETLIPKER-------LEAIVERTRKGGGEIVGLLGNGSA    230
205919     GGHGDDMVPLVRYSYAGGIPLETLIPKER-------LEAIVERTRKGGGEIVGLLGNGSA    230
260799     GGHGDDMVPLVRYSYAGGIPLETLIPKER-------LEAIVERTRKGGGEIVGLLGNGSA    230
261591     GGHGDDMVPLVRYSYAGGIPLETLIPKER-------LEAIVERTRKGGGEIVGLLGNGSA    230
222523     GGHGDDMVPLVRYSYAGGIPLETLIPKER-------LEAIVERTRKGGGEIVGLLGNGSA    230
405532     GGHGDDMVPLVRYSYAGGIPLETLIPKER-------LEAIVERTRKGGGEIVGLLGNGSA    230
269801     GGHGDDMVPLVRYSYAGGIPLETLIPKER-------LEAIVERTRKGGGEIVGLLGNGSA    230
412694     GGHGDDMVPLVRYSYAGGIPLETLIPKER-------LEAIVERTRKGGGEIVGLLGNGSA    230
226900     GGHGDDMVPLVRYSYAGGIPLETLIPKER-------LEAIVERTRKGGGEIVGLLGNGSA    230
413999     GEHGDSEIATWSVTNIQNIKIDEYANKENLEYNDNFRKEVYENVKNAAYEVINR--KGAT    234
195103     GEHGDSEITAWSLTNIAGANVEEYCKTVCANCDGSFKKELPEKVKNAAYEIINS--KGYT    233
282458     GEHGDTELPVWSHANIAGQPLKTLLEQRPEG--KAQIEQIFVQTRDAAYDIIQA--KGAT    232
272626     GEHGDTQFPAWSHTTVGGLPITEWINEDEQG----AMDTIFVSVRDAAYEIINK--KGAT    227
169963     GEHGDTEFPAWSHTTVGGLPITEWISEDEQG----AMDTIFVSVRDAAYEIINK--KGAT    227
           * ***  .     :   .  :                . :   .: .. :::    :* :
```

```
            *. * .:    .**. :.  ::     .:. *.  .: .. * ::. .* .. ..: :
CpsE        YYAPAASLVEMTEAILKDQRRVLPAIAYLEGEYGYSDLYLGVPVILGGNGIEKIIELELL    290
176279      YYAPATAIYETIDAIFNDRKRLLPSIAYLEGEYGCSDICFGVPTIIGYQGIEKIIEVDMN    289
101031      YYAPAAALIEMAEAIIKDQKRILPSIAYLEGEYGYSDLYLGVPTLLGENGIEKIFELELT    290
66692       YYAPAASLTQMVEAILKDKKRIIPTIAYLEGEYGQHDLYLGVPTILGGDGIEKVIELDLT    290
272558      YYAPAASLAEMVEAILKDKKRVLPTIAYLEGEYGYEDIYVGVPTILGGDGIEKVIELDLT    290
221109      YYAPAASLTVMAEAILKDQRRVLPTIAYLEGEYGYQDIYLGVPTILGGEGIEEIIELDLT    290
315750      YYAPAASLVEMVEAILKDQRRVMPTIAYLEGEYGYEGIYLGVPTIVGGNGLEQIIELELT    290
279010      YYAPAASLTEMVEAILKDQRRVLPTIAYLEGEYGYEGIYLGVPTIIGGNGLEQIIELELT    290
224308      YYAPAASLTEMVEAILKDQRRVLPTIAYLEGEYGYEGIYLGVPTIVGGNGLEQIIELELT    290
326423      YYAPAASLTEMVEAILKDQRRVLPTIAYLEGEYGHEGIYLGVPTIIGGNGLEQIIELELT    290
491915      YYAPAASLAEMVEAIVKDQRRVLPAIAYLEGEYGYEGIYLGVPTILGGNGIEKVIELELT    290
235909      YYAPAASLVEMVEAILKDQRRILPAIAYLEGEYGYEGIYLGVPTILGGNGIEKVIELELT    290
420246      YYAPAASLAEMVEAIVKDQRRILPAIAYLEGEYGYEGIYLGVPTILGGNGIEKVIELELT    290
315749      YYAPAASLVEMTEAILKDQRRVLPAIAYLEGEYGYRDLYLGVPVILGGNGIEKVIELELR    290
315730      YYAPAASLVEMTEAILKDQRRVLPAIAYLEGEYGYSDLYLGVPVILGGNGIEKIIELELL    290
288681      YYAPAASLVEMTEAILKDQRRVLPAIAYLEGEYGYSDLYLGVPVILGGNGIEKIIELELL    290
281309      YYAPAASLVEMTEAILKDQRRVLPAIAYLEGEYGYSDLYLGVPVILGGNGIEKIIELELL    290
261594      YYAPAASLVEMTEAILKDQRRVLPAIAYLEGEYGYSDLYLGVPVILGGNGIEKIIELELL    290
198094      YYAPAASLVEMTEAILKDQRRVLPAIAYLEGEYGYSDLYLGVPVILGGNGIEKIIELELL    290
280355      YYAPAASLVEMTEAILKDQRRVLPAIAYLEGEYGYSDLYLGVPVILGGNGIEKIIELELL    290
280477      YYAPAASLVEMTEAILKDQRRVLPAIAYLEGEYGYSDLYLGVPVILGGNGIEKIIELELL    290
280354      YYAPAASLVEMTEAILKDQRRVLPAIAYLEGEYGYSDLYLGVPVILGGNGIEKIIELELL    290
205919      YYAPAASLVEMTEAILKDQRRVLPAIAYLEGEYGYSDLYLGVPVILGGNGIEKIIELELL    290
260799      YYAPAASLVEMTEAILKDQRRVLPAIAYLEGEYGYSDLYLGVPVILGGNGIEKIIELELL    290
261591      YYAPAASLVEMTEAILKDQRRVLPAIAYLEGEYGYSDLYLGVPVILGGNGIEKIIELELL    290
222523      YYAPAASLVEMTEAILKDQRRVLPAIAYLEGEYGYSDLYLGVPVILGGNGIEKIIELELL    290
405532      YYAPAASLVEMTEAILKDQRRVLPAIAYLEGEYGYSDLYLGVPVILGGNGIEKIIELELL    290
269801      YYAPAASLVEMTEAILKDQRRVLPAIAYLEGEYGYSDLYLGVPVILGGNGIEKIIELELL    290
412694      YYAPAASLVEMTEAILKDQRRVLPAIAYLEGEYGYSDLYLGVPVILGGNGIEKIIELELL    290
226900      YYAPAASLVEMTEAILKDQRRILPAIAYLEGEYGYSDLYLGVPVILGGNGIEKIIELELL    290
413999      FYAIALAVTRIVKAILGDEKTILPVSTLVENYYGIKDVYLGMPCIVGGSGIEKALSIDLN    294
195103      NYAVALAVTRIVEAILRDENAILTVSSLFEGQYGIDNVYLAMPTIVDRSGARQILDVPIS    293
282458      YYGVAMGLARITEAIFRNEDAVLTVSALLEGEYDEEDVYIGVPAVINRNGIRNVVEIPLN    292
272626      FYGVAAALARITKAILNNENAILPLSVYLDGHYGMNDIYIGAPAVVNRQGVRHIVEMNLN    287
169963      FYGVAAALARITKAILNNENAILPLSVYLDGHYGMNDIYIGAPAVVNRQGVRHIVEMNLN    287
            *. * .:    .**. :.  ::     .:. *.  .: .. * ::. .* .. ..: :
```

```
                 *   : .* . :
CpsE          ADEKEALDRSVESVRNVMKVLV----     312
176279        NDEYQQLQHSAQAVSEVKNSLKFK--     313
101031        DKEKAALDQSADAVRNVMKILV----     312
66692         EEEKAQLDKSVQSVRNVMAALPAE--     314
272558        DEEKATFAKSIESVRNVMSALPKE--     314
221109        KEEKAQLDKSADSVKNVLNVLQ----     312
315750        EEERSQLDRSVESVKNVMKVLS----     312
279010        ETEKSQLDKSVESVKNVMKVLS----     312
224308        DYERAQLNKSVESVKNVMKVLS----     312
326423        DYEKAQLSKSVESVKNVMKVLS----     312
491915        EEEKAALAKSVESVKNVMKVLQ----     312
235909        EEEKAALAKSVESVKNVMRMLE----     312
420246        EDEKAALAKSVESVKNVMRVLE----     312
315749        EEEKMALDRSVESVRNVMEILS----     312
315730        ADEKEALDRSVESVRNVMKVLV----     312
288681        ADEKEALDRSVESVRNVMKVLV----     312
281309        ADEKEALDRSVESVRNVMKVLV----     312
261594        ADEKEALDRSVESVRNVMKVLV----     312
198094        ADEKEALDRSVESVRNVMKVLV----     312
280355        ADEKEALDRSVESVRNVMKVLV----     312
280477        ADEKEALDRSVESVRNVMKVLV----     312
280354        ADEKEALDRSVESVRNVMKVLV----     312
205919        ADEKEALDRSVESVRNVMKVLV----     312
260799        ADEKEALDRSVESVRNVMKVLV----     312
261591        ADEKEALDRSVESVRNVMKVLV----     312
222523        ADEKEALDRSVESVRNVMKVLV----     312
405532        ADEKEALDRSVESVRNVMKVLV----     312
269801        ADEKEALDRSVESVRNVMKVLV----     312
412694        ADEKEALDRSVESVRNVMKVLV----     312
226900        ADEKEALDRSVESVRNVMKVLV----     312
413999        KTEASKLVKSAETLKNTLNNASCL--     318
195103        NEEKENLIKSAEILKGHIANSELD--     317
282458        DEEQSKFAHSAKTLKDIMAEAEELK-     317
272626        DKEKEQMKNSADTLKKVLDDAMKQID     313
169963        DKEKEQMKNSADTLKKVLDDAMKQID     313
                 *   : .* . :
```

Fig. S7: Multiple sequences alignments between of CpsE from *Bacillus tropicus* AOA-CPS1 with closely related neighbourhoods' whole-genome. Where the sequences are similar were indicated with asterisk (*) shaded in green. L-lactate dehydrogenase (EC 1.1.1.27) from *Clostridium botulinum* A str. ATCC 3502 (413999), *C. perfringens* ATCC 13124 (195103), *Listeria innocua* Clip11262 (272626), *Listeria monocytogenes* EGD-e (169963), *Staphylococcus aureus* subsp. *aureus* MRSA252 (282458)and malate dehydrogenase (EC 1.1.1.37) from *Staphylococcus epidermidis* RP62A (176279), *Bacillus thuringiensis* serovar konkukian str. 97-27 (281309), *B. subtilis* subsp subtilis str. 168 (224308), *B. cereus* ATCC 14579 (226900), *B. tropicus* AOA-CPS1 (CpsE), *B.* anthracis str. Ames (198094), *B.clausii* KSM-K16 (66692), *B. halodurans* C-125 (272558), *B. licheniformis* ATCC 14580 (279010), *Anoxybacillus flavithermus* WK1 (491915), *Geobacillus kaustophilus* HTA426 (235909), *G. thermodenitrificans* NG80-2 (420246), *Oceanobacillus iheyensis* HTE831 (221109), *B. amyloliquefaciens* FZB42 (326423), *B. anthracis* str. A1055 (280355), *B. anthracis* str. Australia 94 (280477), *B. anthracis* str. CNEVA-9066 (280354), *B. anthracis* str.

Kruger B (205919), *B. anthracis* str. Sterne (260799), *B. anthracis* Vollum (261591), *Bacillus* B-14905 (101031), *B. cereus* ATCC 10987 (222523), *B. cereu*s B4264 (405532), *B. cereus* e33L (288681), *B. cereus* G9241 (269801), *B. weihenstephanensis* KBAB4 (315730), *B. thuringiensis* str. Al Hakam (412694), *B. pumilus* SAFR-032 (315750), and *B. cereus* subsp. cytotoxis NVH 391-98 (315749).

CHAPTER TEN

10.0 CONCLUDING REMARKS

10.1 The research in perspective

Pentachlorophenol (PCP) is a toxic compound produced in the 1930s, marketed with different names and used extensively as herbicides, fungicides, disinfectants, algaecide and as an ingredient in antifouling paints (Kim et al., 2019). PCP was specifically used as a wood preservative to combat wood pests and fungi (IARC, 2019). PCP has not only been listed as a priority pollutant (ATSDR, 2017), due of their toxicological traits (Igbinosa et al., 2013), it has also been designated as a human carcinogen (IARC, 2019b, 2019a; Stockholm convention, 2019). Unfortunately, PCP is still been used by certified industries (Kim et al., 2019), to protect products that would as a matter of time come in contact with humans, foods, and animals.

For instance, PCP, dibenzo-p-dioxins and polychlorinated dibenzofurans were recently found in surface soil surrounding PCP-treated utility poles on the Kenai National Wildlife Refuge, Alaska USA (Verbrugge et al., 2018). The surface soil levels of PCP and polychlorinated dibenzofurans in the sampling area exceeded both human health and ecological risk-based screening levels, like another project in Montreal, Quebec in Canada, indicating that millions of similarly treated utility poles in North America may be point sources of PCP (Verbrugge et al., 2018). The re-occurrence of chlorophenolic compounds including PCP is not limited to the European soils (Thompson, et al. 2017; Silva et al. 2019), many cases were reported in South Africa (WRC 2015; Yahaya et al. 2019; Nthunya et al. 2019; Olatunji 2019; Gabuka et al. 2018) and other parts of Africa recently (Chiaia-Hernandez et al. 2017).

PCP is a synthetic compound, very toxic to many microorganisms and resistant to biodegradation, *Bacillus cereus* strain AOA-CPS1 was able to efficiently degrade PCP and some congeners in co-metabolism as described in chapter three. The ability of the isolate to degrade PCP, the biokinetic parameters, metabolic genes detected, presence of two different pathways for PCP degradation in the genome of the organism and its relatedness to other organism that have been used in various remediation processes showed that it has potential for use in bulk bioremediation of PCP.

Optimization significantly enhanced cell growth and PCP degradation compared to the unoptimized condition in chapter three, the low-affinity coefficient and high inhibition constant obtained in chapter four showed that the bacterium has a high affinity and tolerance to PCP, which could be explored for the biotechnological applications of the organism for bulk PCP transformation.

The detection, cloning and overexpression of catabolic genes involved in PCP degradation in *Bacillus cereus* strain AOA-CPS1 also showed that the isolate has evolved pathway for the complete mineralization of PCP in accordance with another PCP degrading bacterium (Chanama & Chanama, 2011). The biological roles of some of the catabolic enzymes shown in chapters' six to nine showed that the enzymes recruited into PCP degradation pathway of *Bacillus cereus* are metabolically active and well adapted to the environment of the organism.

In chapters six and seven, another role of bacterial phenylalanine hydroxylating system in xenobiotic degradation was discovered and the first experimental evidence that the operon (bacterial phenylalanine hydroxylating system) encoding CpsB and CpsD enzymes recruited into *Bacillus cereus* AOA-CPS1 PCP degradation pathway play crucial roles in PCP degradation. CpsB is involved in hydroxylation of PCP to Tet-CBQ similar to PCP-4-monooxygenase of *S. chlorophenolicum* (Hlouchova et al., 2012) while CpsD is a Tet-CBQ reductase that catalysed the reduction of Tet-CBQ to Tet-CHQ also functionally similar to *S. chlorophenolicum* Tet-CBQ reductase PcpD (Chen & Yang, 2008; Dai et al., 2003).

In chapter seven, metabolically efficient Cytochrome P450 overexpressed heterologous was also found to degrade PCP in a possible three steps reaction, confirming the presence of the two pathways for PCP degradation as proposed in chapter three. This further strengthen the bioremediation potential of the isolate.

This study also discovered (for the first time) the biological role of a hypothetical protein CpsA (whose function has been at the prediction level), belonging to the *Bacillus* multispecies glyoxalase family protein and provided experimental evidence that CpsA catalysis the aromatic ring-cleavage of 2,6-dichloro-*p*-hydroquinone (a metabolite in the PCP degradation pathway) and convert the substrate to 2-chloromaleylacetate. Since CpsA is most similar to 2,6-dichloro-*p*-hydroquinone 1,2-dioxygenase of *Sphingobium chlorophenolicum* (Hayes et al., 2013; Machonkin & Doerner, 2011; Sun et al., 2011), the study proposed that CpsA name 2,6-dichloro-

p-hydroquinone 1,2-dioxygenase and it should be added to the group of hydroxyl quinone 1,2-dioxygenase cluster in the vicinal cholate superfamily as suggested in chapter nine.

The expression (heterologous) levels of each of the PCP-degrading enzymes of *B. cereus* strain AOA-CPS1 were relatively high, which means they could be produced in large quantity for bulk bioremediation process. Interestingly, metal ions such as iron (FE) stimulates the catalytic effectiveness of the PCP-degrading enzymes, which means that a concoction of the enzymes can be develop and used in bulk bioremediation of PCP co-contaminated with iron sites.

The whole-genome sequence analysis and metabolic model of *Bacillus cereus* AOA-CPS1 showed that the bacterium has been exposed to various kinds of agrochemicals including chlorophenolic compounds in the time past, in tandem with the reported incidences of chlorophenolic and other toxic priority pollutants in various samples across South Africa. The presence of stress response proteins in the genome of the organism, its ability to transform PCP efficiently, and the presence of many catabolic enzymes for degradation of several other xenobiotics in its metabolic model showed that the isolate has potential for bulk bioremediation of many recalcitrant compounds.

10.2 Limitations of the current study

- Though the isolate degraded PCP and congeners efficiently in both optimized and unoptimized minimal salt medium, in laboratory settings, however, the current study did not conduct any field test to evaluate the degradation potential of the strain on chlorophenol impacted sites.
- Though, the heterologous expression of the PCP-degrading enzymes of *B. cereus* strain AOA-CPS1 were relatively high, which meaning they could be produced in large quantity for bulk bioremediation process. However, most of the PCP-degrading enzymes are unstable, they lose activities in time-dependent manner, which may hamper their long time use in a continuous bioremediation and other biotechnological processes.
- The current study also lacks the experimental evidences to support the authenticity of the active sites of the PCP-degrading enzymes detected via structural and homology modelling.

10.3 Potential for future development of the study

The recent reports on the incidences of persistent organic compounds in different environments in Durban and other provinces in South Africa and other countries is worrisome and needs attention. The indigenous *Bacillus cereus* strain AOA-CPS1 evaluated in this study proved to have potential for bulk bioremediation of PCP and congeners based on laboratory studies, it would be interesting to evaluate the degradation potential of this isolate in soil and water mesocosms and on polluted sites and compare with the laboratory study. Regulation of the expression of the catabolic genes needs to be investigated.

Future research can be geared towards site directed mutagenesis of the proteins to ascertain their catalytic mechanisms and other factors that can stimulate or represses their expression. The pure proteins can also be crystalized to gain more knowledge of the structure of the enzymes. Future study can also be geared towards cloning all the catabolic genes into one vector using Gibson assembly (Gibson et al., 2009; Li & Lu, 2018) or golden gate cloning (Chiasson et al., 2019; Engler & Marillonnet, 2014) and use the recombinant clone for microbial degradation or expressed the protein heterologous.

The degrading enzymes can be optimized to improve their stability. The enzyme can also be immobilized to increase their re-usability. The immobilized enzymes can be used to make a concoction that can be use in enzymatic degradation of PCP and congeners, which is the goal of this research, since enzymatic degradation is more economical, efficient, faster and cheaper than the microbial degradation. The degrading enzymes can also be impregnated into a nano fibre and incorporate the fibre into a polluted site for bioremediation.

10.4 References

Adu-Kumi, S., Kawano, M., Shiki, Y., Yeboah, P. O., Carboo, D., Pwamang, J., … Suzuki, N. (2010). Organochlorine pesticides (OCPs), dioxin-like polychlorinated biphenyls (dl-PCBs), polychlorinated dibenzo-p-dioxins and polychlorinated dibenzo furans (PCDD/Fs) in edible fish from Lake Volta, Lake Bosumtwi and Weija Lake in Ghana. *Chemosphere*, *81*(6), 675–684. https://doi.org/10.1016/j.chemosphere.2010.08.018

ATSDR. (2017). Agency for Toxic Substances and Disease Registry. Substance priority list (candidates for toxicological profiles)., (October), 190. Retrieved from t www.atsdr.cdc.gov/SPL/resources.%0Aiii

Chanama, M., & Chanama, S. (2011). Expression of pentachlorophenol-degradative genes of Sphingobium chlorophenolica ATCC39723 in Escherichia coli. *Asian Journal of Public Health*, *2*(2), 78–83.

Chen L, & Yang J. (2008). Biochemical characterization of the tetrachlorobenzoquinone reductase involved in the biodegradation of pentachlorophenol. *International Journal of Molecular Sciences*, *9*(3), 198–212. https://doi.org/10.3390/ijms9030198

Chiaia-Hernandez, A. C., Keller, A., Wächter, D., Steinlin, C., Camenzuli, L., Hollender, J., & Krauss, M. (2017). Long-Term Persistence of Pesticides and TPs in Archived Agricultural Soil Samples and Comparison with Pesticide Application. *Environment International Journal*, *51*(18), 10642–10651. https://doi.org/10.1021/ACS.EST.7B02529

Chiasson, D., Giménez-Oya, V., Bircheneder, M., Bachmaier, S., Studtrucker, T., Ryan, J., … Parniske, M. (2019). A unified multi-kingdom Golden Gate cloning platform OPEN. *Scientific Reports*, *9*(10131), 1–11. https://doi.org/10.1038/s41598-019-46171-2

Dai, M. H., Rogers, J. B., Warner, J. R., & Copley, S. D. (2003). A previously unrecognized step in pentachlorophenol degradation in Sphingobium chlorophenolicum is catalyzed by tetrachlorobenzoquinone reductase (PcpD). *Journal of Bacteriology*, *185*(1), 302–310. https://doi.org/10.1128/JB.185.1.302-310.2003

Engler, C., & Marillonnet, S. (2014). Golden Gate cloning. *Methods in Molecular Biology*, *1116*, 119–131. https://doi.org/10.1007/978-1-62703-764-8_9

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343–345. https://doi.org/10.1038/nmeth.1318

Gong, X.; Qi, S.; Wang, Y.; Julia, E.B. & Lv, C. (2007). Historical contamination and sources of organochlorine pesticides in sediment cores from Quanzhou Bay, Southeast China. *Marine Pollution Bulletin*, *54*, 1434–1440.

Hayes, R. P., Green, A. R., Nissen, M. S., Lewis, K. M., Xun, L., & Kang, C. (2013). Structural characterization of 2,6-dichloro-p-hydroquinone 1,2-dioxygenase (PcpA) from Sphingobium chlorophenolicum, a new type of aromatic ring-cleavage enzyme. *Mol Microbiol*, *88*(3), 523–536. https://doi.org/10.1111/mmi.12204

Hlouchova, K., Rudolph, J., Pietari, J. M. ., Behlen, L. S., & Copley, S. D. (2012). Pentachlorophenol hydroxylase, a poorly functioning enzyme required for degradation of pentachlorophenol by Sphingobium chlorophenolicum. *Biochemistry*, *51*(18), 3848–3860. https://doi.org/10.1021/bi300261p

Hung, C.; Gong, C.; Chen, H.; Hsieh, H.; Santschi, P.H.; Wade, T.L. & Sericano, J.L. (2007). Relationships between pesticides and organic carbon fractions in sediment of the Danshui River estuary and adjacent areas of Taiwan. *Environmental Pollution*, *148*, 546 – 554.

IARC. (2019a). *International Agency for Research on Cancer monographs on the identification of carcinogenic hazards to humans. In: Agents Classified by the IARC Monographs, Volumes 1–124.* Retrieved from https://monographs.iarc.fr/agents-classified-by-the-iarc/

IARC. (2019b). *Pentachlorophenol and some related compounds. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans.* (Vol. 117). Retrieved from http://publications.iarc.fr/Book-And-Report-Series/Iarc-Monographs-On-The-Identification-Of-Carcinogenic-Hazards-To-Humans/Pentachlorophenol-And-Some-Related-Compounds-2019

Igbinosa, E. O., Odjadjare, E. E., Chigor, V. N., Igbinosa, I. H., Emoghene, A. O., Ekhaise, F. O., … Idemudia, O. G. (2013). Toxicological profile of chlorophenols and their derivatives in the environment: the public health perspective. *TheScientificWorldJournal*, *2013*, 1–11. https://doi.org/10.1155/2013/460215

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., … Bolton, E. (2019). PubChem 2019 update: improved access to chemical data. Nucleic Acids Res. 2019 Jan 8; 47(D1):D1102-1109. doi:10.1093/nar/gky1033. [PubMed PMID: 30371825].

Li, L., Jiang, W., & Lu, Y. (2018). A modified gibson assembly method for cloning large DNA fragments with high GC contents. In *Methods in Molecular Biology* (Vol. 1671, pp. 203–209). Humana Press Inc. https://doi.org/10.1007/978-1-4939-7295-1_13

Machonkin, T. E., & Doerner, A. E. (2011). Substrate Specificity of Sphingobium chlorophenolicum 2,6-Dichlorohydroquinone 1,2-DioxygenaseMachonkin, T.E., Doerner, A.E., 2011. Substrate Specificity of Sphingobium chlorophenolicum 2,6-Dichlorohydroquinone 1,2-Dioxygenase. Biochemistry 50, 8899–8913. *Biochemistry*, *50*(41), 8899–8913. https://doi.org/10.1021/bi200855m

Quinn, L. P., de Vos, B. J., Fernandes-Whaley, M., Roos, C., Bouwman, H., & Van den Berg, J. . (2011). *Pesticide use in South Africa: One of the largest importers of pesticides in Africa*. *Pesticides in the Modern World - Pesticides Use and Management*. https://doi.org/10.5772/16995

Silva, V., Mol, H. G. J., Zomer, P., Tienstra, M., Ritsema, C. J., & Geissen, V. (2019). Pesticide residues in European agricultural soils – A hidden reality unfolded. *Science of the Total Environment*, *653*, 1532–1545. https://doi.org/10.1016/j.scitotenv.2018.10.441

Stockholm convention. (2019). Stockholm convention on persistent organic pollutants. In *Draft guidance on best available techniques and best environmental practices for the production and use of pentachlorophenol listed with specific exemptions under the Stockholm Convention. (UNEP/POPS/COP.9/INF/16)*. Geneva. Retrieved from http://chm.pops.int/TheConvention/ConferenceoftheParties/Meetings/COP9/tabid/7521/Default.aspx

Sun, Wanpeng., Sammynaiken, Ramaswami., Chen, Lifeng., Maley, Jason., & Schatte, Gabriele. (2011). Sphingobium Chlorophenolicum Dichlorohydroquinone Dioxygenase (PcpA) Is Alkaline Resistant and Thermally Stable. *J Environ Qual.*, *7*, 1171–1179.

Thompson, L. A., Darwish, W. S., Ikenaka, Y., Nakayama M. M. S, Mizukawa, H., & Ishizuka, M. (2017, April 1). Organochlorine pesticide contamination of foods in Africa: Incidence and public health significance. *Journal of Veterinary Medical Science*. Japanese Society of Veterinary Science. https://doi.org/10.1292/jvms.16-0214

Verbrugge, L. A., Kahn, L., & Morton, J. M. (2018). Pentachlorophenol, polychlorinated dibenzo-p-dioxins and polychlorinated dibenzo furans in surface soil surrounding pentachlorophenol-treated utility poles on the Kenai National Wildlife Refuge, Alaska USA. *Environmental Science and Pollution Research International*, *25*(19), 19187–19195. https://doi.org/10.1007/s11356-018-2269-7