Biochemical and Structural characterization of ClpK from *Klebsiella pneumoniae.*

By
Tehrim Motiwala

Submitted in fulfilment of the requirement for the degree of

Master of Science

in

Biochemistry

in the
College of Agriculture, Engineering and Sciences at the
University of KwaZulu-Natal

Supervised by Dr Thandeka Khoza

Preface

The experimental work described in this dissertation was carried out in the School of Life Science, University of KwaZulu-Natal, Pietermaritzburg, from January 2020 to December 2021, under the supervision of Dr. Thandeka Khoza. The studies represent original work by the author and none of this work has been submitted for the award of any degree or examination at any university. All authors of data and any other information has been acknowledged accordingly by reference.

Student: Tehrim Motiwala (216013319)

Signature: 

Date: 10/02/2022

As the candidate's supervisor, I have approved this thesis for examination.

Supervisor: Dr. Thandeka Khoza

Signature: 

Date: 10/02/2022

Plagiarism Declaration

I, **Tehrim Motiwala**, declare that:

1. The research reported in this dissertation, except where otherwise indicated or acknowledged, is my original work.

2. This thesis has not been submitted in full or in part for any degree or examination to any other university.

3. This thesis does not contain other persons' 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, their writing has been placed inside quotation marks, and referenced.

5. This dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the Reference sections.

Student Name: Tehrim Motiwala

Signature: 

Date: 10/02/2022

Abstract

Klebsiella pneumoniae is a multi-drug resistant pathogen which is included into a group referred to as ESKAPE pathogens. This group consists of *Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, Pseudomonas aeruginosa* and *Enterobacter*. These bacterial species are responsible for more than 40% of infections in intensive care units therefore causing economic burden especially in developing countries. Pathogens have developed various virulence factors and mechanisms to survive harsh environmental conditions. One such mechanism is the use of caseinolytic (Clp) proteins which are found in bacteria, plants and fungi. These proteins function as a complex of catalytic (ClpP) and regulatory subunits (Clp ATPases) and play a role in maintaining cell protein homeostasis by unfolding and reactivating damaged or misfolded proteins. ClpK is a Clp ATPase which was recently identified from a clinical *K. pneumoniae* isolate which was responsible for an outbreak in Denmark. This newly identified protein was found to be present on a heat resistant locus and was therefore associated with the survival of *K. pneumoniae* at high temperatures. *In silico* and *in vitro* analysis were performed to investigate ClpK to understand the structure and functions of the protein. Firstly, bioinformatic analysis was used to investigate the distribution of Clp ATPases and the spread of ClpK across the *Klebsiella* species. Out of the investigated strains, only 34% contained the *clpk* gene. Homology modelling and circular dichroism spectroscopy (*in silico* and *in vitro*) showed that the protein was mainly α-helical. Additionally, homology modelling showed that ClpK was trimeric and contained a middle domain which was almost half the size of the ClpB middle domain. Molecular dynamic simulations performed to confirm the structural stability of modelled ClpK showed that the structure was stable, highly dynamic and flexible. *In silico* studies also showed that less than 40% of the ClpK protein sequence was disordered suggesting that ClpK could be expressed recombinantly and this was an important observation given that this protein has not been previously expressed, purified and characterised. The *clpk* gene was therefore synthesised and cloned into a pCold1 vector for recombinant expression. This was followed by the successful expression of ClpK in the presence of 0.25 mM IPTG at 15^{o}C as indicated by the band at 105 kDa which corresponds to the molecular weight of ClpK. ClpK was then purified to homogeneity using affinity and anion exchange chromatography. ClpK was confirmed to be biologically active using an

ATPase assay. This biological assay showed that ClpK hydrolysed ATP in a concentration dependent manner. The binding of ATP to ClpK was also confirmed using extrinsic fluorescence studies and the binding of fluorescent probes (mant-ATP and ANS) resulted in an increase in fluoresce quantum yield in the presence of ATP. Secondary structure analysis using circular dichroism showed that the binding of ATP did not cause a significant change in protein structure. The thermal stability profile of ClpK showed the melting temperature of ClpK to be approximately 65^{o}C.

Declaration – Research Outputs

Publication:

Motiwala, T.; Akumadu, B.O.; Zuma, S.; Mfeka, M.S.; Chen, W.; Achilonu, I.; Syed, K.; Khoza, T. Caseinolytic Proteins (Clp) in the Genus Klebsiella: Special Focus on ClpK. Molecules 2022, 27, 200. https://doi.org/10.3390/molecules27010200

In writing: **Motiwala, T.**; Mthethwa, Q.; Khoza, T. Targetting caseinolytic proteins in ESKAPE pathogens. Antibiotics.

Conference:

Biochemical and Structural characterization of ClpK from *Klebsiella pneumoniae.* **Tehrim Motiwala**, Thandeka N. Khoza, Ikechukwu Achilonu and Syed Khajamohiddin. SASBMB Conference Oral presentation (24/01/2021).

Student Signature: 

Date: 10/02/2022

Supervisor Signature: 

Date: 10/02/2022

Acknowledgements

First and foremost; I am thankful to my grandfather, Iqbal Modi, for his continued support and encouragement to pursue my dreams. His absence is greatly missed but his teachings will always guide me through.

I would also like to thank my grandmother, Mehrunissa Modi, for always believing that I am capable of achieving my dreams. I admire her strength and hope to be able to go through life with the same strength and resilience. Thank you to my parents, Majid and Siffiyya Mehmood for their constant support and encouragement.

I would like to thank my supervisor Dr Thandeka Khoza for her continued assistance, hard work, and guidance through the course of this study.

I would like to thank the National Research Foundation for their financial assistance.

I would also like to thank Professor Syed Khajamohiddin, Dr Ikechukwu Achilonu and all co- authors for sharing their expertise and giving me an opportunity to learn and master various skills in the field of bioinformatics and molecular biology at large.

I am also appreciative of the postgraduate students in Lab 39. Thank you for the hours we spent in the laboratory together. Our time spent together made my working environment extremely pleasant.

Table of Contents

List of Figures

Figure **Page no.**

Appendix

List of Tables

Table **Page no.**

Appendix

List of Abbreviations

2xYT	2 x yeast tryptone
3D	Three dimensional
°C	Degrees Celsius
x g	Relative centrifugal force
ATP	Adenosine triphosphate
ADP	Adenosine diphosphate
α	alpha
Cα	Alpha carbon atoms
aa	Amino acids
ANS	8-anilino-1-naphthalenesulfonic acid
Å	Armstrong
AAA+	**A**TPases **a**ssociated with diverse cellular **a**ctivities
bp	Base pairs
β	beta
BSA	Bovine serum albumin
CPS	Capsule polysaccharide
C-terminal	Carboxy terminal
CRE	Carbapenem-resistant *Enterobacteriaceae*
Clp	Caseinolytic proteins
CD	Circular dichroism
dNTPs	Deoxynucleoside triphosphate
DNA	Deoxyribose nucleic acid
D1	Domain 1
D2	Domain 2
ESKAPE	***E****nterococcus faecium,* ***S****taphylococcus aureus,* ***K****lebsiella pneumoniae,* ***A****cinetobacter baumannii,* ***P****seudomonas aeruginosa* and ***E****nterobacter*
EDTA	Ethylenediaminetetraacetic acid
ESBL	Extended-spectrum β-lactamase
γ	Gamma
g	Gram
Hsp	Heat shock protein

h	hour
iTOL	Interactive Tree of Life
IPTG	Isopropyl-β-D-thiogalactopyranoside
K	Kelvin
kb	Kilobases
kcal/mol	Kilocalories per molar
kDa	Kilodaltons
KPC	*Klebsiella pneumoniae* carbapenemase
LPS	Lipopolysaccharide
LB	Lysogeny Broth
.mae	Maestro
$MgAc^2$	Magnesium acetate
θMRE	Mean residue ellipticity
Tm	Melting temperature
Mant-ATP	Methylanthraniloyl-ATP
µg/mL	Microgram per millilitre
µL	Microlitre
µM	Micromolar
θmdeg	Millidegrees ellipticity
mg/mL	Milligram per millilitre
mL	Millilitre
mm	Millimetre
mM	Millimolar
min	Minute
M	Molar
MD	Molecular Dynamics
MWM	Molecular Weight Marker
MDR	Multi-drug resistant
MCS	Multiple cloning site
nm	Nanomolar
ns	Nanosecond
NCBI	National Center for Biotechnology Information
NDM	New Delhi metallo-*β*-lactamase

N-terminal	Amino terminal
NBD	Nucleotide Binding Domain
NPT	Isothermal-isobaric ensemble
NVT	Canonical ensemble
OD	Optical density
Omp	Outer membrane proteins
ps	Picosecond
PCR	Polymerase chain reaction
PDB	Protein data base
Rg	Radius of gyration
rpm	Revolutions per minute
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
sec	seconds
NaCl	Sodium Chloride
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Tris	tris(hydroxymethyl)aminomethane
units/mg	Units per milligram
units/L	Units per litre
WHO	World Health Organisation

Chapter 1: Introduction

1.1. *Klebsiella* species

The *Klebsiella* species are members of the *Enterobacteriaceae* family which consists of Gram-negative pathogens such as *Salmonella, Escherichia coli,* and *Shigella* (Paterson, 2006). The *Klebsiella* spp*.* are Gram-negative and ubiquitous in nature, they can be found in a variety of places, including; soil, sewage, plants, animals, on medical devices and they form a part of the natural human microbiome (Calfee, 2017, Martin and Bachman, 2018, Podschun and Ullmann, 1998). Infections caused by the *Klebsiella* spp. are associated with a significant mortality and morbidity rate and are responsible for 8% and 3% of endemic hospital infections and epidemic outbreaks, respectively (Podschun and Ullmann, 1998, Bengoechea and Sa Pessoa, 2018). Another major health concern with the *Klebsiella* spp. is the increase in the rates of carbapenem-resistant *Enterobacteriaceae* (CRE) among healthcare-associated *Enterobacteriaceae* species (Ma *et al.*, 2020). CRE have been listed as one of the top three critical-priority pathogens by the World Health Organisation (WHO) (Ma *et al.*, 2020). The *Klebsiella* genus includes three subspecies (subspecies *pneumoniae*, *ozaenae,* and *rhinoscleromatis*), *K. pneumoniae*, *K. oxytoca, K. aerogenes, K. michiganensis, K. granulomatis, K. mobilis, K. terrigena, K. quasipneumoniae* and *K. variicola* (Wareth and Neubauer, 2021, Barrios-Camacho *et al.*, 2019).

1.1.1. Klebsiella pneumoniae

Klebsiella pneumoniae are common aerobic, opportunistic, rod-shaped, and capsule enveloped pathogens which were first isolated and described by Carl Friedlander in 1882 (Doorduijn *et al.*, 2016, Ko *et al.*, 2002, Martin and Bachman, 2018, Li *et al.*, 2014). The *K. pneumoniae* genome consists of approximately 5000 to 6000 genes, about 1700 genes form a part of the core genome and are conserved in all members of the species (Wyres *et al.*, 2020). The remaining genes form a part of the accessory genome which divides the strains into three subsets, namely; opportunistic, hypervirulent and multi-drug resistant groups (Martin and Bachman, 2018, Li *et al.*, 2014, Wyres *et al.*, 2020). Additionally, genes in the accessory genome encode for functions such as nitrogen fixation, virulence factors and antibiotic-resistant enzymes and can be transferred through horizontal gene transfer in bacterial species (Martin and Bachman, 2018).

Opportunistic *K. pneumoniae* are responsible for infections such as urinary tract infections, and abdominal infections in immunocompromised patients in the nosocomial environment (Doorduijn *et al.*, 2016, Ko *et al.*, 2002, Martin and Bachman, 2018, Li *et al.*, 2014). In such environments, *K. pneumoniae* is transmitted through person-to-person contact or via contaminated hospital equipment such as endoscopes, catheters or endo tracheal tubes (Montgomerie, 1979). Between 2011 and 2014, *K. pneumoniae* and *Klebsiella oxytoca* were reported as being the third most common pathogens responsible for causing central-line associated bloodstream infections, ventilator-associated pneumonia, and surgical site infections (Calfee, 2017, Martin and Bachman, 2018). This highlights the global challenges of *Klebsiella* species concerning nosocomial infections.

Hypervirulent *K. pneumoniae* include multi-drug resistant (MDR) *K. pneumoniae* strains which are highly invasive and cause community-acquired infections such as meningitis and liver abscess in healthy people (Martin and Bachman, 2018, Li *et al.*, 2014). *K. pneumoniae* nosocomial infections spread effectively in a clinical environment and are chronic for the following reasons: firstly; the pathogen can form biofilms *in vivo* to protect itself against the host immune system and antibiotics, and secondly; many *K. pneumoniae* strains isolated from the nosocomial environment are found to contain multidrug-resistant phenotypes (Li *et al.*, 2019, Martin and Bachman, 2018). A recent study conducted in South Africa showed that extended-spectrum β-lactamase (ESBL) producing *K. pneumoniae* caused hospital-acquired, and healthcare-associated bloodstream infections in young children and infants, with high mortality rates (Buys *et al.*, 2016). It is worth noting that *K. pneumoniae* is second to *Escherichia coli* in causing bloodstream infections resulting in a case mortality rate of about 20% to 30%, and a population mortality rate of about 1.3 per 10 000 people (Martin and Bachman, 2018, Podschun and Ullmann, 1998). Another concerning threat with regards to *K. pneumoniae* infection is the identification of *K. pneumoniae* clinical isolates with genes conferring resistance to antibiotics (Safavi *et al.*, 2020). For example, *K. pneumoniae* strains containing the New Delhi metallo-*β*-lactamase (NDM) gene were identified in Asia (mainly China, India and Saudi Arabia), America and Africa (mainly South Africa) (Safavi *et al.*, 2020). This gene confers resistance to most β-lactams antibiotics (Safavi *et al.*, 2020). Furthermore, other studies have shown that some *K. pneumoniae* have also been identified as being resistant to “last resort”

antibiotics such as colistin and tigecycline (Van Der Weide *et al.*, 2020). The continuous increase in the number of *K. pneumoniae* strains (both opportunistic and hypervirulent) producing ESBLs and other antibiotic resistant genes such as carbapenemases and lactamases is a threat to global health since it results in unsuccessful treatment of *K. pneumoniae* infection using current conventional antibiotics (Ainoda *et al.*, 2019, Doorduijn *et al.*, 2016).

In addition to being resistant to a wide range of antibiotics, *K. pneumoniae* expresses two outer membrane porins which allow the entry of hydrophilic substances such as antibiotics or nutrients into the cell (Li *et al.*, 2014). In conjunction with the porins, the pathogen expresses AcrAB, an efflux pump which exports antibiotics and host-derived antimicrobial agents out of the cells, therefore antibiotics that may have entered through the porins are effectively removed (Li *et al.*, 2014).

The resistance of *K. pneumoniae* to different classes of antibiotics has led to the inclusion of this pathogen into a group of pathogens referred to as ESKAPE pathogens (Mulani *et al.*, 2019, Pendleton *et al.*, 2013). This group consists of ***E****nterococcus faecium,* ***S****taphylococcus aureus,* ***K****lebsiella pneumoniae,* ***A****cinetobacter baumannii,* ***P****seudomonas aeruginosa* and ***E****nterobacter* (Pendleton *et al.*, 2013, Ramsamy *et al.*, 2018, Santajit and Indrawattana, 2016). These bacterial species are responsible for more than 40% of infections in intensive care units therefore causing an economic burden especially in developing countries (Pendleton *et al.*, 2013, Ramsamy *et al.*, 2018, Santajit and Indrawattana, 2016). Furthermore, they pose a global health threat as these pathogens are known to evade a range of antibiotic biocidal effects (Pendleton *et al.*, 2013, Ramsamy *et al.*, 2018, Santajit and Indrawattana, 2016).

1.1.2. Other *Klebsiella* species

Recently, there is increasing evidence which shows that there are other members of the *Klebsiella* species causing a range of infections and these members have also been identified to be resistant to a number of antibiotics (Table 1).

Table 1: Members of the *Klebsiella* species which have been identified as infection causing and antibiotic resistant.

***Klebsiella* species**	**Infections caused**	**Antibiotic resistance**	**References**
Klebsiella aerogenes	Pneumonia Urinary and surgical site infections	MDR[1]	(Wesevich *et al.*, 2020, Ma *et al.*, 2020)
Klebsiella michiganensis	Nosocomial infections especially in at-risk infants	MDR Contain plasmid encoding the KPC[2]-3 gene Contain Chromosomally-encoded OXY-type 97 β-lactamase	(Chapman *et al.*, 2020, King *et al.*, 2021, Chen *et al.*, 2020)
Klebsiella oxytoca	Blood and respiratory infections in immunocompromised patients Development of antibiotic-associated haemorrhagic colitis in adults and youngsters	β-lactamases	(Darby *et al.*, 2014, Alexander *et al.*, 2020, Beaugerie *et al.*, 2003)
K. pneumoniae subsp. *pneumoniae*	Pneumoniae and mastitis in pigs Pneumonia, septicaemia and urinary infections in humans	Drug resistance capabilities have not been identified	(Bidewell *et al.*, 2018, Kwon *et al.*, 2016)
Klebsiella quasipneumoniae	Prevalent in human disease	KPC[1]-antimicrobial resistance gene	(Mathers *et al.*, 2019, Long *et al.*, 2017a)
K. variicola	Blood and urinary tract infections Infect surgical wounds	KPC[1] and NDM[3]-1 antimicrobial resistance genes	(Barrios-Camacho *et al.*, 2019, Rodríguez-Medina *et al.*, 2019, Long *et al.*, 2017a)

[1]Multi-drug resistant
[2] *Klebsiella pneumoniae* carbapenemase
[3] New Delhi metallo-*β*-lactamase

Wesevich *et al.* (2020) showed that 70% of patients with blood stream infections caused by *Klebsiella aerogenes* had poor clinical outcome. The patient either died before being discharged from the hospital, had recurrent infection, or experienced complications due to the infections (Wesevich *et al.*, 2020). Data from studies such as those of Wesevich *et al.* (2020) show that it is important to identify novel drug targets in pathogens. These targets can be proteins which are essential for the survival and pathogenicity of the pathogen and successfully inhibiting the activity of these proteins may lead to killing the pathogen.

1.2. Virulence factors

In order for infections to successfully occur, pathogens need to overcome immune defences which are inclusive of mechanical, chemical and cellular barriers (Bengoechea and Sa Pessoa, 2018, Paczosa and Mecsas, 2016). The pathogen overcomes these barriers in the body by using a variety of strategies such as cell surface modulation, mimicking host molecules and virulence factors (Finlay and Mcfadden, 2006). Virulence factors can either be membrane-associated, secretory or cytosolic (Sharma *et al.*, 2017). Membrane-associated virulence factors help the bacterium adhere to and evade host cells whereas secretory factors aid the pathogen in the evasion of the immune system (Sharma *et al.*, 2017). Cytosolic factors help the pathogen undergo metabolic, physiological and morphological changes (Sharma *et al.*, 2017).

Currently, there are four extensively studied virulence factors of *K. pneumoniae* namely; capsule polysaccharide (CPS), lipopolysaccharide (LPS), pili, and siderophores (Figure 1). Of these; CPS, LPS and pili are membrane-associated virulence factors and siderophores are cytosolic factors (Sharma *et al.*, 2017). Other virulence factors which have been identified but not sufficiently characterised include; outer membrane proteins (Omp), porins and efflux pumps (Paczosa and Mecsas, 2016).

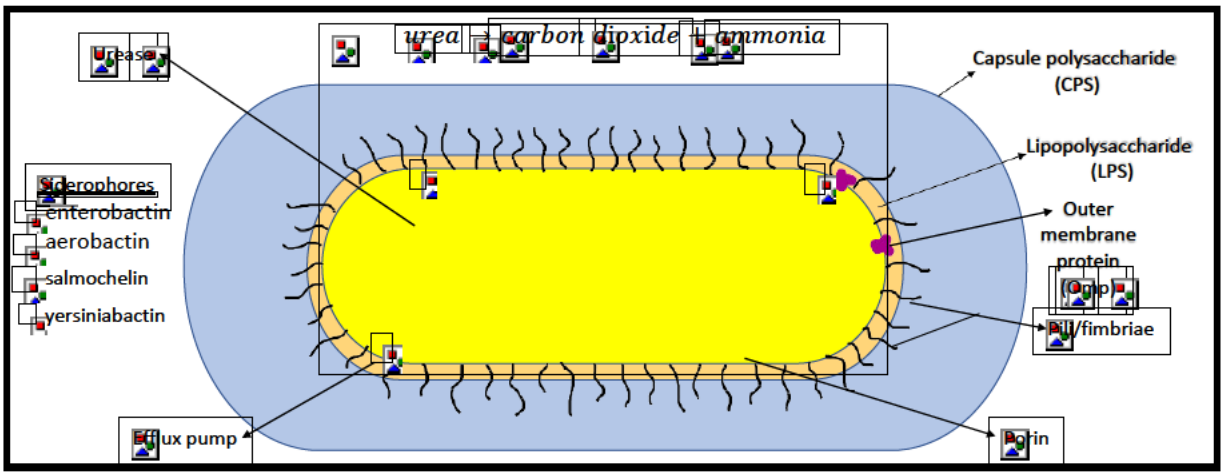


Figure 1: Virulence factors identified in *K. pneumoniae*. To overcome mechanical, chemical, and cellular barriers in the host, the pathogen expresses virulence factors.

The first stages of infection involve pathogen adhesion, invasion and colonisation of host cells (Ribet and Cossart, 2015). Pathogen adhesion involves interaction between pathogen surface proteins and host cells (Ribet and Cossart, 2015). These surface proteins play vital roles in cell adhesion, evasion and/or modulation of the host immune system (Ribet and Cossart, 2015). The interaction between the pathogen's surface proteins and host cells is important for disease establishment and it is therefore important to continuously study this interaction (Ribet and Cossart, 2015). To assist with bacterial adhesion, *K. pneumoniae* contains four types of structural proteins known as pili (also known as the fimbriae) on their surface, namely; type 1, type 3, Kpc and KPF-28 adhesin (Schroll *et al.*, 2010). Type 1 pili mediate the adhesion of the bacterial cell to structures which contain mannose on the host cell surface, these pili are expressed by 80% of all *K. pneumoniae* strains and have been reported to be important for the establishment of urinary tract infections (Li *et al.*, 2019, Wareth and Neubauer, 2021). Like type 1, type 3 pili is expressed by the majority (85%) of *K. pneumoniae* strains and in vitro studies have shown that type 3 pili facilitate binding of *K. pneumoniae* to epithelial cells, kidney and lung tissue to aid in the establishment of infections (Li *et al.*, 2019). Di Martino *et al.* (1996) showed that KPF-28 adhesion pili play a role in mammalian intestine colonization of *K. pneumoniae* cells using the human Caco-2 cell model.

Following the successful colonisation of host cells, pathogens have employed various mechanisms to evade the host defence system (Bengoechea and Sa Pessoa, 2018). To respond to infection and eradicate the invading pathogen, the host immune system dispatches phagocytic cells to engulf and destroy the pathogen and/or activate complement pathways to mediate bacterial cell lysis (Shao *et al.*, 2019, Ribet and Cossart, 2015). One of the strategies that *K. pneumoniae* uses to evade phagocytotic cells such as macrophages, dendritic cells and neutrophils is coating or enveloping the cell with an extracellular polysaccharide matrix referred to as CPS (Podschun and Ullmann, 1998). This acidic polysaccharide shields the pathogen in a number of ways (Podschun and Ullmann, 1998, Wu *et al.*, 2008). For example, it prevents opsonisation and phagocytosis, it hinders the bactericidal action of antimicrobial peptides such as defensins and lactoferrin secreted by host immune cells, and lastly it blocks complement-mediated lysis (Wu *et al.*, 2008). Studies on animal models have shown that acapsular *K. pneumoniae* strains are dramatically less virulent than hypercapsule encapsulated strains (Wu *et al.*, 2008). This was supported by low bacterial loads in the lungs, and an inability of the bacteria to spread systemically in animals challenged with acapsular *K. pneumoniae* (Wu *et al.*, 2008). Also, lower mortality was seen in animals challenged with the acapsular strain than those with the hypercapsule encapsulated strains (Wu *et al.*, 2008). This suggests that virulence may be more dependent on the thickness of the capsule than it is on the chemical composition of the capsule (Wu *et al.*, 2008). Bacteria which survive phagocytosis and intracellular killing then travel around the body within the phagocytic cells and cause infection in organs such as the liver or the lungs (Li *et al.*, 2014). The pathogen may also secrete free CPS which traps antimicrobial peptides secreted by host immune cells to decrease the amount of peptides that reach the bacterial surface (Li *et al.*, 2014).

In addition to being coated with CPS, *K. pneumoniae* also contains a layer of LPS (Figure 1) which inhibits complement-mediated killing and the action of antimicrobial peptides (Li *et al.*, 2019). LPS demonstrates the complexity of the defence system *K. pneumoniae* uses against the host immune system. LPS is made up of three components, namely; hydrophobic lipid A anchored to the outer membrane, the O-antigen, and a core polysaccharide which connects Lipid A and the O-antigen (Li *et al.*, 2014). The O-antigen inhibits the access of the host immune cells to activators on the bacterial cell surface, and therefore plays a role in bacterial resistance (Li *et al.*,

2014). To date there are nine types of O-antigens identified in *K. pneumoniae*, and each antigen plays a significant role in the protection of the pathogen against complement-mediated killing (Li *et al.*, 2019, Wareth and Neubauer, 2021).

After escaping immune system cells, pathogens need to survive inside the host (Bengoechea and Sa Pessoa, 2018). Bacteria need iron to grow and have developed mechanisms to obtain iron in mammalian hosts where iron is generally bound to various transport proteins (Li *et al.*, 2019). One such mechanism is the production of various siderophores. Siderophores are molecules that have a higher affinity for iron than transport proteins do therefore, these molecules help pathogens salvage iron from transport proteins (Bengoechea and Sa Pessoa, 2018, Li *et al.*, 2019). *K. pneumoniae* secretes a siderophore known as enterobactin to access iron off iron-rich host cells (Bengoechea and Sa Pessoa, 2018, Paczosa and Mecsas, 2016). The host in turn secretes lipocalin 2 to inhibit the binding of the siderophores to the host cells (Bengoechea and Sa Pessoa, 2018). To “outrun” this host defence, *Klebsiella* strains additionally secrete other siderophores namely; aerobactin, salmochelin and yersiniabactin (Bengoechea and Sa Pessoa, 2018). Enterobactin and aerobactin solubilize and import the iron ions needed during infection (Wareth and Neubauer, 2021). Strains which produce only yersiniabactin are unlikely to cause infections of the lungs as the activity of this particular siderophore is inhibited when the host produces transferrin (Paczosa and Mecsas, 2016). Analysis of blood cultures and urine in hospitalized patients in Munich showed that yersiniabactin was produced by 17.7% of *K. pneumoniae* strains (Wareth and Neubauer, 2021). Additionally, *K. pneumoniae* produces cytoplasmic urease to suffice itself with nitrogen for growth much like many other gut pathogens (Li *et al.*, 2014). Although we have a brief understanding of the virulence factors of *K. pneumonia,* it is important to further investigate potential targets which could be virulence or survival factors which could be used to inhibit or reduce infections caused by this pathogen (Li *et al.*, 2014).

1.3. Caseinolytic proteins

Caseinolytic proteins (Clp) are found in bacteria, fungi, in the mitochondria of eukaryotes and in the chloroplast of plants. These proteins play a role in cell protein quality control and are therefore important in living cells (Ahyoung *et al.*, 2015). To maintain a balance of proteins, all living cells use various mechanisms to eliminate

misfolded or abnormal proteins, maintain amino acid pools, and reactivate proteins damaged due to stressful conditions (Ahyoung *et al.*, 2015, Capestany *et al.*, 2008, Frees *et al.*, 2014).

The association of misfolded or partially folded polypeptides is known as protein aggregation, protein aggregates usually form due to cellular stress such as heat shock or a change in pH (Zolkiewski *et al.*, 2012, Frees *et al.*, 2007). A build-up of protein aggregates reduces the ability of protease systems to degrade proteins and can result in cell death (Zolkiewski *et al.*, 2012). It is therefore important for the cell to remove unfolded proteins for cellular functions and growth to continue (Frees *et al.*, 2014). To do this, cells contain Clp proteins which function as a complex of catalytic (ClpP) and regulatory subunits (further referred to as Clp ATPases) (Ahyoung et al., 2015).

Most bacteria contain a single ClpP subunit, however *Mycobacterium tuberculosis* has been found to contain ClpP homologs which are referred to as *clpp1* and *clpp2* (Raju *et al.*, 2012). ClpP is a serine protease composed of two heptameric rings (Raju *et al.*, 2012). These rings form a barrel-shaped structure referred to as a tetradecamer which encloses a protease active site (Brötz-Oesterhelt and Sass, 2014, Capestany *et al.*, 2008). The catalytic triads in the active site consist of three amino acids, namely serine, histidine and aspartic acid (Brötz-Oesterhelt and Sass, 2014). ClpP can adopt two conformations, that is; the closed/inactive and open/active conformation (Brötz-Oesterhelt and Sass, 2014, Capestany *et al.*, 2008). In the closed conformation the enzyme is inactive because the catalytic triad is misaligned therefore no protein degradation may take place (Capestany *et al.*, 2008). In order for ClpP to adopt an active conformation, it needs to be bound to a Clp ATPase (Capestany *et al.*, 2008) (Figure 2). The binding of a Clp ATPase to ClpP results in the alignment of the catalytic triad residues, this results in conformational changes which leads to the opening of the cavity to allow for the substrate to access the active site for protein degradation to occur (Brötz-Oesterhelt and Sass, 2014). Therefore, Clp ATPases play a very important role in the function and regulation of ClpP activity and ultimately protein degradation (Brötz-Oesterhelt and Sass, 2014, Capestany *et al.*, 2008).

Figure 2: The function of Clp proteins. The target protein is tagged for recognition by the Clp ATPases by degrons. The tagged protein is recognised by a Clp/Hsp 100 subfamily member, **a)** a member of the ClpB/Hsp104 subfamily along with the DnaK system, unfolds the protein and refolds it into its functional conformation, **b)** the protein is either refolded into functional conformation by a member of the ClpA subfamily, or the protein is directed to ClpP for degradation. Adapted from Maurizi and Xia (2004).

Clp ATPases are divided into subfamilies, namely; the ClpA subfamily and the ClpB/Hsp104 subfamily (Maurizi and Xia, 2004). Axiomatically, enzymes that catalyse the reaction to break down adenosine triphosphate (ATP) to form adenosine diphosphate (ADP) and free phosphate ions are referred to as ATPases (Abad, 2011). Furthermore, these enzymes use the energy released from the breakdown of ATP to perform other cellular reactions (Abad, 2011). The members of the ClpA subfamily form hexameric complexes when bound to ATP (Miller *et al.*, 2018). The ATP bound conformation of Clp ATPases adopts a central channel of the ring structure which then binds, unfolds and reactivates tagged proteins (Miller *et al.*, 2018) (Figure 2). To unfold protein aggregates, Clp ATPases use the energy generated from ATP hydrolysis to generate a tugging force on the protein to be unfolded (Miller *et al.*, 2018). The members of the ClpA subfamily also form complexes with ClpP where the Clp ATPase unfolds protein aggregates and directs unfolded proteins to the ClpP proteolytic barrel for degradation (Maillard *et al.*, 2011, Marsee *et al.*, 2018, Capestany *et al.*, 2008) (Figure 2).

Members of the ClpB/Hsp104 subfamily assist in the resolubilisation and reactivation of proteins damaged due to stressful conditions (Ahyoung *et al.*, 2015). ClpB proteins do not contain the tripeptide consensus sequence (IGF) needed for interaction with ClpP, it is therefore believed that these proteins act as chaperones under stressful conditions to prevent protein unfolding or assist in protein disaggregation (Frees *et al.*, 2014, Sauer and Baker, 2011) (Figure 2).

1.3.1. Structure of Clp ATPases

Clp proteins belong to the AAA+ (**A**TPases **a**ssociated with diverse cellular **a**ctivities) superfamily of proteins (Zolkiewski *et al.*, 2012). Proteins belonging to the AAA+ superfamily are characterised by (a) their ATP dependent mechanisms in unfoldase activity (Frees *et al.*, 2014, Neuwald *et al.*, 1999, Maillard *et al.*, 2011), (b) a family-specific N-terminal domain (Wojtyra *et al.*, 2003, Thibault *et al.*, 2006) and (c) the number of conserved ATPase/AAA+ domains (Sauer and Baker, 2011). The AAA+ superfamily are further divided into families or subgroups based on structural organisation and structural motifs in their sequence (Schirmer *et al.*, 1996). Clp ATPases are grouped into the heat shock protein (Hsp) 100 family which have the unique ability to promote the resolubilisation of protein aggregates (Ahyoung *et al.*, 2015, Schirmer *et al.*, 1996).

The Clp ATPase protein sequence contains highly conserved ATPase/AAA+ domains which are approximately 200 to 250 amino acids (aa) in length (Zolkiewski *et al.*, 2012, Thibault *et al.*, 2006, Snider *et al.*, 2008). This conserved structural feature is used to group the Clp ATPases into two classes based on the number of ATPase domains present in their sequence (Schirmer *et al.*, 1996). Clp ATPase class I contains two ATPase domains whereas class II contains one ATPase domain (Zolkiewski *et al.*, 2012, Thibault *et al.*, 2006, Snider *et al.*, 2008). Therefore, members of class I are considerably larger (68 to 110 kDa) compared to members of class II (40 to 50 kDa) (Dods, 2020). Members belonging to class I and class II are shown in Table 2.

Table 2: Clp ATPases identified across different species, exhibiting diverse functions.

Class I			
Clp regulatory subunits	**Species**	**Functions**	**References**
ClpA[1]	Gram-positive Proteobacteria	Protein quality control	(Ingmer *et al.*, 1999)
ClpB	Prokaryotes, yeast, and plants	Disaggregation of stress-damaged proteins	(Lee *et al.*, 2003, Pietrosiuk *et al.*, 2011, Zheng *et al.*, 2002)
	Porphyromonas gingivalis	Intracellular replication and survival	(Capestany *et al.*, 2008)
ClpC	Gram-positive bacteria (Firmicutes and Actinobacteria) and Cyanobacteria	Protein quality control, red blood cell lysis, regulate expression of virulence factors	(Ingmer *et al.*, 1999, Nair *et al.*, 2001)
ClpD	Chloroplasts of higher plants	Molecular chaperone	(Zheng *et al.*, 2002)
ClpE	Firmicutes	Thermotolerance, cell division and virulence	(Ingmer *et al.*, 1999)
ClpK	*K. pneumonia*	Thermotolerance	(Bojer *et al.*, 2013)
ClpL	*Streptococcus pneumoniae*	Nucleotide phosphohydrolase activity, stabilises unfolded proteins, prevents protein aggregation	(Park *et al.*, 2015)
ClpV	Gram-negative bacteria	Component of the type V1 secrection system	(Thibault *et al.*, 2006)
Class II			
Clp regulatory subunits	**Species**	**Functions**	**References**
ClpM	*Mus musculus*	Protein quality control	(Ali and Baek, 2020, Zheng *et al.*, 2002)
ClpN	*Pseudomonas aeruginosa*	Cell division	(Ali and Baek, 2020, Zheng *et al.*, 2002)
ClpX	Proteobacteria, Firmicutes and Thermatogae	Protein quality control, cell division, heat tolerance and virulence	(Ingmer *et al.*, 1999, Capestany *et al.*, 2008)
ClpY	Gram-positive Proteobacteria	Cell division, heat shock response, and capsule transcription	(Ingmer *et al.*, 1999)

[1]ClpA and ClpC are orthologs, bacteria usually contain either one of these (Ingmer *et al.*, 1999).

ATPase domains are also known as nucleotide binding domains (NBD) and are divided into two subdomains, namely; the large **α/**β domain and the small **α** domain (Maurizi and Xia, 2004, Walker *et al.*, 1982) (Figure 3). These ATPase domains are essential for stability, substrate recognition, ATP hydrolysis and oligomerization in Clp

ATPases (Walker *et al.*, 1982). Furthermore, the ATPase domains contain Walker A and B motifs which are common in all ATPase proteins and the domains contain additional functional motifs known as sensor motifs (Walker *et al.*, 1982, Maurizi and Xia, 2004) (Figure 3). The Walker A motif binds the ATP phosphate group, whereas Walker B motif is involved in binding of metal and plays a role in ATP catalysis (Walker *et al.*, 1982, Wojtyra *et al.*, 2003). Additionally, the Walker A motif forms the floor of the nucleotide-binding pocket, while the Walker B region forms a loop that overlays the pockets (Schirmer *et al.*, 1996). As mentioned before, proteins in Clp ATPase class I contain two NBDs, the domain situated closer to the N-terminus is referred to as Domain 1 (D1), and the domain situated closer to the C-terminus is referred to as Domain 2 (D2) and the Clp ATPase class II contains D1 (Maurizi and Xia, 2004, Wojtyra *et al.*, 2003, Miller *et al.*, 2018, Schirmer *et al.*, 1996) (Figure 3).

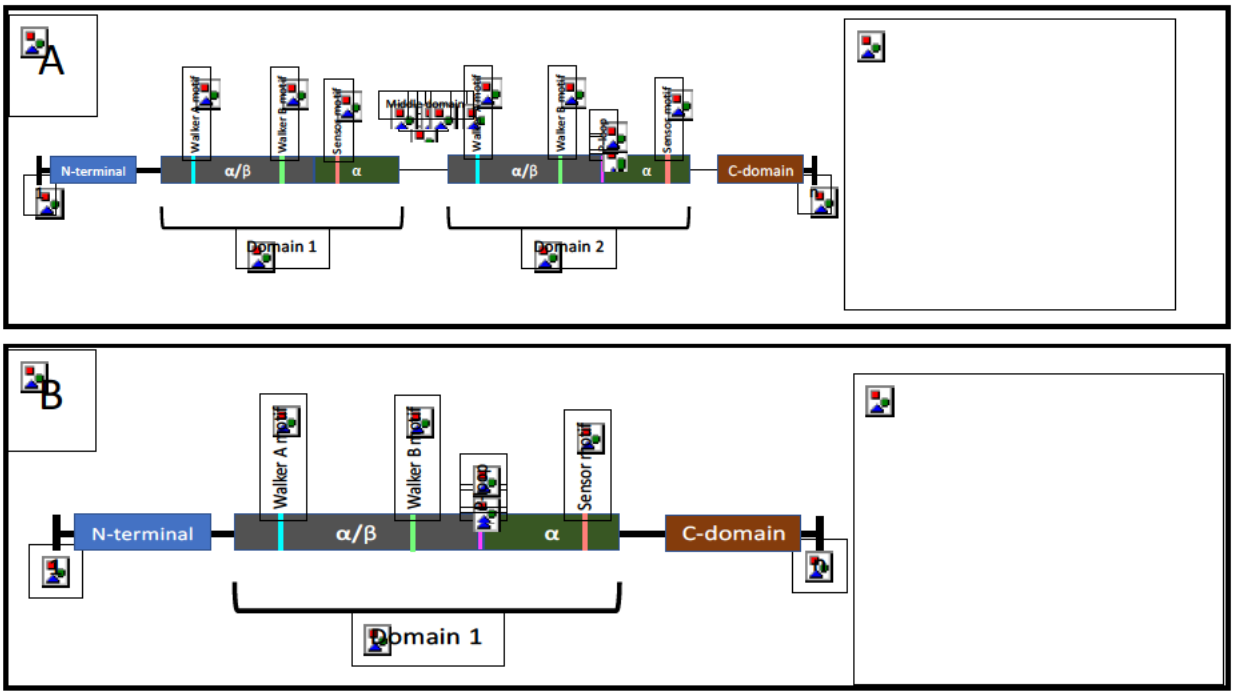


Figure 3: Schematic representation of the general structural features of Class I and Class II Clp ATPases from prokaryotes. A) Class I Clp ATPases contain two nucleotide binding domains (NBD) referred to as domain 1 and domain 2. A middle domain has been identified to be present in certain Clp ATPases and play a role in separating the two NBDs. The tertiary structure (PBD: 1KSF) shows that the protein is mainly α-helical. **B)** Class II Clp ATPases contain one NBD. The tertiary structure (PBD: 6SFW) shows that the protein is mainly α-helical. Each domain contains Walker A and Walker B motifs. Certain Clp regulatory protein contain a tripeptide (represented in pink) required for interaction with ClpP. n represents the number of amino acids in the sequence. Adapted from Maurizi and Xia (2004), Baker and Sauer (2012). Tertiary structures; red- N-terminal domain, blue- Domain 1, orange- Domain 2, green- C-terminal domain.

As much as the Clp ATPase domains are conserved across species, the following has been observed, firstly; there is a slight difference in the amino acid sequences of the NBDs between class I and class II (Schirmer *et al.*, 1996) and secondly; Domain 2 in class I Clp ATPases has been found to contain an extended carboxyl terminal domain which contains conserved serine and threonine residues that partake in nucleotide hydrolysis (Bajaj and Batra, 2012). The carboxyl terminal domain is crucial for the oligomerization of ATPase to a hexamer, and therefore also called the oligomeric domain (Bajaj and Batra, 2012, Kar *et al.*, 2008). These subtle differences in the NBD between classes results in different ATP binding profiles and ATPase activity (Maurizi and Xia, 2004, Wojtyra *et al.*, 2003, Miller *et al.*, 2018, Schirmer *et al.*, 1996).

The N-terminal domain of Clp ATPases precedes the first NBD and is approximately 150 aa in length (Mogk *et al.*, 2003) (Figure 3). Studies have suggested that the N-terminal domain may play a role in substrate binding. Additionally, studies have also found that in the absence of this domain, protein activity is decreased, suggesting that the N-terminal domain is important in the structure thus the function of Clp ATPases (Lo *et al.*, 2001). However other studies have shown that even in the absence of the N-terminal domain, ClpB undergoes oligomerisation and performs disaggregation functions therefore, a conclusive decision on the function and importance of the N-terminal domain is yet to be made (Mogk *et al.*, 2003). Interestingly, the N-terminal domain has been reported to have the ability to fold independently of the mainly α-helical ATPase into an all α helical domain dimer (Wojtyra *et al.*, 2003, Schirmer *et al.*, 1996).

1.3.2. Clp ATPase in *Klebsiella pneumoniae*

There are five chromosomally encoded Clp proteins usually expressed by Gram-negative bacteria namely; ClpA, ClpB, ClpX, ClpP and ClpK (Bojer *et al.*, 2013). ClpK (K for *Klebsiella*) is a recent member of the Clp ATPase family which was isolated from *K. pneumoniae* strain C132-98 from an outbreak in Denmark (Bojer *et al.*, 2010, Bojer *et al.*, 2013, Jørgensen *et al.*, 2016). Genome analysis of this isolate identified the *clpK* gene to be situated on a conjugative plasmid which confers multiple antibiotic resistances (resistance towards tetracycline, trimethoprim, neomycin, third-generation cephalosporins and sulfamethoxazole) in *K. pneumonia* (Bojer *et al.*, 2010, Bojer *et al.*, 2013). Based on the observation that the *clpk* gene was situated in the heat

resistant locus and belonged to the Hsp family based on sequence analysis, Bojer *et al.* (2010) suggested that the protein (ClpK) encoded by this gene could be responsible for the thermoresistance capability of *K. pneumoniae* which caused the outbreak. This suggestion by Bojer *et al.* (2010) can further be supported by the fact that Hsp proteins play a role in assisting cells recover from stressful environmental conditions such as high temperature (Jolly and Morimoto, 2000). These researchers further reasoned that ClpK may allow *K. pneumoniae* strains to survive in the hospital environment or in other environments where its thermostable properties would serve as an advantage since the *clpk* gene was found in 30% of a collection of clinical isolates (Bojer *et al.*, 2010, Bojer *et al.*, 2013). To confirm the thermotolerance properties conferred by *clpk,* Bojer *et al.* (2010) incubated *K. pneumoniae* with and without *clpk* at 53^{o}C over a period of 50 minutes and investigated the percentage survival of the bacteria. The *K. pneumoniae* cells containing *clpk* survived the heat shock treatment in comparison to those that did not express *clpk* (Bojer *et al.*, 2010). This finding was important given that past research had demonstrated thermostability of *K. pneumoniae* although thermostability properties had not been attributed to a particular gene or protein at that time (Verrips *et al.*, 1979).

A closer inspection of the primary amino acid sequence of ClpK shows that the protein shares identity with bacteria from all proteobacterial subdivisions, except division ε, suggesting that ClpK evolved due to horizontal transfer (Bojer *et al.*, 2010). The primary amino acid sequence also shows that ClpK is a class I Clp ATPase as it contains two NBDs however it also contains structural features which are unique to this protein. Firstly, it contains a 100 amino acid N-terminal extension which is unique to the protein (Figure 4). As mentioned previously the primary function of the N-terminal domain is to stabilise Clp ATPases (Lo *et al.*, 2001). However, the function of the 100 amino acid N-terminal extension is yet to be investigated. Secondly, ClpK contains an intermediate length linker region which separates the two NBDs (Bojer *et al.*, 2013) (Figure 4). Like the N-terminal extension, the role of this linker region is yet to determined. Thirdly, like ClpB, ClpK lacks the ClpP tripeptide sequence which is essential for ClpP interaction (Amor *et al.*, 2019) (Figure 4). The absence of this motif may suggest that (a) ClpK does not interact with ClpP to degrade proteins that cannot be unfolded or (b) ClpK contains an alternative site for ClpP interaction and this site is yet to be identified (Amor *et al.*, 2019). It could be speculated that the N-terminal

extension may play a role in ClpK and ClpP interaction. Another interesting observation on the ClpK sequence, is the presence of the zinc binding motif which is only present in Gram-positive bacteria (Wojtyra *et al.*, 2003) and it is worth mentioning that no ClpK homologues were identified in Gram-positive bacteria (Bojer et al., 2010) (Figure 4). This aligns with the suggestion that the proteobacterial ClpC evolved into ClpK due to the absence of the ClpP binding motif, and the presence of a zinc binding pattern in both proteins (Bojer *et al.*, 2010, Miller *et al.*, 2018). The zinc binding domain in the N-terminal domain of Clp ATPase stabilises the N-terminal domain which is important for substrate recognition and binding (Maurizi and Xia, 2004).

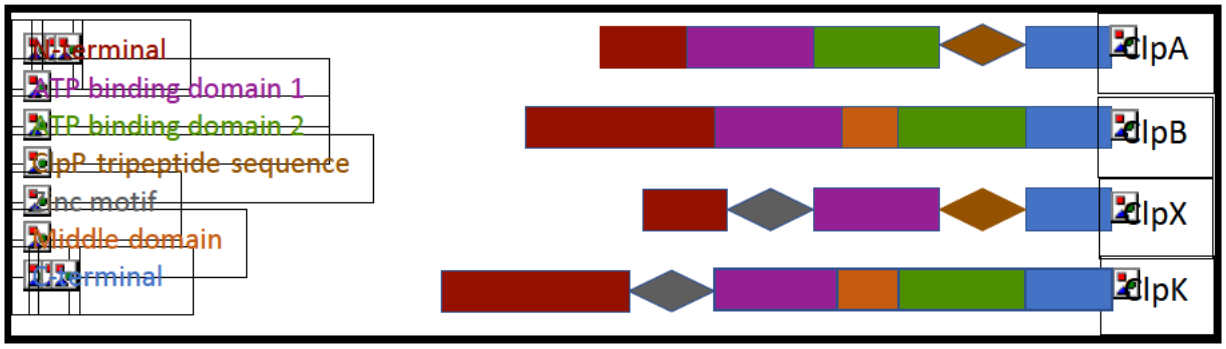


Figure 4: Primary domain structures of four Clp ATPases. ClpB and ClpX are found in both Gram-positive and Gram-negative bacteria. ClpA and ClpK is found in Gram-negative bacteria (Bojer *et al.*, 2010, Miller *et al.*, 2018).

1.4. Clp ATPases as drug targets

Clp ATPases contribute towards the survival of pathogenic bacterial species and are therefore ideal drug targets to treat a number of diseases (Bhandari *et al.*, 2018). Till date, compounds with either anti-virulent or antibacterial activities have been identified to target Clp ATPases in *Mycobacterium tuberculosis*, *Plasmodium falciparum* and *Staphylococcus aureus* (Bhandari *et al.*, 2018, Gao *et al.*, 2015, Kazmaier and Junk, 2021, Fetzer *et al.*, 2017). The identified compounds target Clp ATPases differently, for example; one compound inhibits the interaction between a Clp ATPase and ClpP while another compound targets the N-terminal domain of a Clp ATPase (Kazmaier and Junk, 2021, Gao *et al.*, 2015).

Compounds developed to be potential targets for Clp ATPases include; 334, cyclomarin A, lassomycin, ecumicin, rufomycin, hydantoin analogs and armeiaspirols (Table 3).

Table 3: Compounds developed as potential targets for Clp ATPases.

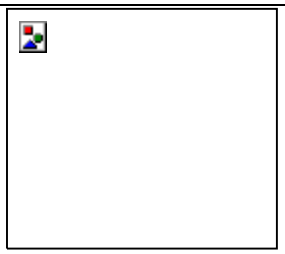
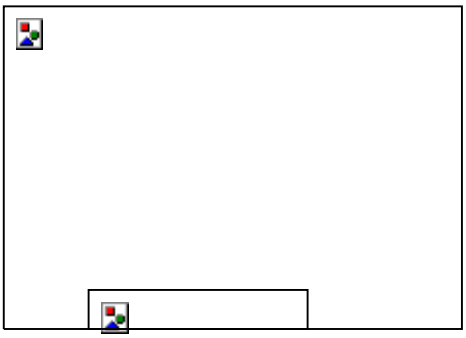
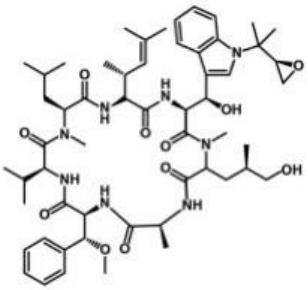
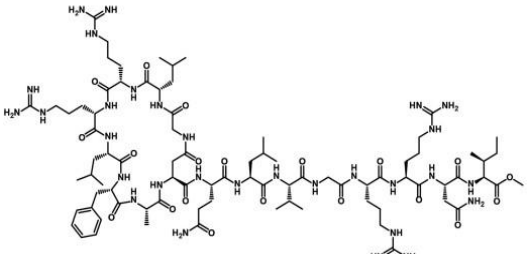
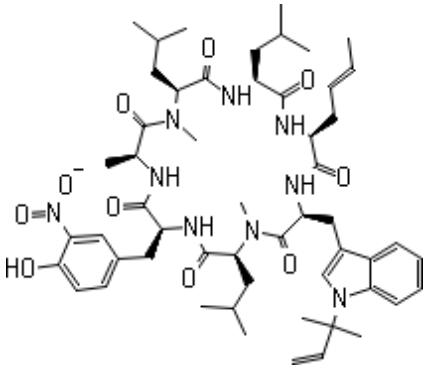
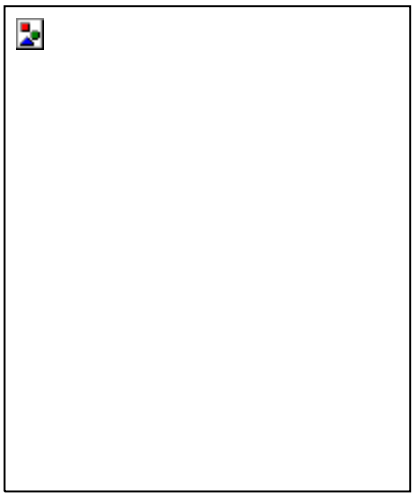
Compound	Structure[1]	Mechanism of action	References
334		Deoligomerization of *S. aureus* ClpX, disrupts the ClpXP complex and blocks ClpX ATPase activity. *S. aureus* produces lower levels of toxin in its presence.	(Fetzer *et al.*, 2017, Bhandari *et al.*, 2018)
Ecumicin		Binds to the N-terminal domain of ClpC1 of *M. tuberculosis.* Stimulates the ATPase hydrolysis activity of *M. tuberculosis* ClpC1 and at the same time decouples ClpC1 and ClpP therefore inhibiting proteolytic activity and resulting cell death.	(Bhandari *et al.*, 2018, Gao *et al.*, 2015)
Cyclomarin A		Binds to the N-terminal domain of ClpC1 of *M. tuberculosis* and prevents the movement of the N-terminal domain. Causes excessive proteolysis.	(Gavrish *et al.*, 2014, Bhandari *et al.*, 2018, Wolf *et al.*, 2019)
Lassomycin		Binds to an acidic N-terminal pocket on ClpC1. Stimulates ATPase activity of ClpC1 from *M. tuberculosis,* however it also inhibits ATP-dependent degradation of proteins. Uncouples ClpC1 from ClpP1 and ClpP2, resulting in the death of the cell as unnecessary protein build up.	(Gavrish *et al.*, 2014, Bojer *et al.*, 2010)

Table 3 continued...

Rufomycin		Interacts with the N-terminal domain of ClpC1 of *M. tuberculosis.* Decreases the proteolytic activity of the ClpC1 and ClpP complex therefore resulting in the build-up of proteins in the cell.	(Wolf *et al.*, 2019)
Armeiaspirols		Inhibits ClpXP and ClpYQ in *Bacillus subtilis* by binding to the ATPase domains and therefore inhibits the function of the complexes. Inhibits ATP hydrolysis and proteolysis.	(Labana *et al.*, 2021)
Hydantoin analog		Inhibits the ClpXP complex. Binds to a binding pocket on ClpP and impairs complex substrate turnover.	(Fetzer *et al.*, 2019)

[1]Adapted from Labana *et al.* (2021), Meßner *et al.* (2021), Fetzer *et al.* (2017), Bhandari *et al.* (2018), Fetzer *et al.* (2019).

Many of the compounds being investigated are highly specific for *Mycobacteria* Clp ATPases. For example lassomycin was active against multi-drug resistant and extensively drug resistant *M. tuberculosis* however, the compound was inactive against *S. aureus, Bacillus anthracis* and *K. pneumoniae* (Gavrish *et al.*, 2014). This is an interesting observation as one would expect Clp ATPases across different species to be inhibited as Clp ATPases are known to contain conserved regions (Schirmer *et al.*, 1996). To target Clp ATPases across species, a compound which targets specific motifs on Clp ATPases rather than trying to target a specific Clp ATPase could potentially be developed.

1.5. Objectives

K. pneumoniae is a multi-drug resistant pathogen which is responsible for a number of nosocomial infections, particularly those which effect immunocompromised patients. Infections caused by this pathogen have high morbidity and mortality rates and it is therefore important to find potential targets to inhibit the growth of this pathogen. One such target is a novel Clp ATPase which was identified as being present in *K. pneumoniae.* This Clp ATPase (ClpK) was found to confer thermoresistance to a clinical isolate. The first step to developing compounds to target the protein includes protein investigation to understand protein function.

To understand protein function, the protein would have to be expressed and purified before it is subjected to biophysical characterisation. Biophysical characterisation of proteins is important as it allows for the development of drugs and targets that could be used to inhibit the protein and potentially inhibit the pathogen in which it is identified (Neet and Lee, 2002, Vedadi *et al.*, 2010). Furthermore, biophysical characterisation also allows us to understand the structure, stability, and function of the protein in question (Vedadi *et al.*, 2010).

Seeing the importance of ClpK in the survival of *K. pneumoniae*, this study aimed to investigate the novel ClpK ATPase. The objectives of the study were as follows:

- Investigating the presence of Clp ATPases and the spread of ClpK amongst the *Klebsiella* species
- Modelling the theoretical structure of ClpK
- Molecular dynamic simulations to confirm the modelled structure and assess the dynamic nature of the protein
- Overexpression and purification of recombinant ClpK
- Biochemical and physical characterisation of ClpK

Chapter 2: Materials and Methods

2.1. Species and Databases

The National Center for Biotechnology Information (NCBI) genome database was used to obtain protein genomes of 100 *Klebsiella* strains (7 species, complete draft). The 100 strains included in analysis were: eight *K. aerogenes* strains, one *K. michiganensis* strain, seven *K. oxytoca* strains, 57 *K. pneumoniae* strains, 15 *K. pneumonia* subsp. *pneumoniae* strains, three *K. quasipneumoniae* strains, and nine *K. variicola* strains. The list of the *Klebsiella* species, strains, web-links, and references are presented in the appendix (Table A1).

2.2. Genome data mining and annotation of Clp ATPases

The protein genomes obtained from the NCBI genome database were used to mine Clp ATPases. Each genome file was individually searched for the presence of Clp ATPases, the sequence for each Clp ATPase was separated from the main file and used for further analysis.

2.3. Phylogenetic analysis of Clp ATPases

The method described by Ngcobo *et al.* (2020) was used for the phylogenetic analysis of the separated Clp ATPases. Briefly, the protein sequences were aligned using MAFFT v6.864 embedded on the Trex Web Server (Letunic and Bork, 2019, Boc *et al.*, 2012). The alignments were automatically deduced and optimized by the Trex Web server, the file for the best tree was then visualized and colored using the Interactive Tree of Life (iTOL) server (Mullins, 2012).

2.4. Clp ATPase homology analysis

The percentage identity between different protein codes within the Clp ATPases were analyzed using Clustal Omega. Full length Clp ATPase sequences were subjected to Clustal analysis to obtain the percentage identity amongst the proteins as identity matrix results. The results were then laid out on an Excel spreadsheet for analysis.

2.5. Homology modelling of ClpK

To model the structure of ClpK (Uniprot: E0W6V3), a template search was done on PYHRE (Kelley *et al.*, 2015), Itasser (Yang *et al.*, 2015b, Roy *et al.*, 2010, Zhang, 2008) and NCBI (Camacho *et al.*, 2009). The template was selected based on high sequence identity, coverage parameters and a high determination resolution. Thus, *Thermus thermophilus* ClpB (1QVR-B) was selected sharing 52% identity and 83% query coverage with the target protein. The modelled ClpK and template ClpB was submitted to ProCheck for stereochemical analysis using Ramachandran analysis (Laskowski *et al.*, 1993). The MolProbity server was used to obtain the Rama Z-score for the trimeric ClpK structure (Williams *et al.*, 2018). The model was then submitted to the Maestro v12.2 molecular modelling algorithm (Schrödinger, 2021). To pre-process the protein structure; bond orders were assigned, hydrogen atoms were added, zero bond orders to metals and disulphide bonds were created, and water molecules that were 5 Å from heteroatoms were deleted. Additionally, the PROPKA algorithm was used at pH 7.0 to optimize hydrogen bonding network by sampling the orientation of water molecules. Lastly, the OPLS_2005 force field was used to refine the structure through minimisation. The stereochemistry of the side chains was checked to ensure that no major perturbations have been induced while preparing the structure. The minimized structure was saved as a Maestro (.mae) file for subsequent prediction analysis.

The ClpK and ClpB sequences were aligned on TCOFFEE to prepare for modelling (Di Tommaso *et al.*, 2011). The position of Walker A and Walker B motifs on the template protein was assigned as shown in Lee *et al.* (2003). The final ClpK structure was then visualized using PyMol, and the root mean square deviation (RMSD) value was obtained (Sievers *et al.,* 2011). The monomeric ClpK and template ClpB structures were processed through ProteinPlus (https://proteins.plus) to identify the position of binding pockets in each protein structure (Schöning-Stierand *et al.*, 2020).

2.6. Molecular Dynamic simulation

Molecular dynamic (MD) simulations were carried out on Maestro v12.2 using the implemented GPU-enabled Desmond molecular dynamics simulation engine. The ClpK trimeric protein or the template ClpB protein were saved as Protein Data Base

(PDB) files and submitted to the Linux (Ubuntu) desktop server for the Desmond MD simulations studies. The ClpK trimeric protein or ClpB template protein were placed in an orthorhombic box (distance from the box face to the outermost protein atom was set 10 Å, the box angle was α = β = γ = 90°). The volume box containing the ClpK trimeric protein or ClpB template protein was minimized, and counter ions were added to neutralize the system. In addition, 0.15 M NaCl was added into the solvent box for physiological conditioning. Thereafter, the system was submitted for the MD simulation. MD simulation is divided into eight distinct stages in which the simulation parameters are specified for each stage. Stages 1-7 is the equilibration, which is made up of short simulation steps, and stage 8 is the final long range 100 ns simulation stage. The type and parameters of the solvated system were detected in stage 1. In stage 2, a 100 ps simulation was carried out using Brownian Dynamics under canonical ensemble (NVT) condition at 10 K with restraints placed on the solute heavy atoms. Stage 3 involved a 12 ps NVT condition at 10 K with restraints on heavy atoms. Stages 4, 6 and 7 (the pocket solvation at stage 5 was skipped) employed short simulation steps (12, 12 and 24 ps, respectively) under isothermal-isobaric ensemble (NPT) conditions (at 10 K and restraints on heavy atoms for stages 4 and 6). No restraints were placed on heavy atoms at stage 7. The final MD production stage at constant temperature of 300 K was carried out at stage 8.

2.7. Post dynamic analysis

Post dynamic analyses of the trajectories derived from the MD simulation studies were carried out using Schrodinger Maestro v12.2 or the Bio3D R-Statistical package for comparative analysis of protein structures. Firstly; Simulation Quality Analysis (implemented in Maestro v12.2) was used to analyse the quality of simulations which analyse the average energy, pressure, temperature and volume. Secondly; Simulation Interaction Diagram algorithm (implemented in Maestro v12.2) was used to analyse the RMSD of the alpha carbon atoms (Cα), the root mean square fluctuation (RMSF) of the residues, and secondary structure element analysis. Lastly; the Simulation Events Analysis algorithm (implemented in Maestro v12.2) was used to calculate the radius of gyration (Rg) and atomic distance.

2.8. Protein Disorder and Circular Dichroism analysis

The primary amino acid sequence of ClpK (Figure A1) was used to predict protein disorder and perform a virtual Circular Dichroism (CD). The sequence was subjected to analysis on the IUPred2A server for protein disorder prediction (Mészáros *et al.*, 2018). The sequence was also subjected to analysis on DichroCalc which is found on The Hirst Group Home Page (https://comp.chem.nottingham.ac.uk/dichrocalc/) server. The CD graph obtained from DichroCalc was then analyzed for secondary structures through the DicroWeb server (Lobley *et al.*, 2002).

2.9. Plasmid construction

The sequence of ClpK (Uniprot: E0W6V3_KLEPN) was sent to GenScript (NJ, USA) for synthesis. At Genscript the gene was optimized for expression in *Escherichia coli* cells, and it was inserted into a pColdI vector using BamH1 and Sal1 to obtain the pColdI-ClpK plasmid construct.

2.10. Competent cell preparation

Vials of non-competent *E. coli* JM109 cells were obtained and grown overnight in 2xYT media containing 16 g tryptone, 10 g yeast extract and 5 g sodium chloride. From the overnight culture, 2 ml was added to 200 ml 2xYT media and incubated at 37°C until an optical density (600 nm) between 0.3 and 0.4 was obtained. The cells were then centrifuged (2700 x g, 4°C, 10 minutes), and the pellet was resuspended in 0.1 M $MgCl_2$ and 0.1 M $CaCl_2$. The cells were centrifuged (2700 x g, 4°C, 10 minutes) again and the pellet was resuspended in 0.1 M $CaCl_2$. An equal amount of 60% glycerol was added to the resuspended cells, the cells were then rapidly frozen in liquid nitrogen and stored at -80°C. Competent *E. coli* BL21 cells were prepared in a similar manner.

2.11. Transformation of recombinant plasmid

The pColdI-ClpK construct was transformed into competent *E. coli* JM109 cells. Briefly, 1 µL of pColdI-ClpK construct was gently mixed with 20 µL competent JM109 cells. The mixture was incubated on ice for 30 minutes, it was then was heat shocked for 90 seconds at 42°C and 80 µl of 2xYT media was added. The cells were then incubated at 37°C (200 rpm) for 1 hour before they were plated onto 2xYT agar plates

containing ampicillin (100 µg/mL) and incubated overnight. Competent *E. coli* BL21 cells were transformed in a similar manner.

2.12. Plasmid Isolation

The recombinant plasmid was isolated using the GeneJET Plasmid Miniprep Kit. Briefly, following transformation, a JM109 colony was resuspended in 2xYT medium containing ampicillin (100 mg/mL) and grown overnight at 37ºC (200 rpm). Glycerol stocks were made by combining equal volumes of the overnight culture and 60% glycerol and stored at -80°C. The remaining culture was centrifuged (6000 x g; 15 minutes, 4°C) and used to isolate recombinant plasmid as per the manufacturer's instructions. The concentration of the isolated plasmid was quantified using a Thermo Scientific NanoDrop 2000 Spectrophotometer.

2.13. Clone confirmation

To confirm that the ClpK gene had been successfully cloned into the plasmid we carried out plasmid confirmation using restriction digestion and colony PCR.

2.13.1. Restriction digest

The isolated plasmid was combined with CutSmart, BamHI, Sal1 (obtained from New England, Bio labs) and Milli-Q water. The mixture was incubated at 37°C for 1 hour, and then 65°C for 20 minutes. The digest was then analysed in 1% agarose gel containing 3 µL ethidium bromide. The gel was viewed under the ultraviolet light using the Syngene GBox.

2.13.2. Colony PCR

JM109 and BL21 colonies grown following transformation were picked using a sterile pipette tip and resuspended in sterile Milli-Q water. A polymerase chain reaction (PCR) master mix containing 1X One*Taq* Standard Reaction buffer, 200 µM dNTPs, 0.2 µM pColdI forward primer (5'ACGCCATATCGCCGAAAGG 3'), 0.2 µM pColdI reverse primer (5' GGCAGGGATCTTAGATTCTG 3'), and One*Taq* Standard polymerase was assembled. The master mix was added to 1 µL of the resuspended colony and the reaction volume was made up to 25 µL using nuclease free water. The 25 µL reaction

was then placed in a Bio-Rad PCR thermocycler under the following conditions; initial denaturation (94°C for 30 sec), denaturation (94°C for 30 sec), annealing (50°C for 30 sec), extension (68°C for 60 sec), and final extension (68°C for 5 min) for 30 cycles. The PCR samples were analysed on a 1% agarose gel with 3 µL of ethidium bromide. The gel was visualized under the ultraviolet light using the Syngene GBox.

2.14. ClpK expression and purification

Transformed *E. coli* BL21 pColdI-ClpK cells were used to inoculate 10 mL Lysogeny Broth (LB) containing 100 mg/mL Ampicillin (10 µL). The cells were grown overnight (16 h) at 37 ˚C. Cells from the overnight flask (1 mL) were then used to inoculate 100 ml LB media containing 100 mg/mL Ampicillin (100 µL). The cells were grown to an OD of 0.4 to 0.6 before they were cold-shocked in ice for 30 minutes. The culture was then induced with 0.25 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and incubated at 15 ˚C for 24 h. Uninduced cells served as the control and were grown along with the induced cells.

Both uninduced and induced cells were harvested by centrifugation at 6800 x g for 15 min. After centrifugation, the resulting pellets were resuspended in binding buffer (20 mM sodium phosphate, 20 mM imidazole, pH 7.4). The cells were then ruptured by sonication (15 min, 30 sec on, 20 sec off, 50% amplification) followed by centrifugation. Both the soluble and insoluble fractions were analyzed using reducing 12.5% sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli, 1970). The protein concentration was determined using Thermofisher NanoDrop 2000 at 280 nm (Kruger, 1994).

Purification was carried out using firstly; a 5 mL HiTrap Q HP anion exchange column (GE Healthcare), and secondly; a 1 mL HisTrap HP column (GE Healthcare). The HiTrap Q HP column was equilibrated with binding buffer. The soluble cell fraction was passed through the anion exchange column. Unbound proteins were removed from the column using binding buffer, and remaining protein was eluted with Buffer B (20 mM imidazole, 20 mM sodium phosphate, 0.2 M NaCl, pH 7.4).

Fractions containing the eluted protein were pooled and passed through the HisTrap HP column which had been equilibrated with Buffer B (20 mM imidazole, 20 mM sodium

phosphate, 0.2 M NaCl, pH 7.4). Unbound proteins were removed from the column using buffer B, and the remaining proteins was eluted with Buffer C (192 mM imidazole, 20 mM sodium phosphate, 0.5 M NaCl, pH 7.4), the elution samples were collected and analyzed using reducing 12.5% SDS-PAGE gels (Laemmli, 1970).

2.15. Enzyme activity assay

ATP hydrolysis was measured using the ATPase/GTPase Activity Assay Kit (Lot# 113BI08A16) from Sigma-Aldrich according to the manufacturers instruction. Briefly; 0 to 0,006 mg/ml ClpK was mixed with 30 µl of reaction buffer (40 mM Tris, 80 mM NaCl, 8 mM $MgAc^2$, 1 mM EDTA, and 4 mM ATP, pH 7.5) and incubated at room temperature for 30 minutes. The ATPase activity was followed by measuring the release of phosphate ions spectrometrically at 620 nm, as a result of the conversion of ATP to ADP. Measurements were performed in triplicate.

2.16. Spectroscopical analysis

2.16.1. Far-UV circular dichroism spectroscopy

Far-UV CD was used to investigate the secondary structure of ClpK (0.5 µM) in 20 mM sodium phosphate and 0.02% sodium azide (pH 7.4). This was performed both in the absence and presence of 0.1 mM ATP. The Jasco J-1500 spectropolarimeter was used to conduct the experiment at 20^0C using a 2 mm quartz cuvette. Spectra measurement was collected in 7 accumulations from 250 to 180 nm at 2.5 nm band width, 0.2 nm data pitch and 1 sec response time. Buffer contributions were subtracted from the final spectrum. The CD spectra was recorded in millidegrees ellipticity (θmdeg) and was converted to mean residue ellipticity [θMRE] (deg.cm2.dmol-1) using the equation below:

$$[\theta_{MRE}] = \frac{100 \times \theta_{mdeg}}{cnl}$$

Where θmdeg is the signal of measured ellipticity in mdeg, c is the protein concentration measured in mM, n is the number amino acid residues of the protein, and l is the pathlength in cm. The data was analysed using the CONTIN parameter on Dichroweb (Lobley *et al.*, 2002).

2.16.2. Extrinsic fluorescence spectroscopy

The binding of ATP-substrate to ClpK was probed using methylanthraniloyl-ATP (mant-ATP). A concentration of 10 µM mant-ATP each was added to the protein (0.5 µM) and was excited at 355 nm, the excitation and emission band width were fixed at 5 nm. The emission was collected between 400 – 600 nm wavelengths in three accumulations. Buffer contributions were subtracted from the final spectrum.

Extrinsic 8-anilino-1-naphthalenesulfonic acid (ANS) fluorescence was carried out to investigate the presence of hydrophobic pockets on the protein. ANS (200 µM) was incubated with the protein in the dark, and the samples were excited at 390 nm wavelength. Emission spectra were collected between 400 – 600 nm wavelengths. The excitation and emission band widths were fixed at 5 nm and 10 nm, respectively using 200 nm/min scanning speed. All fluorescence samples were prepared with 0.5 µM protein concentration in the presence and absence of 2 mM ATP using 10 mM sodium phosphate pH 7.4 and 5 mM MgCl2 for ATP-bound samples. Buffer contributions were subtracted from each final spectrum. The experiments (ANS and mant-ATP fluorescence) were performed using Jasco FP-6300 fluorescence spectrophotometer at 20 °C using a 10 mm quartz cuvette.

2.17. Thermal analysis

Protein thermal analysis was performed in two ways. Firstly; purified protein (0.3 mg/ml) was incubated at temperatures varying from 10 ^{0}C to 100 ^{0}C for 1 h. The protein was then combined with non-reducing treatment buffer and analysed on a SDS-PAGE gel. Secondly; a baseline thermal study (from 20 ^{0}C to 80 ^{0}C) was carried out on ClpK (2 µM) in 20 mM sodium phosphate and 0.02% sodium azide (pH 7.4) using the Jasco J-1500 spectropolarimeter.

Chapter 3: Results

3.1. Caseinolytic ATPase in *Klebsiella* species

The recent discovery of ClpK in the *Klebsiella* species led to the investigation of the presence and diversity of Clp ATPases in *Klebsiella* strains using genome wide analysis. It was observed that 98% of the strains studied contained ClpA with the exception of *K. oxytoca* CAV1335 and *K. variicola* KP5-1 (Figure 5). A similar observation was noted for ClpB where out of the studied species, only two (*K. variicola* KP5-1 and *K. variicola* DX120-E) did not contain ClpB (Figure 5). With respect to ClpX, *K. pneumoniae* CAV1217 was the only species that lacked the ClpX gene (Figure 5). Interestingly, it was further observed that unlike ClpA, ClpB and ClpX, ClpK was only found to be present in 34% of the studied species. Also, only four out of the seven *Klebsiella* species contained ClpK with the highest number of ClpK being found within the *K. pneumoniae* species (Figure 6).

The data obtained from the genome analysis shows a considerably low number of ClpK identified among the *Klebsiella* strains compared to the number of ClpA, ClpB and ClpX proteins. Out of the 57 *K. pneumoniae* strains which were analyzed, only 27 (47%) contained ClpK (Figure 5, Figure 6). Interestingly it was also observed that none of the studied species contained ClpC.

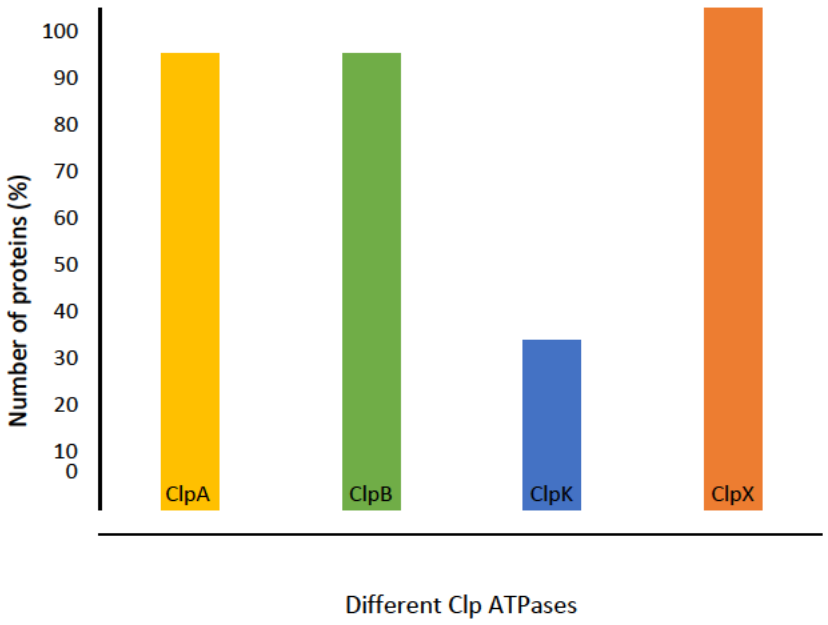


Figure 5: Clp ATPases identified in the investigated *Klebsiella* strains. The total number of Clp ATPases obtained from the National Center for Biotechnology Information (NCBI) Genome database was tallied.

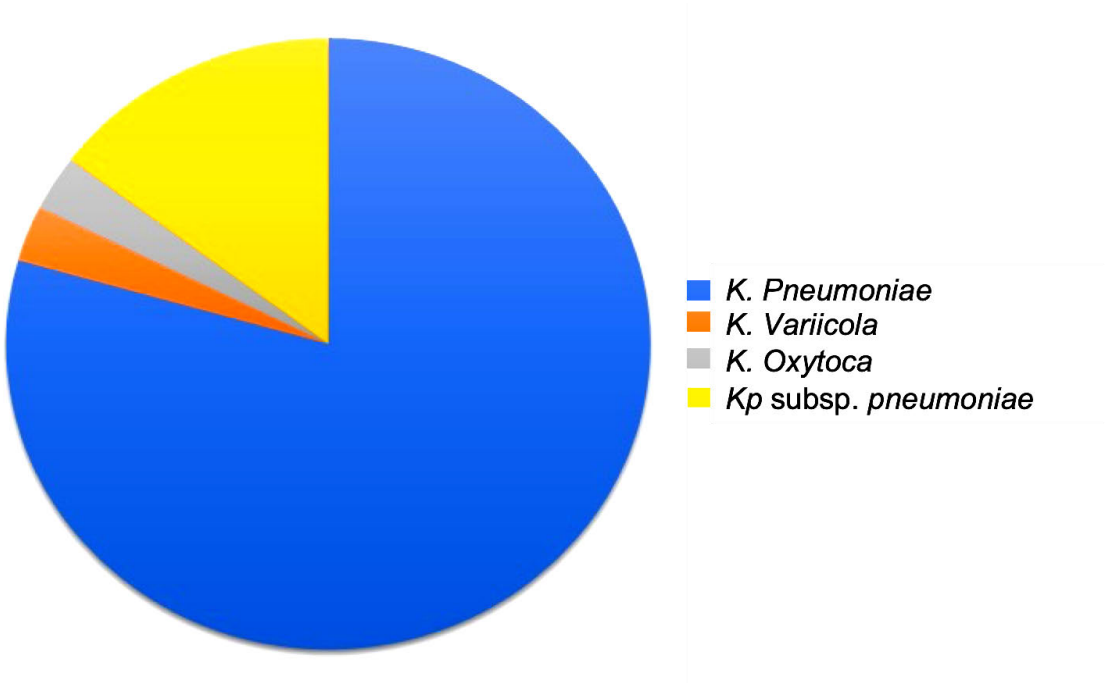


Figure 6: The distribution of ClpK among the investigated *Klebsiella* species. The total number of ClpK was tallied from data obtained using the NCBI Genome database. The highest number of ClpK was found in *K. pneumoniae* (27 strains), followed by *Kp* subsp. *pneumoniae* (5 strains), *K. variicola* (1 strain) and *K. oxytoca* (1 strain).

3.2. Phylogenetic Analysis

A Clustal and phylogenetic tree analysis was performed to establish the relationship between the investigated *Klebsiella* Clp ATPases. Clustal analysis of the studied proteins was performed to evaluate the percentage identity shared within each group (Table 4). The percentage identity gives an estimation of the percentage residues that match up amongst proteins of interest (Newell *et al.*, 2013). Table 4 shows the high percentage identity of the compared protein sequences. This indicates that the proteins identified across the *Klebsiella* species are functionally similar (Hark Gan *et al.*, 2002).

Table 4: Percentage identity of the various Clp ATPases obtained through Clustal analysis.

Clp ATPase	Identity (%)
ClpA	97.1-100
ClpB	96.5-100
ClpX	98.3-100
ClpK	94.1-100

The phylogenetic tree of the investigated Clp ATPases is shown in Figure 7. The root of the phylogenetic tree is in the middle with branches radiating out in different directions. The branch of ClpX irradiates directly from the root. The branches of ClpA, ClpB and ClpK connect before divergence suggesting that the ClpA, ClpB and ClpK protein share a common ancestor. Figure 7 also shows divergence within the arrangement of the ClpK proteins suggesting that changes have occurred at an amino acid level therefore there may be slight functional modifications among the ClpK proteins identified from the various *Klebsiella* strains. Clustal analysis of ClpK proteins namely, *K. pneumoniae* FDAARGOS 566, *K. pneumoniae* KPNIH39, *K. pneumoniae* 2-1, *K. pneumoniae* WCHKP020098, *K. pneumoniae* J1, *K. pneumoniae* FDAARGOS 444, *K. pneumoniae* CAV1417, and *Kp*. subsp. *pneumoniae* KPNIH32 from the various branches, showed an identity of 93.48-100% (Figure 7).

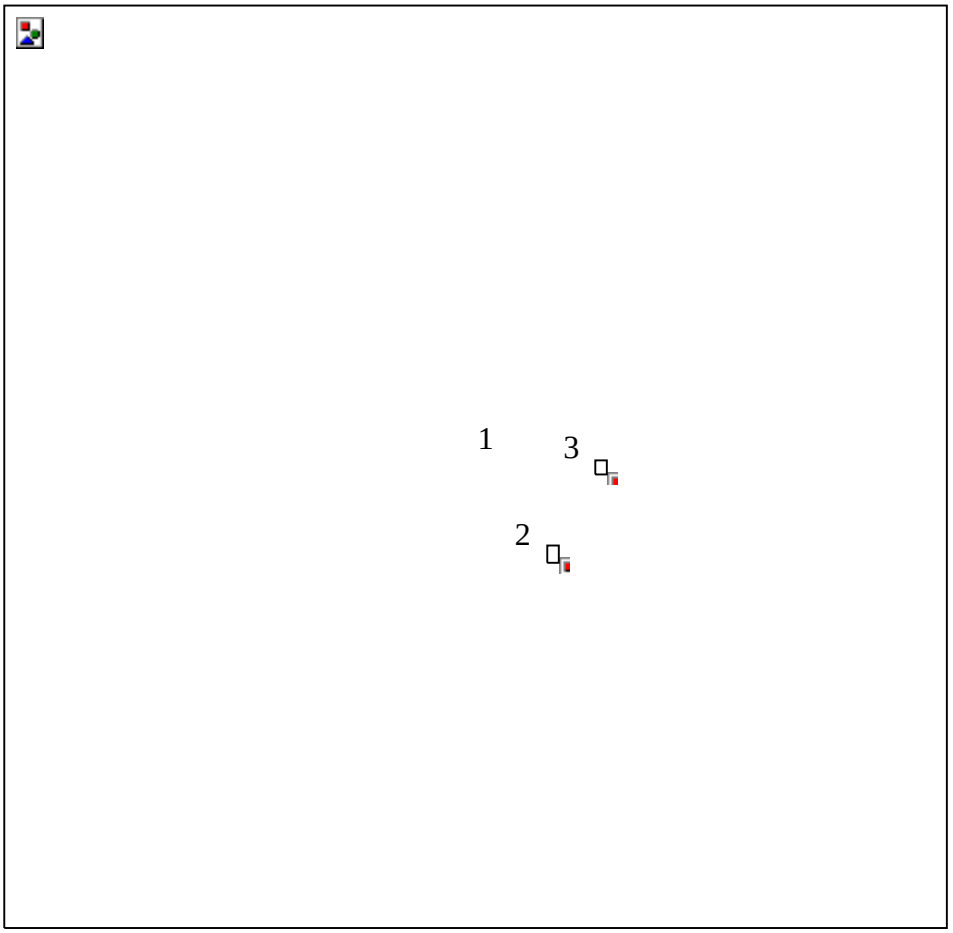


Figure 7: Phylogenetic tree of Clp ATPases found among the seven *Klebsiella* species. MAFFT embedded in Trex servers was used to align the tree. Different Clp ATPases are indicated in different colors. Centre of tree (1), point of ClpK and ClpB divergence (2), ClpA (3). Tree distance scale: 0.1.

3.3. Hypothetical ClpK structure

To successfully model a protein structure, a suitable template for modelling needs to be identified (Schwede *et al.*, 2003, Brenner, 2001). This template should be an experimentally solved three-dimensional structure with more than 40% amino acid sequence identity to the target sequence (Schwede *et al.*, 2003, Brenner, 2001). Subsequently a BLAST search and sequence alignment was performed to identify a suitable template for modelling ClpK. ClpB (1QVR-B) was identified as an appropriate template to model the structure of ClpK since it had a sequence identity and coverage query of 52% and 83%, respectively (Figure 8). The percentage query coverage indicates how much of the query sequence is included in the alignment, the higher the query coverage, the better the match (Newell *et al.*, 2013). Furthermore, the structure of ClpB was determined at 3.00 Å and this was the best resolution compared to the resolution of other structures with similar percentage identity. This resolution is considered to be fairly good as it allows for the visualization of well-defined water molecules and provides a fairly good idea about the shape of the macromolecule (Wlodawer *et al.*, 2008).

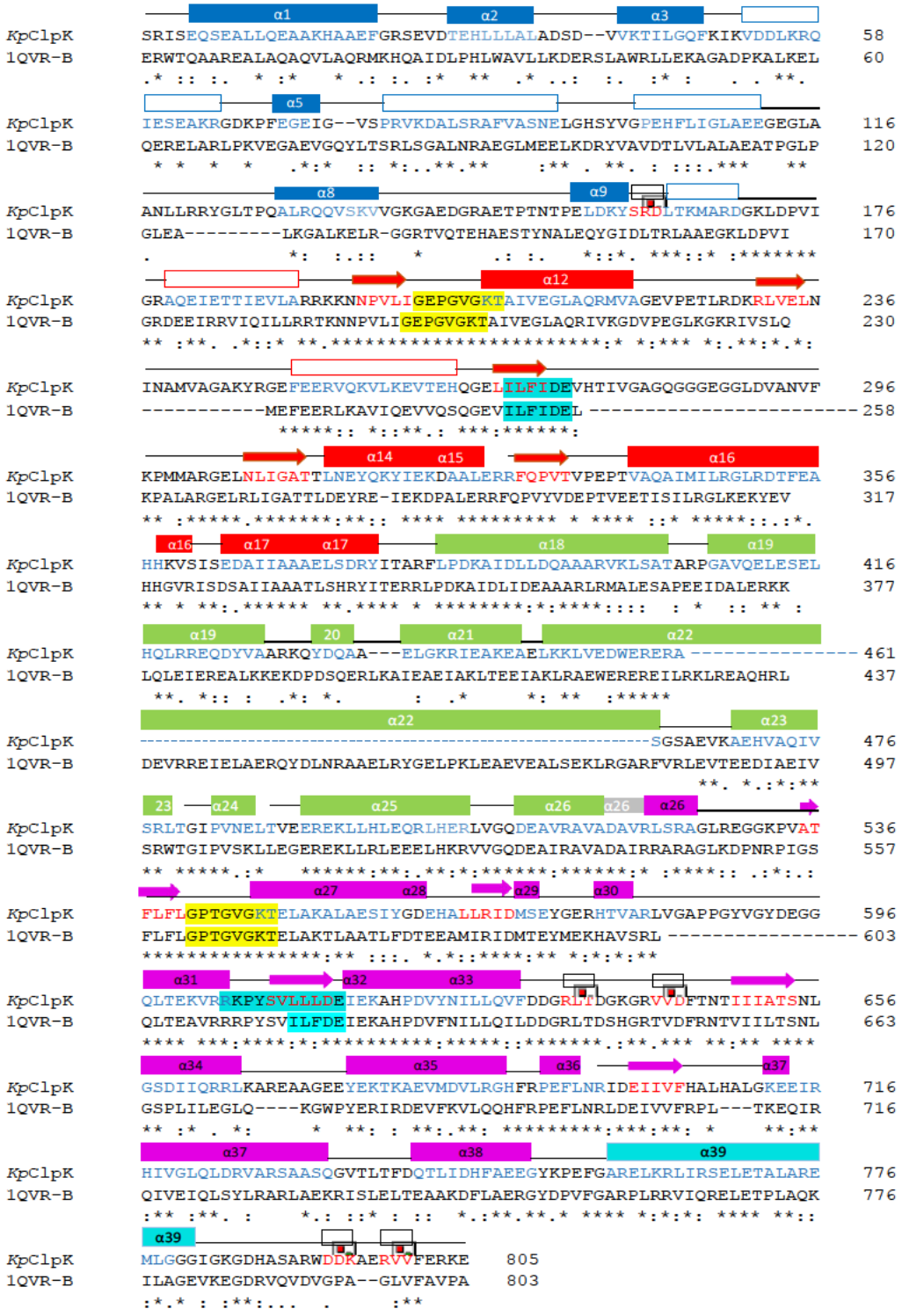


Figure 8: Alignment of ClpK with ClpB. The annotations were done according to sequence ClpB (1QVR-B). The α-helices and β-sheet are shown as rectangles and arrows, respectively. ClpK residues colored blue represent the position of the α-helices whereas red residues represent the position of the β-sheet. Asterisks (*) denote the identical residues, (:) represents sequence homologies, (.) represents weak similarity. The five domains are shown as follows: N-terminal domain (blue); D1 large domain (red); D1 small domain (green); short linker region (grey); D2 large domain (pink); D2 small domain/C-terminal domain (cyan). The secondary structure elements are colored according to the domain. Walker A and Walker B motifs are highlighted yellow and cyan respectively. The alignment was performed using T-COFFEE (Di Tommaso *et al.*, 2011).

The alignment of ClpB and ClpK shows the five domains namely; the N-terminal domain, D1-large domain, D1-small domain, D2-large domain and D2-small domain which are conserved within Class I proteins (Figure 8). The Walker A motif (GXXXGK[T/S]-X represents any residue) is 100% identical in the aligned sequences. The Walker B motif (hhhhD[D/E]-h represents hydrophobic residues) is 100% identical in NBD1 of the aligned sequence, while it only has 73% identity in NBD2 (Figure 8).

The modelled ClpK structure was validated using the ProCheck server and is shown in Figure 9A. ATPases can exist in a monomeric, dimeric and trimeric state in the absence of nucleotides, therefore the hypothesized trimeric structure of ClpK is not alarming (Zheng *et al.*, 2002). Ramachandran analysis of the trimeric ClpK (90.10%) and template ClpB (83.10%) proteins showed that the majority of their protein residues lie within the most favoured regions (Figure 10, Table 5). A Rama Z-score of -0.75 ± 0.16 was obtained for the trimeric ClpK structure from the MolProbity server, this value was within the accepted Z-score range. The structural alignment between the monomeric ClpK and ClpB (Figure 9B) gave a root mean square deviation (RMSD) value of 0.300 Å which is indicative of the two structures adopting a similar conformation (Carugo and Pongor, 2001). The modelled structure of ClpK is consistent with other known structures of ATPases which contain a mixture of mainly α-helices and some β-sheets (Figure 8, 9). Furthermore, this modelled structure agrees with the virtual CD data obtained for ClpK using the DichroCalc server and DicroWeb analysis (Figure 11, Table 6). The spectrum shows that ClpK displays one ellipticity maxima at about 190 nm, and one ellipticity minima at about 220 nm (Figure 11).

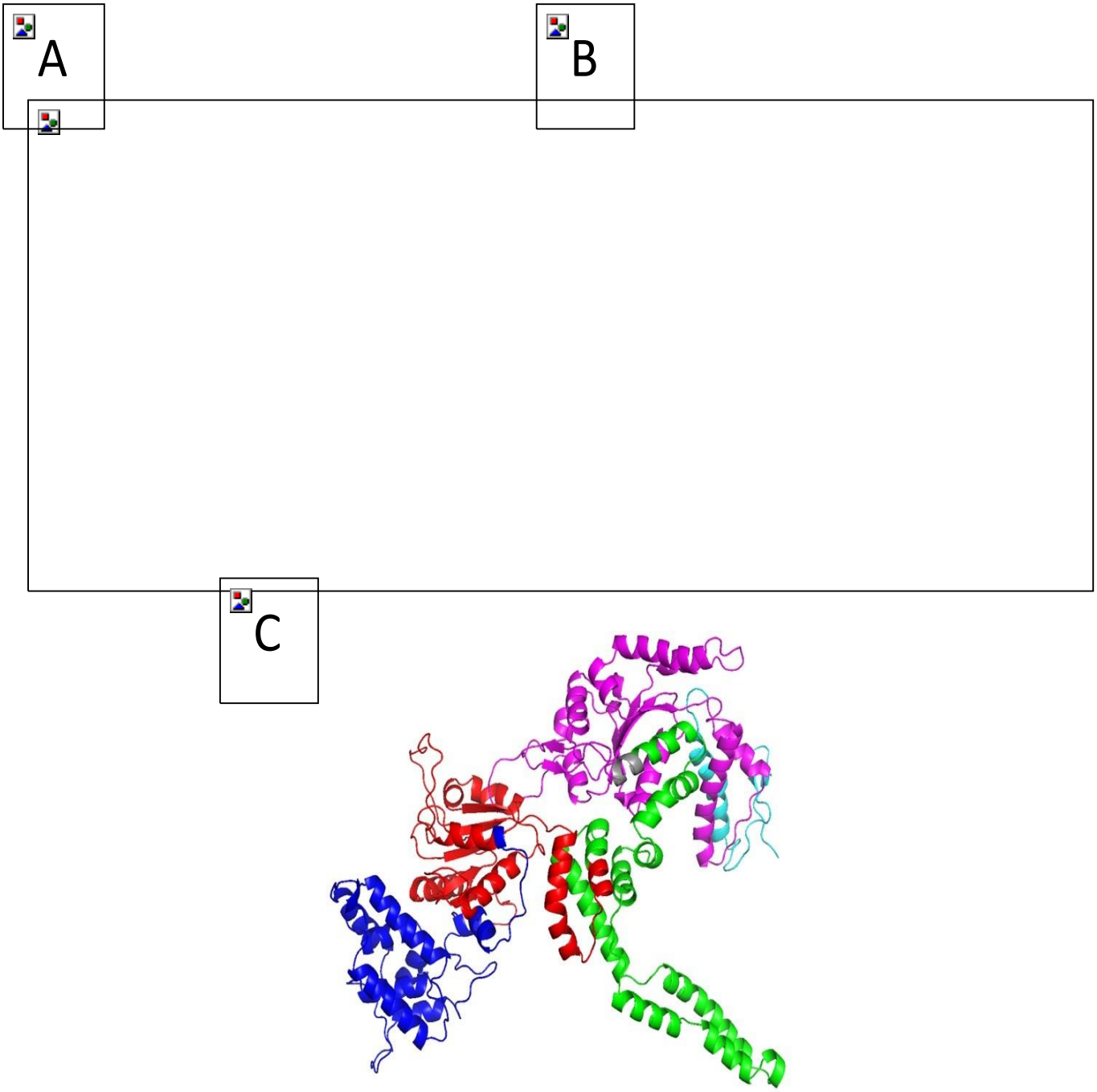


Figure 9: Hypothetical ClpK structure modelled using ClpB as a template. (A): Secondary trimeric structure of ClpK with chains A, B, and C shown as green, cyan, and pink, respectively. (**B):** Superimposed monomeric secondary structure of ClpK (red) and ClpB (1QVR-B) template (blue). **C:** Secondary structure elements of ClpK domains colored as follows: Blue; N-terminal domain, red; D1 large domain, green; D1 small domain, grey; short linker region, pink; D2 large domain, light blue; D2 small domain/C-terminal domain. The structures were refined using OPLS_2005 force field and visualized using PyMOL (Ko *et al.*, 2012, Schrodinger, Version 2010).

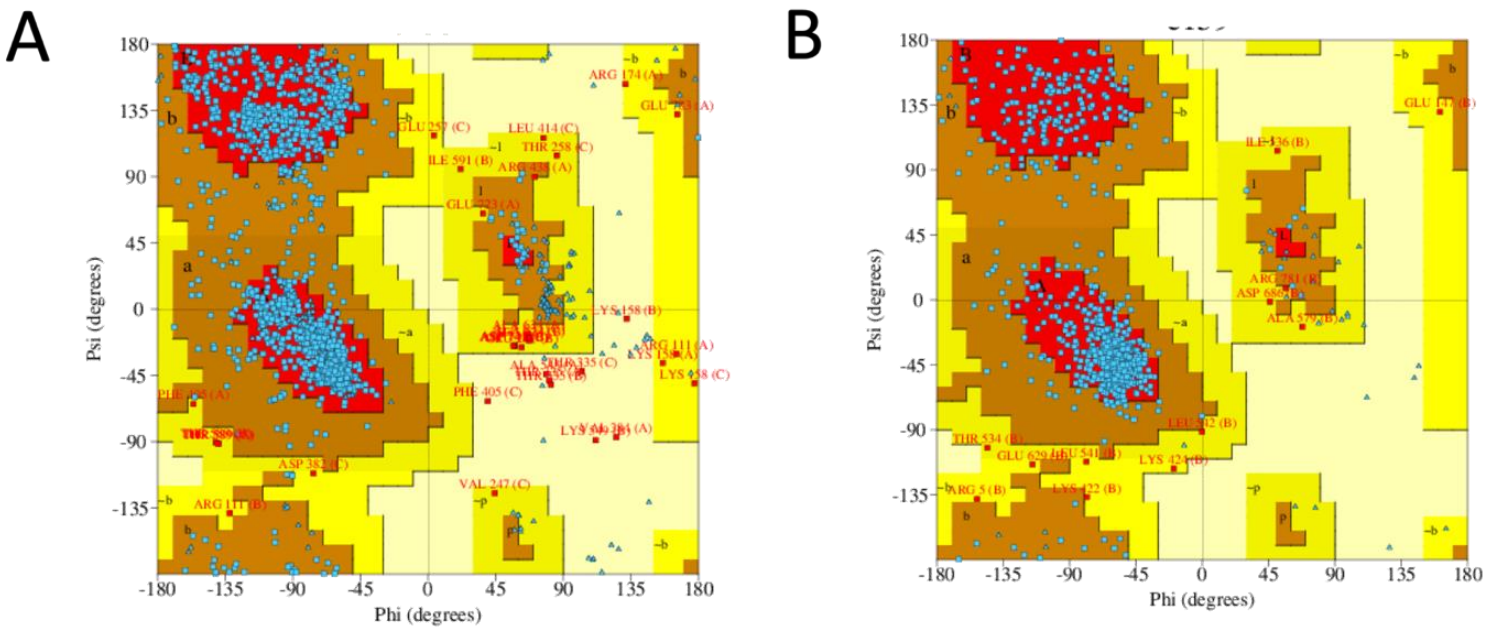


Figure 10: Ramachandran plots of the modelled ClpK protein and the template ClpB protein. (A) ClpK trimeric protein. (**B)** ClpB template protein. The blue dots represent the residues of each protein. The regions are denoted as follows; the most favored regions (red), additional allowed regions (brown), the generously allowed regions (yellow) and the disallowed regions (light yellow). The figures were obtained from the ProCheck server.

Table 5: Ramachandran analysis of the template ClpB and modelled trimeric ClpK protein obtained from ProCheck.

Residues in	Ramachandran statistics	
	1QVR	KpClpK Model
Most favoured regions	83.10%	90.10%
Additional allowed regions	15.20%	8.30%
Generously allowed regions	1.70%	1.20%
Disallowed regions	0.00%	0.40%
Total residues	717	2100

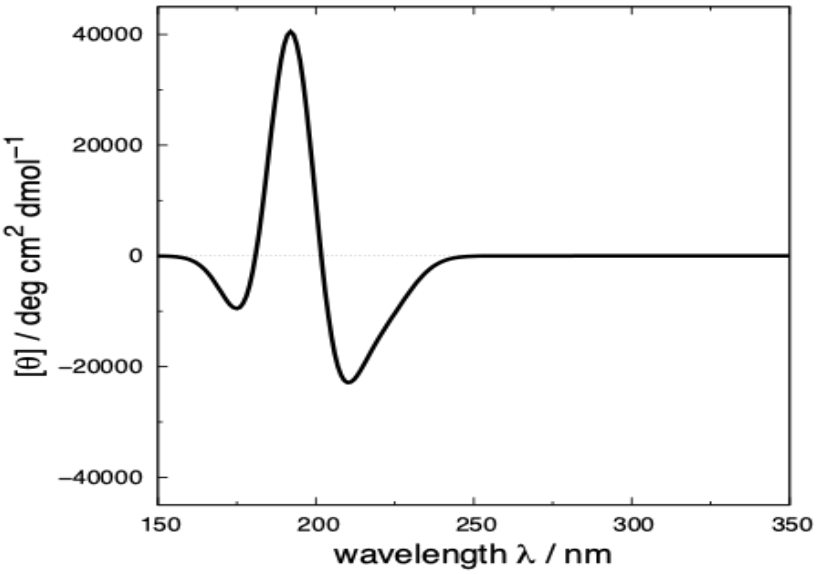


Figure 11: Virtual Circular Dichroism (CD) spectrum of ClpK. The primary amino acid sequence of ClpK was passed through the DichroCalc server, the data obtained from the server was then analyzed to understand the secondary structures of ClpK (https://comp.chem.nottingham.ac.uk/dichrocalc/).

Table 6: DichroWeb[1] analysis of the secondary structure of ClpK.

	Helix (%)	Strand (%)	Turns (%)	Unordered (%)	NRMSD[3]
ClpK (952[2])	35.3	22.9	20.7	21	0.12

[1](Lobley *et al.*, 2002)
[2]Number of amino acids
[3]Normalised root mean square deviation (NRMSD) value is independent of dimensions, and a value of <0.1 is considered acceptable.

Structural analysis of ClpB showed that the N-terminal domain contains a substrate binding groove which is known to recognize hydrophobic residues of unfolded or aggregated proteins (Rosenzweig *et al.*, 2015). Given that both ClpB and ClpK belong to Clp ATPases class I and could potentially function in a similar manner to recognize substrates; the presence of substrate binding grooves was investigated in ClpK. Interestingly, a similar substrate groove was identified in the ClpK N-terminal domain using the DoGSiteScorer (Figure 12, Table 7). However, the volume of the substrate binding grooves of ClpK and ClpB differ, this could suggest that these grooves bind different substrates. Interestingly, the frequent amino acid in the substrate binding groove of ClpK is glutamine and ClpB is leucine (Table 7). This once again indicates that the substrates binding to the protein grooves may differ.

Table 7: Properties of the binding pockets in the N-terminal region of ClpB and ClpK identified using DoGSiteScorer[1].

	ClpB	ClpK
Volume (Å^3)	1767.82	1511.88
Surface (Å^2)	1967.01	1626.42
Depth (Å)	31.30	20.80
Hydrophobicity ratio	0.34	0.28
Enclosure	0.08	0.07
Frequent amino acid	Leucine	Glutamine

[1](Schöning-Stierand *et al.*, 2020)

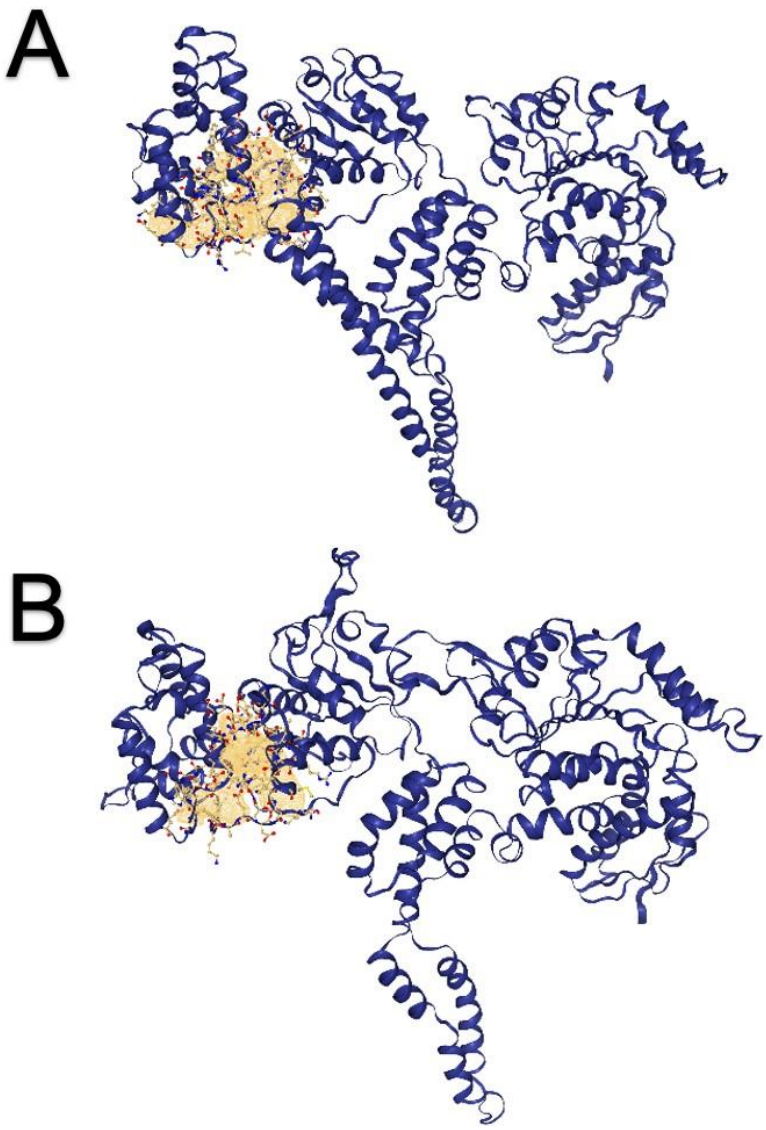


Figure 12: N-terminal binding groove identified in ClpK and ClpB. A) ClpB with the binding groove shown in yellow. **B)** ClpK with the binding groove identified in yellow. The binding grooves were identified using DoGSiteScorer (Schöning-Stierand *et al.*, 2020).

Further analysis of the ClpK N-terminal domain showed that NBD1 and NBD2 of ClpK adopts a RecA-like fold characterized by a central β-sheet flanked by α-helices (Figure 13). This fold is a common structural feature found in most ATPases and assists with the movement of polypeptides into the proteolytic core which is a critical step in protein proteolysis (Ye *et al.*, 2004, Lee *et al.*, 2003). In most Clp ATPases, NBD1 and NBD2 are not separated however in ClpK it was observed that NBD1 and NBD2 are separated by a short linker sequence which adopts a helical structure (Figure 8, 9B). In ClpB, this linker region is termed as "the middle domain" (Lee *et al.*, 2003). It was observed that the linker region of ClpK and ClpB differs in amino acid length, with the linker region of ClpB being almost double the size of ClpK. Although the function of the linker region of ClpK is yet to be investigated, the middle domain of ClpB plays a role in protein stability and interdomain communication between NBD1 and NBD2 (Kedzierska *et al.*, 2003, Cashikar *et al.*, 2002).

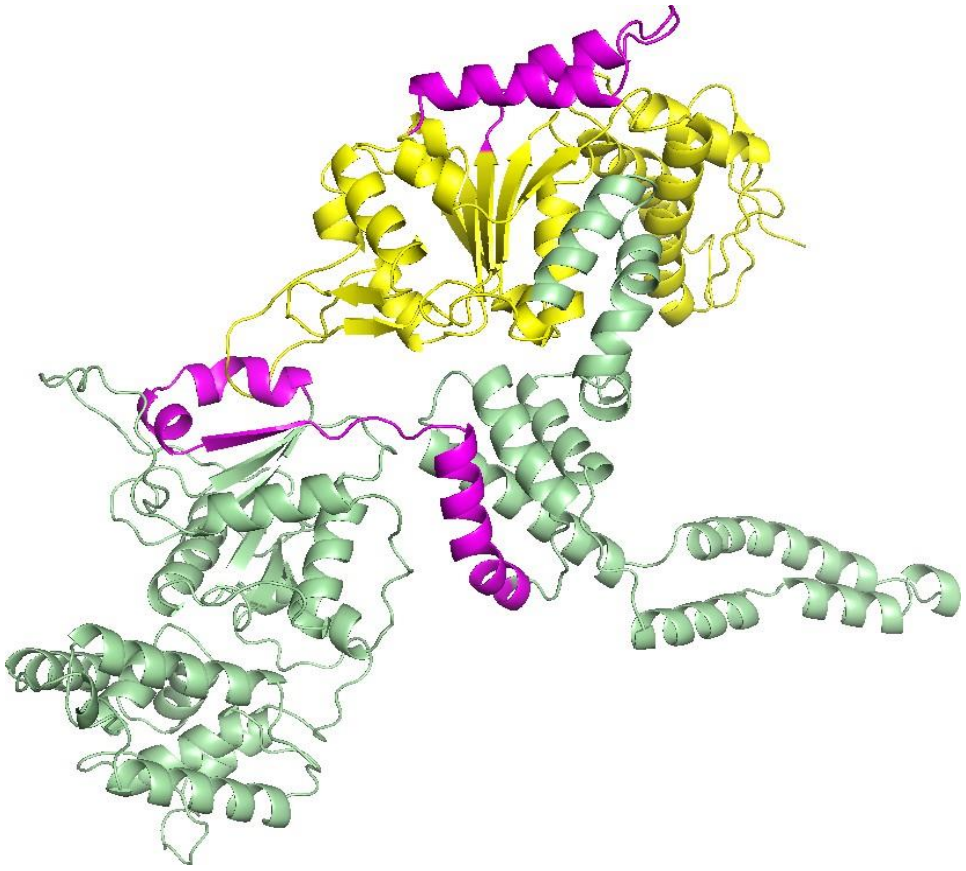


Figure 13: NBDI and NBD2 of ClpK adopts a RecA-like fold. The secondary structural features are coloured as follows; the N-terminal domain and the C-terminal domain are shown as pale green and yellow, respectively. The RecA-like domains are shown in magenta. The protein was visualised using PyMol (Schrodinger, Version 2010).

3.4. Molecular Dynamic Simulation

To confirm that the homology modelling was carried out successfully, the dynamic behavior and stability of the modelled structure was investigated. This was done by performing molecular dynamic (MD) simulations and post-dynamic analyses on the modelled ClpK and ClpB structure. ClpB was used as a control for all the MD simulations and post-dynamic analyses. Figure 14 shows the potential energy profiles to compare the trajectory of the alpha carbons (Cα) within a time frame for ClpK and ClpB. A comparison of the potential energy obtained for ClpK (-401975,3±313,1 kcal/mol) and ClpB (-468132,8±331,6 kcal/mol) shows a slight, insignificant shift which suggests that both the template and modelled protein are relatively stable (Figure 14).

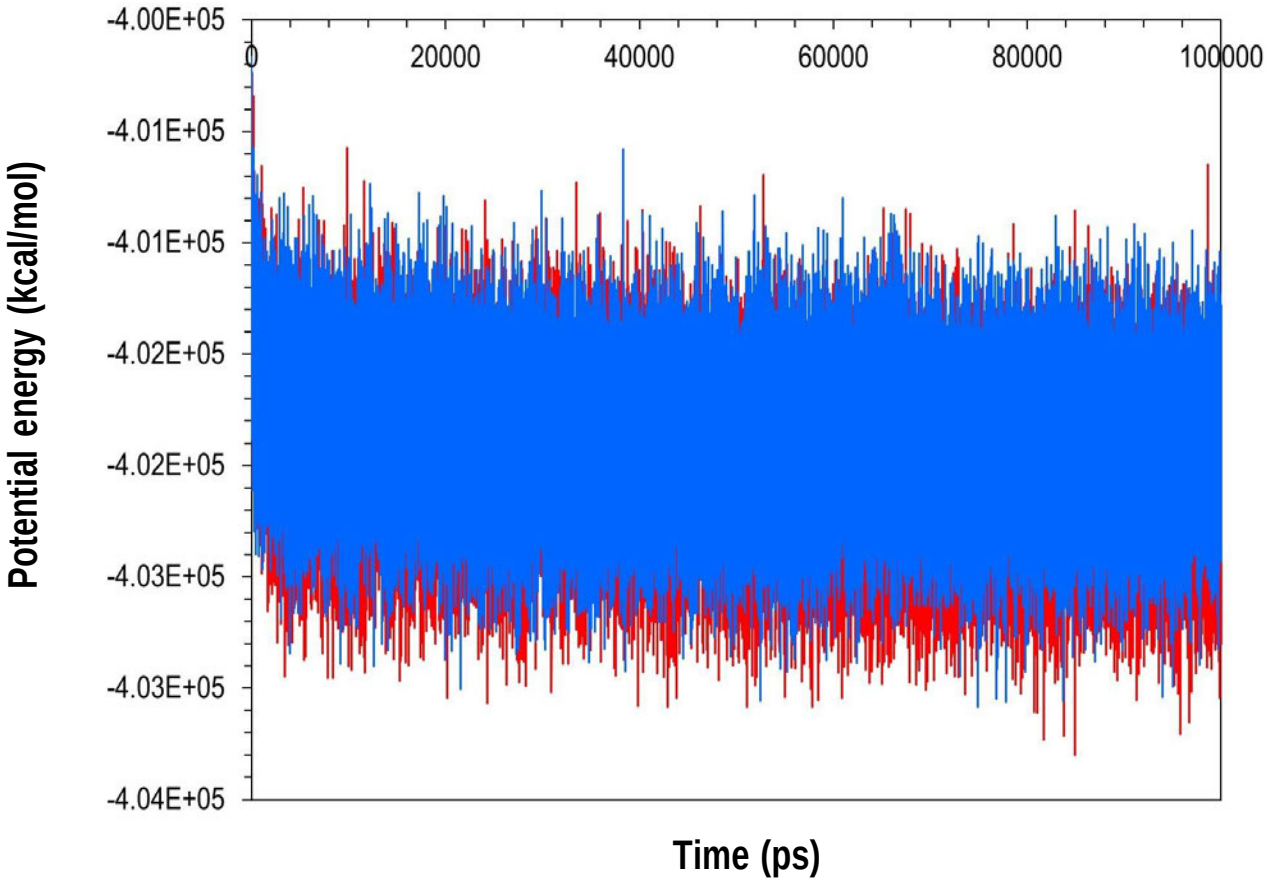


Figure 14: Potential energy profile of ClpK and ClpB during 100 000 ps molecular dynamic simulation. The potential energy for ClpK and ClpB is shown in red and blue, respectively. The image was generated using GraphPad prism.

To assess the dynamic nature of ClpK the root mean squared deviation (RMSD) and the root mean square fluctuation (RMSF) values were calculated. Figure 15A shows the RMSD values of ClpK (7.22±1.52 Å) which increases from 2 Å to 9 Å over 100 ns. A similar increase in RMSD values is observed for ClpB (8.68±2.37 Å) which increases from 2 Å to 11 Å over 100 ns. The increasing RMSD values observed indicate significant conformational changes, which are shown in Figures 15B and 15C for ClpB and ClpK, respectively. Furthermore, the RMSF values were used to identify amino acids which contribute to protein flexibility, the higher the RMSF value the greater the flexibility (Zhao *et al.*, 2015). The D1 small domain and linker region were identified as regions which contribute to the flexibility of ClpK (3.17±1.73 Å) and ClpB (3.27±2.27 Å) (Figure 8, Figure 16).

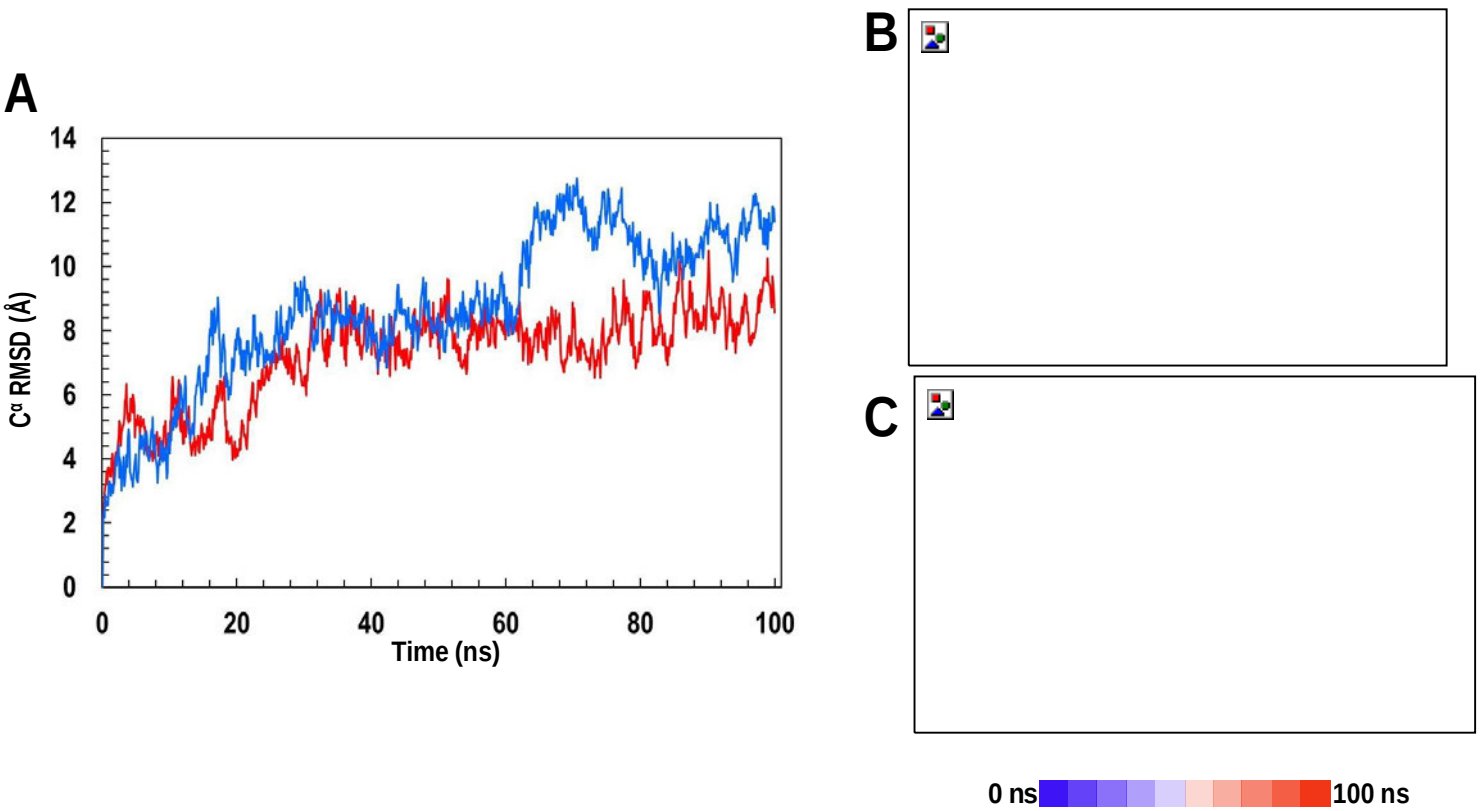


Figure 15: Trajectory analysis showing the RMSD values and the conformational changes of ClpK and ClpB over 100 ns. A) RMSD values of ClpK and ClpB over 100 ns. The trajectory of ClpK is indicated in red and the trajectory of ClpB is indicated in blue. **B)** Conformational changes of ClpB observed over 100 ns. **C)** Conformational changes of ClpK observed over 100 ns.

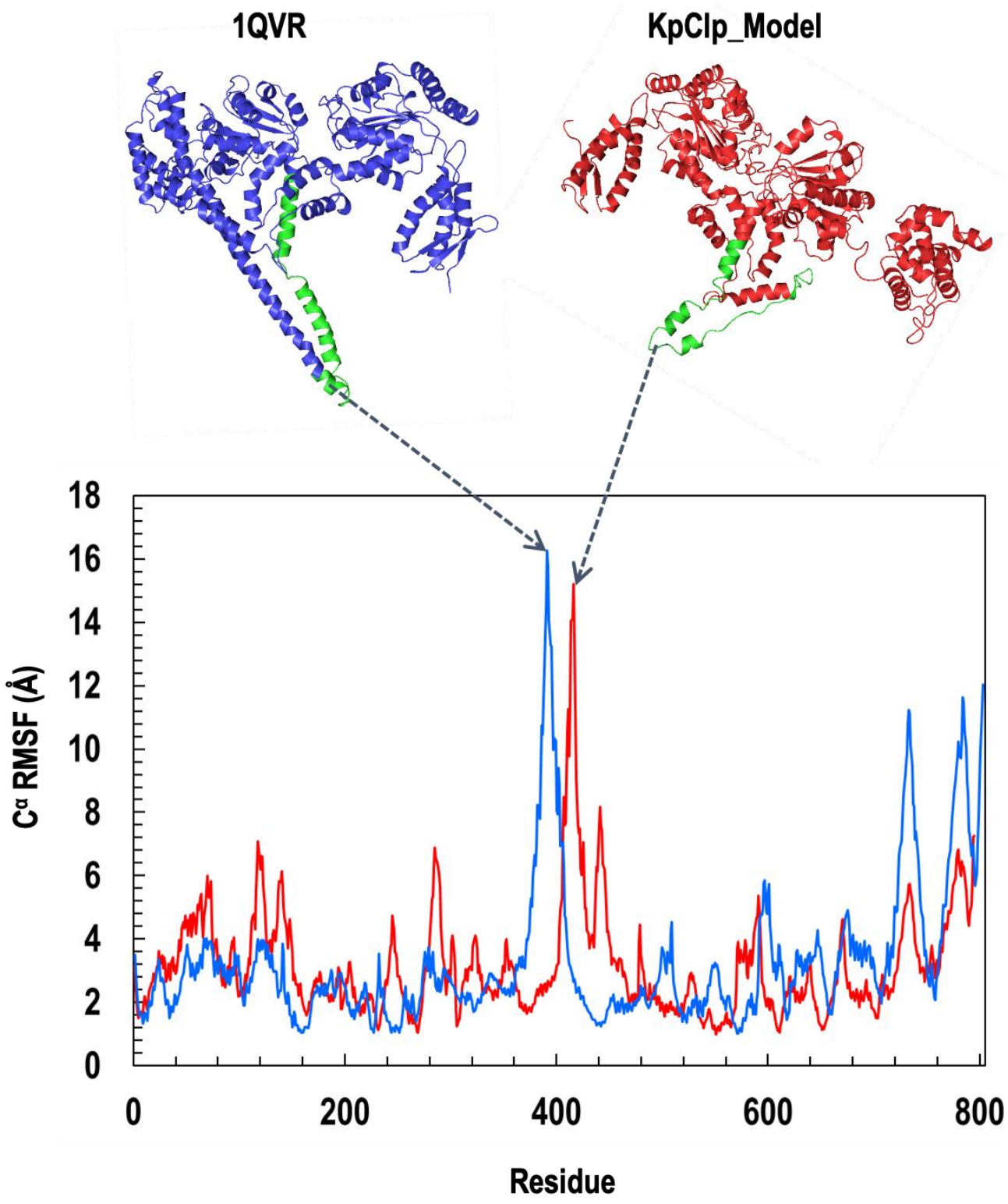


Figure 16: The RMSF of the ClpK and ClpB residues as a function of the 800 ns simulation time. ClpK is represented as red and ClpB is represented as blue on the graph and structures. The position of the peaks are represented on the ClpK and ClpB structures in green. The highest peak for ClpK is seen around residues 410 to 425 while the highest peak for ClpB is seen around residues 379 to 407. To assess the regions of flexibility, we add 110 and 4 to the region values obtained for ClpK and ClpB, respectively, as the structures have been modelled from residue 110 and 4. The graph was generated using Excel. The 3D protein structures were generated using PyMOL.

Protein structure compactness refers to how secondary structures are packed into the tertiary structure (Lobanov *et al.*, 2008). The radius of gyration (Rg) represents the compactness of the protein structure (Pathak *et al.*, 2018). The values obtained for ClpK (40.4±3.93 Å) and ClpB (42,37±4.71 Å) indicate that the proteins do not differ much in terms of structure compactness (Figure 17). Additionally, ClpB seems to undergo transformational change at around 10000 ps, while ClpK only undergoes transformation around 20000 ps (Figure 17). The high Rg values of both ClpB and ClpK confirm the mainly α-helical structure of the proteins as α-helical proteins have the highest Rg profile compared to proteins composed of mainly β-sheets (Lobanov *et al.*, 2008).

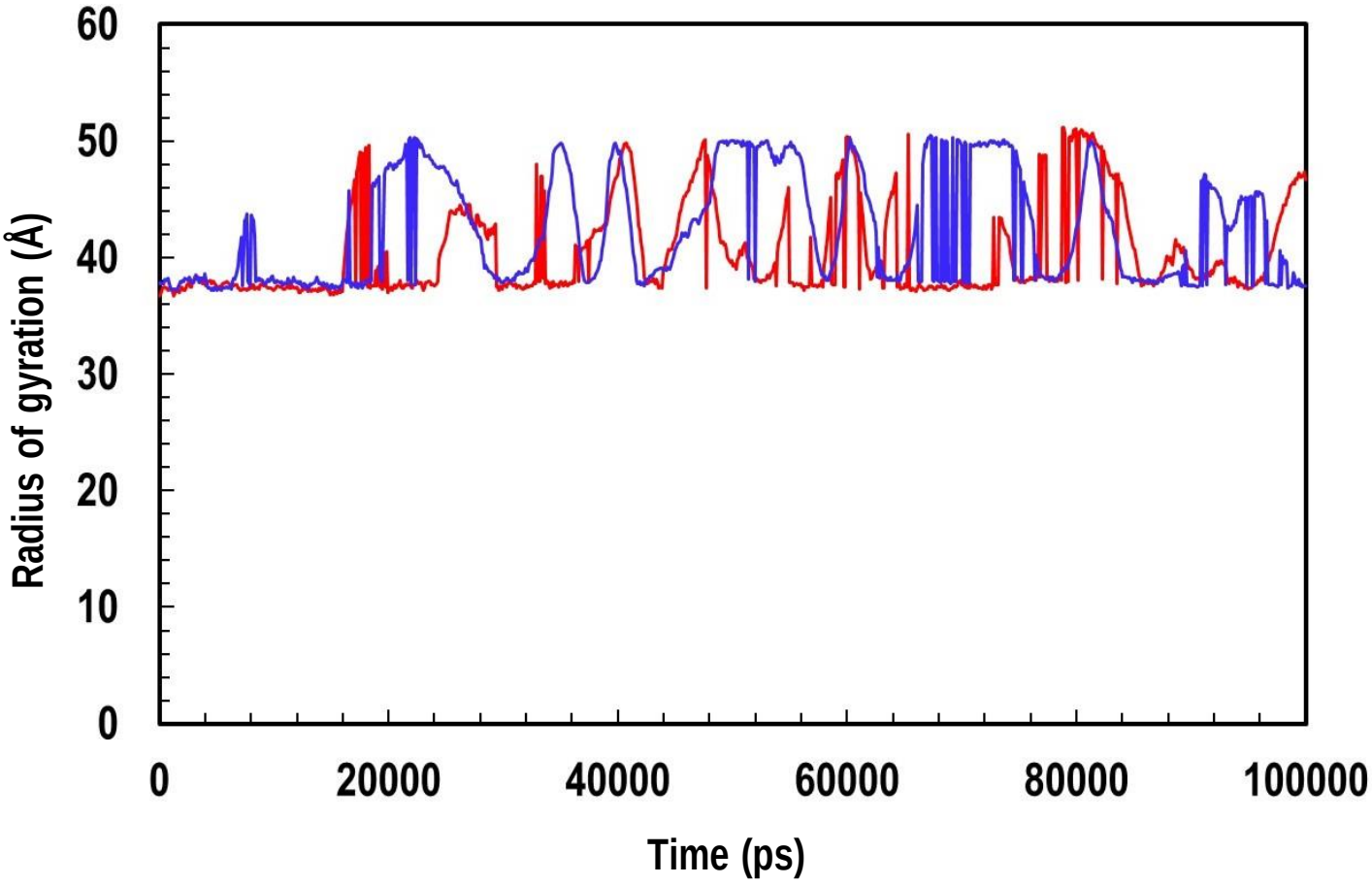


Figure 17: Trajectory analysis showing the Radius of gyration of the alpha carbons of ClpK over 100 000 ps. ClpK is represented as red and ClpB is represented as blue.

3.5. Protein Disorder Prediction

Following homology modelling and MD simulation, the structure of ClpK was assessed for protein and binding disorder. Disordered protein regions do not adopt a stable confirmation and therefore make protein purification, protein-ligand binding studies and crystallization difficult (Babu, 2016). Therefore, assessing ClpK for disorder regions allows one to determine if it is possible to express and purify soluble ClpK as an initial step to protein characterization and protein-drug interaction studies.

Figure 18 shows that less than 40% of the ClpK residues were predicted to be disordered through IUPred2A (red line), suggesting that ClpK can be expressed and purified (Deng *et al.*, 2012). IUPred2A predicts some of the disordered protein residues to be situated in the N-terminal domain (3 small peaks), while most of the disordered protein residues are seen in the C-terminal domain (Figure 18). It has been noted, that proteins with disordered regions carry out important functional roles such as phosphorylation, regulation and protein-DNA binding (Deng *et al.*, 2012). Using Anchor2 server it was observed that a majority of the disordered binding regions were situated in the C-terminal domain and role of C-terminal domain in ClpK is yet to studied (Figure 18).

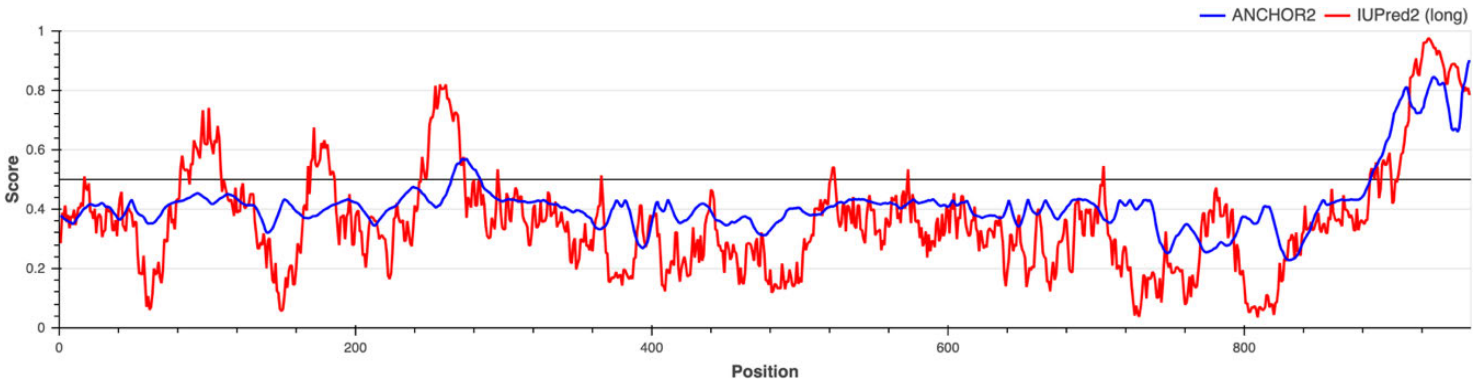


Figure 18: Protein disorder prediction for ClpK. The ClpK protein sequence was analyzed using the IUPred2A server for presence of ordered and/or disordered regions. The black line represents the threshold; the red line represents the protein disorder prediction (IUPred2A), and the blue line represents the binding disorder prediction (Anchor2).

3.6. Clone confirmation

Following *in silico* analysis, *in vitro* analyses were carried out to further investigate the ClpK protein. The expression, purification and characterisation of ClpK is yet to be reported. Subsequently, here the ClpK gene was cloned into a pCold1 vector by GeneScript for expression studies. Prior to expression studies, the ClpK construct was confirmed using restriction digest and colony PCR to ensure that the correct construct had been recieved.

3.6.1. Restriction digestion

The recombinant plasmid (pColdI-ClpK) was transformed into *Escherichia coli* JM109 cells. The construct size was confirmed by digesting the plasmid with BamH1 or SalI (Figure 19). Additionally, the plasmid was double digested with SalI and BamH1 to release the *clpk* gene from the vectors to confirm gene size. Figure 19 shows two bands in Lane 2, one appears below 3000 bp and corresponds to the size of the *clpk* gene and the other band at around 4000 bp corresponds to the size of the vector. Lane 3 and 4 show single bands around 7000 bp, this corresponds to the size of the plasmid with the ClpK insert.

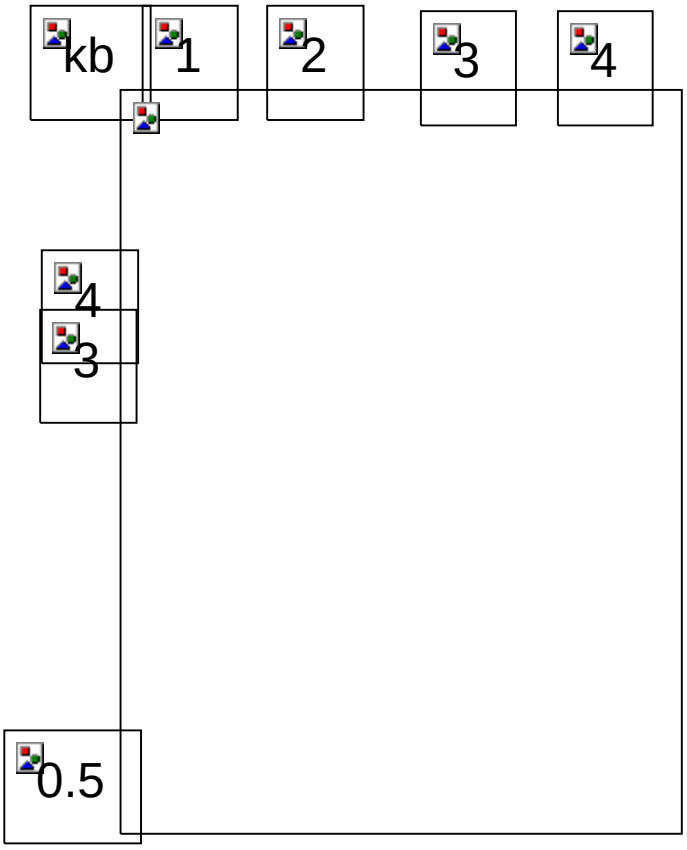


Figure 19: Restriction digest of ClpK construct. The plasmid was incubated in the presence of BamHI and/or SalI at 37^{o}C for 1 h before the reaction was stopped by incubating the samples at 65^{o}C for 20 minutes. Lane 1: Molecular weight marker (MWM), Lane 2: double digest in the presence of BamHI and SalI, Lane 3: single digest in the presence of BamHI, Lane 4: single digest in the presence of SalI.

3.6.2. Colony PCR

Following successful confirmation of the, the ClpK construct, it was transformed into *E. coli* BL21 cells for protein expression. The resulting transformants, were screened for the presence of ClpK plasmid using colony PCR. Figure 20 shows the amplification of the ClpK gene at around 3000 bp in *E. coli* BL21 cells. The observed band is slightly higher than, the expected size of the ClpK gene which is approximately 2859 bp. For this PCR, pCold1 forward and reverse primers which bind to the region flanking the multiple cloning site were used instead of gene specific primers (Figure 21). Therefore, the slight increase in the observed band is due to the amplification of the region flanking the ClpK insert which is about 190 bp.

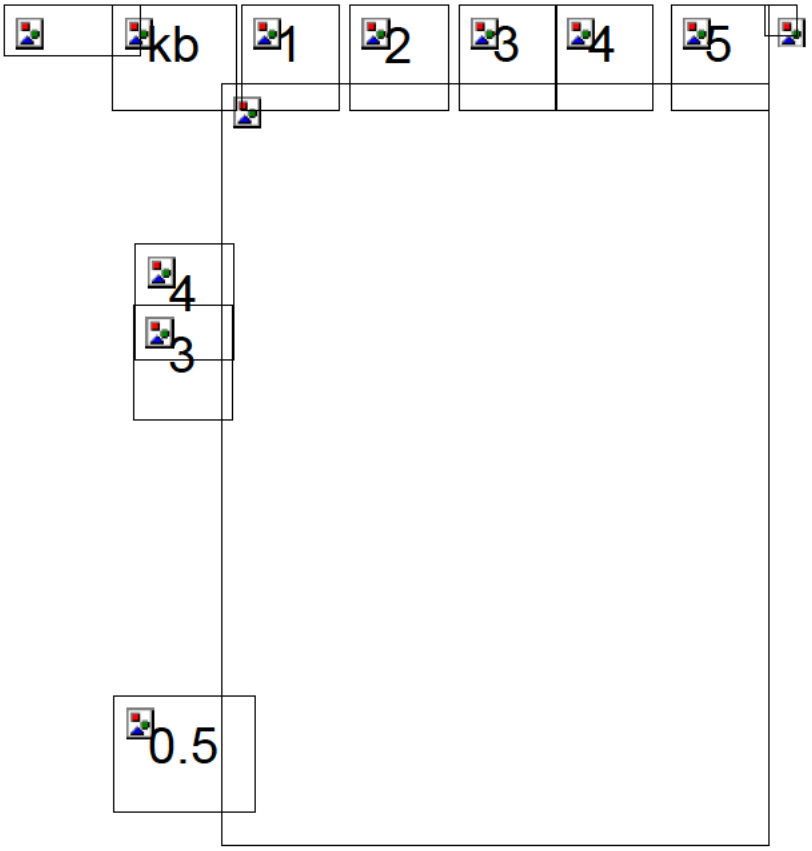


Figure 20: Colony PCR of *E. coli* BL21 cells transformed with the ClpK-pColdI recombinant plasmid. (A) Lane 1: MWM, Lane 2-4: DNA from single colonies of transformed BL21 cells, Lane 5: negative control. PCR products were run on a 1% agarose gel containing ethidium bromide.

Figure 21: Graphical representation of the forward and reverse primers of pColdI flanking the ClpK gene. MCS: multiple cloning site, bp: base pairs.

3.7. Expression and Purification of ClpK

Expression and purification studies of ClpK have not been reported to date. Therefore, to obtain optimum conditions for the expression of ClpK, *E. coli* BL21 cells were transformed with pCold-I plasmid containing the *clpk* gene (ClpK construct). Different expression conditions were tested to determine suitable conditions to express the soluble ClpK protein. Figure 22 shows the successful induction and expression of soluble ClpK using 0.1 mM, 0.25 mM, and 0.5 mM Isopropyl β- D-1-thiogalactopyranoside (IPTG) respectively, as indicated by a protein band corresponding to the molecular weight of ClpK. Based on the band intensity of expressed ClpK, 0.25 mM IPTG was selected as an optimal concentration for expression (Figure 22, Lane 6).

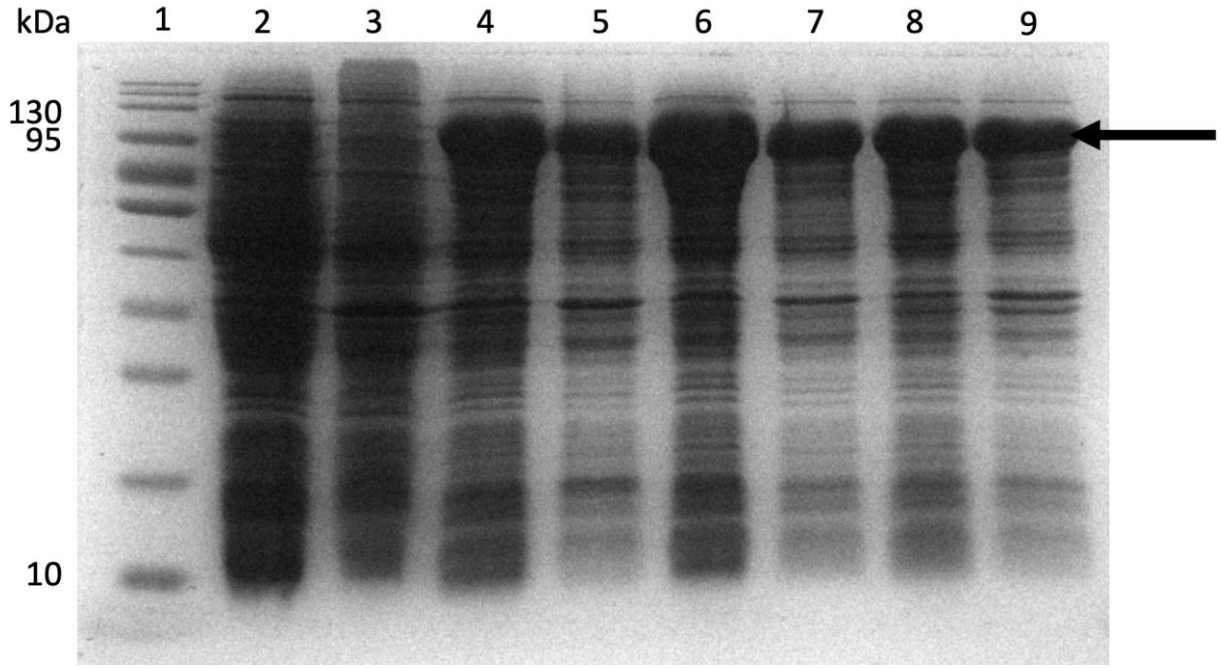
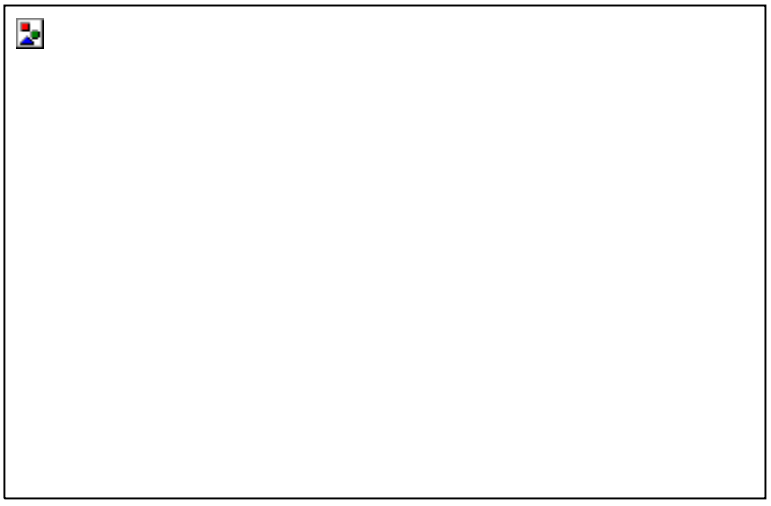


Figure 22: ClpK expression using varying IPTG concentrations. Protein expression was induced with varying IPTG concentrations (0.1 mM, 0.25 mM and 0.5 mM) at 15 ^{0}C for 24 h. Following expression, the culture was harvested by centrifugation. The resulting pellet was resuspended in lysis buffer, sonicated and analysed for soluble and non-soluble protein on a SDS-PAGE gel. Lane 1: Molecular weight marker, Lane 2 and Lane 3: uninduced supernatant (Sn) and uninduced pellet, respectively. Lane 4: 0.1 mM Sn, Lane 5: 0.1 mM pellet, Lane 6: 0.25 mM Sn, Lane 7: 0.25 mM pellet, Lane 8: 0.5 mM Sn, Lane 9: 0.5 mM pellet. The arrow indicates the position of the expected protein.

Following expression, ClpK was purified using ion exchange and affinity chromatography. Initially, the expressed protein was subjected to ion exchange chromatography which is based on the electrostatic interaction between the resin and the protein (Coşkun, 2016). At pH 7.4, ClpK (pI: 5.61) is negatively charged, thereforeit binds to the positively charged anion exchange resin. The protein bound to the resin was eluted with increasing concentration of sodium chloride in Buffer B (Figure 23a). ClpK co-elutes with impurities even at the highest salt concentration, therefore a second purification step was performed. In addition, the recombinant protein only reached 2.7% purity after ion exchange chromatography (Table 8). The eluent from anion exchange was passed through an affinity chromatography column. ClpK was expected to bind to the affinity column resin since it contains a HisTag while the contaminating proteins were expected to flow through (Coşkun, 2016). As shown in Figure 23b and Table 8, the partially purified ClpK was successfully purified to homogeneity (Figure 23b, Lane 6 to 10). The specific activity increases from 0,0423 units/mg in the supernatant to 11.13 units/mg in the homogenous protein sample (Table 8). The specific activity is of importance as it can be used to determine the purity of a protein following a dual purification procedure (Whitfield *et al.*, 1970).

A



B

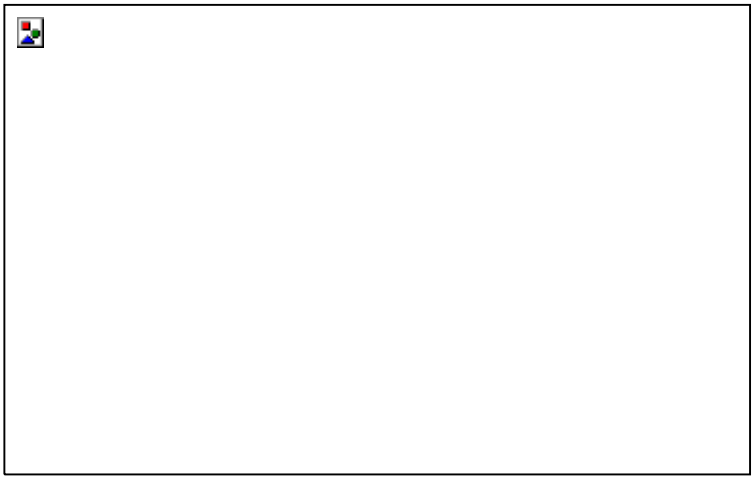


Figure 23: Purification of ClpK using anion exchange and affinity chromatography. A) Anion exchange purification. Lane 1: Molecular weight marker, Lane 2: crude sample, Lane 3: Flow through, Lane 4: Unbound (wash with buffer A), Lane 5: wash with 20% Buffer B, Lane 6-14: elution with 40% Buffer B, Lane 15: elution with 100% Buffer B. Elution samples were collected in 1 ml fractions. **B)** Affinity purification. The eluents from the anion exchange purification were used for further purification. Lane 1: Molecular weight marker, Lane 2: Load, Lane 3: flow through, Lane 4: unbound wash, Lane 5-11; elution with buffer C, Lane 12: elution with Buffer D.

Table 8: Purification table for ClpK

Steps	Volume (mL)	Protein (mg/ml)[b]	Total protein (mg)	Specific activity (units/mg)[c]	Total activity (units)	Yield (%)	Purity (%)
Crude extract[a]	30	0.0345	93.10	0,0423	3.94	100	0.38
Anion eluant[d]	30	0.0122	32.94	0,302	9.95	35.4	2.7
HisTag eluant[e]	10	0.117	1.17	11,13	13.02	1.26	100

[a] Soluble fraction obtained from 0.77 g of wet weight *E. coli* cell pellet (from 1 L of bacterial culture).
[b] Protein concentration determined by Bradford assay using BSA as a standard protein (Bradford, 1976).
[c] Calculated using the ATPase assay; the release of phosphate ions is measures as ATP is converted to ADP (A_{620nm}).
[d] Elution collected from the ion exchange column.
[e] Pooled eluant collected from the HisTag column.

3.8. Biophysical characterisation of ClpK

3.8.1. Enzyme activity assay

ClpK has been identified as an ATPase since it consists of structural motifs associated with the hydrolysis of ATP (Kress *et al.*, 2009, Maurizi and Xia, 2004, Schirmer *et al.*, 1996). However, the ability of ClpK to hydrolyse ATP *in vitro* has not been reported to date. Therefore, the ATPase activity of the purified protein was investigated using an ATPase assay. Figure 24 confirms that ClpK hydrolyses ATP and this is consistent with the primary structure of ClpK (Figure 8) which shows the presence of ATP hydrolysis domains. Furthermore, ClpK ATP hydrolysis is concentration dependant as enzyme activity increases from 0 $\pm$ 0 units/L to 10.43 $\pm$ 0.72 units/L as protein concentration increases from 0 mg/ml to 0.006 mg/ml.

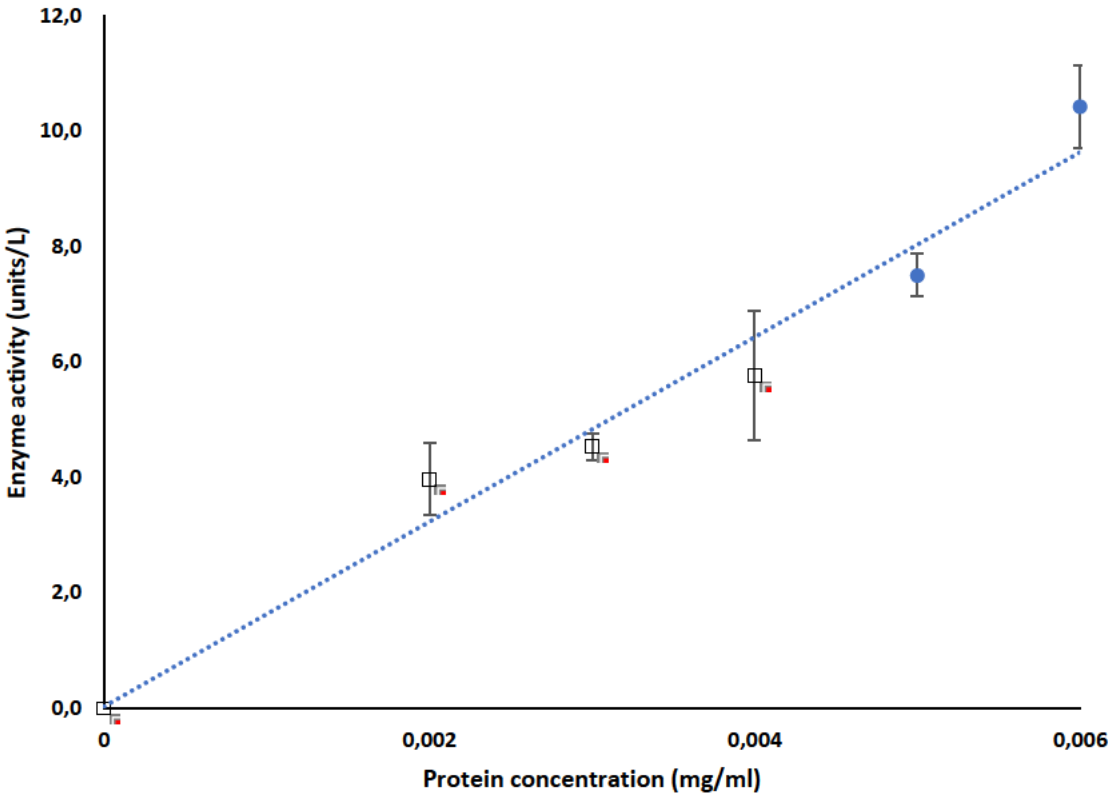


Figure 24: ClpK ATPase activity. ATPase activity was determined by incubating protein at varying concentrations (0-0.006 mg/ml) in reaction buffer containing 40 mM Tris, 80 mM NaCl, 8 mM $MgAc^2$, 1 mM EDTA, and 4 mM ATP (pH 7.5) for 30 minutes. The release of phosphate ions was determined at 620 nm. The absorbance values were then used to calculate enzyme activity. Each data point is an average of three independent measurements and the error bars represent standard deviation. Equation: y=1600,4x+0,0361, R^2: 0,9677.

3.8.2. Spectroscopical analysis

3.8.2.1. Far UV-CD spectroscopy

Circular dichroism (CD) is an optical spectroscopic method that is used to gain an insight into the secondary structure of the protein (Houde and Berkowitz, 2014, Miles and Wallace, 2015). Assessing CD data of a protein over the far-UV range (185 nm – 250 nm) produces a CD spectrum, which is indicative of the secondary structure fingerprint of the protein or peptide (Greenfield, 2006). Therefore, we used CD spectroscopy to characterise the secondary structure of ClpK , ClpK CD spectra shows a negative minima at 208 ± 1 and 221 ± 1 nm, and maxima at 193 ± 1 nm (Figure 25). Furthermore, the absence of a negative band of great magnitude at around 200 nm suggests that the purified ClpK is not highly disordered or unfolded (Rodger, 2013). This obtained CD spectra therefore correlates with the modelled of ClpK structure (Figure 9) and its virtual CD analysis (Figure 11).

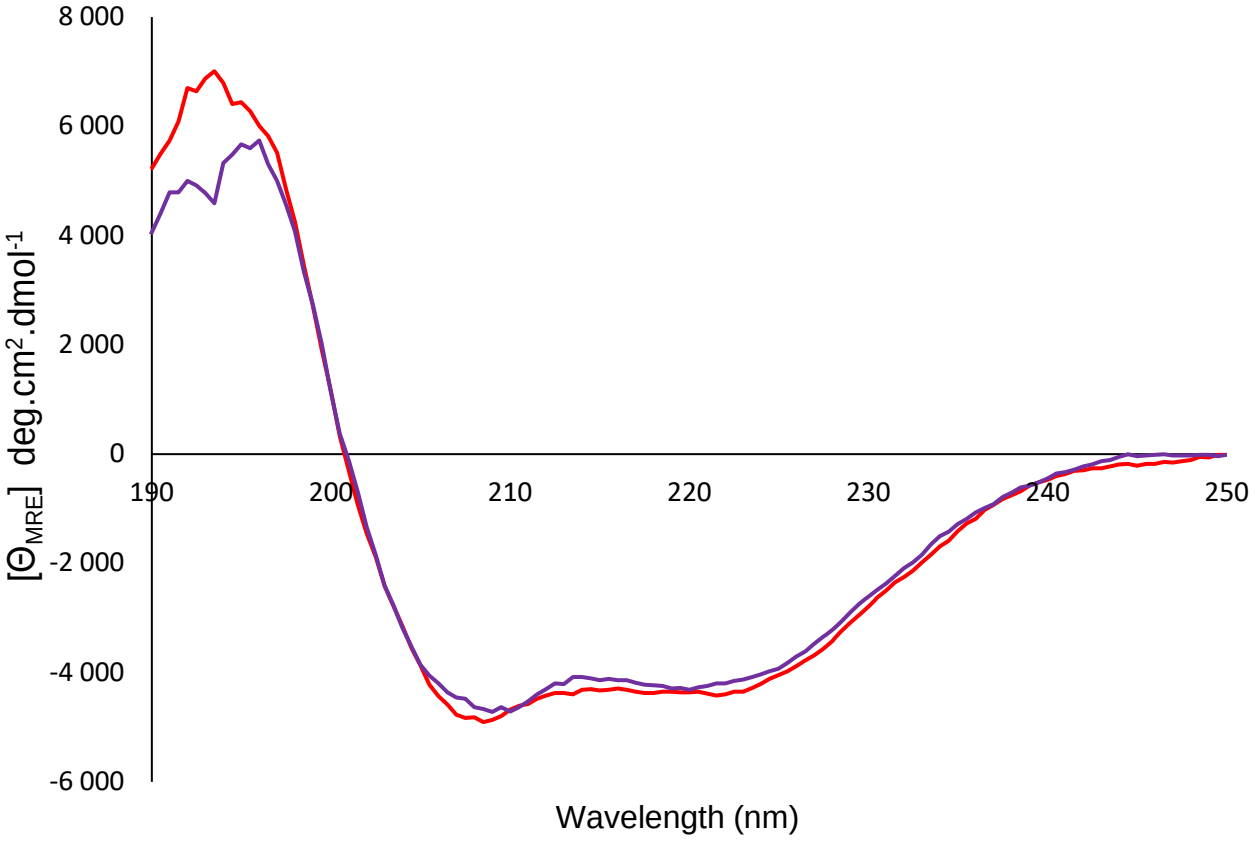


Figure 25: Far-UV CD spectroscopy of ClpK in the presence and absence of ATP. CD spectra of ClpK (2 µM) in sodium phosphate buffer (pH 7.4) in the presence of 0.2 mM ATP (blue) and in the absence of ATP (red). Each spectrum represents an average of five accumulated spectra derived at 20°C using a 2 mm path length quartz cuvette.

Knowing that ClpK binds ATP, CD was used to determine structural changes on ClpK upon ATP binding. In the presence of ATP, there was no shift observed on the CD spectra at 208 ± 1 and 221 ± 1 nm minima however, a slight shift in maxima from 193 ± 1 to 196 ± 1 nm was observed (Figure 25). Based on this observation, there seems to be a slight modification in the secondary structure content upon ATP binding. This indicates that ATP binding does impact the secondary structure content of ClpK however an alternative tool with better resolution needs to be used to determine the exact impact of ATP binding on protein structure and function.

3.8.2.2. Extrinsic fluorescence spectroscopy

Fluorescent spectroscopy was used to further probe the interaction between ClpK and ATP. Fluorescent probes such as 8-anilino-1-naphthalenesulfonic acid (ANS) and 3′-O-(N-methylanthraniloyl)-adenosine 5′-triphosphate (mant-ATP) can be used to investigate the structure and interaction of proteins (Stryer, 1968, Aranovich *et al.*, 2006, Gasymov and Glasgow, 2007). Mant-ATP is a sepctrofluorometric dye which is a derivative of ATP (Aranovich *et al.*, 2006). Figure 26A shows A higher fluorescence quantum yield is observed at 450 nm in the presence of protein and mant-ATP compared to the fluorescence quantum yield observed in the presence of mant-ATP alone. This is expected as mant-ATP exhibits an increased fluorescence intensity upon protein binding (Aranovich *et al.*, 2006). In Figure 26B using ANS as an extrinsic probe showed that ATP binds to ClpK as a hypsochromic shift (blue shift) in emission maxima was observed in the presence of ATP (Figure 26B). This hypsochromic shift from 495 nm to 507 nm indicates an increase in the energy released upon ATP binding (Wypych, 2015). Furthermore, this hypsochromic shift is accompanied with an increase in fluorescence intensity. The increase in fluorescence intensity in the presence of both ANS and mant-ATP suggest that the nucleotide binding domain is hydrophobic of (Aranovich *et al.*, 2006, Gasymov and Glasgow, 2007).

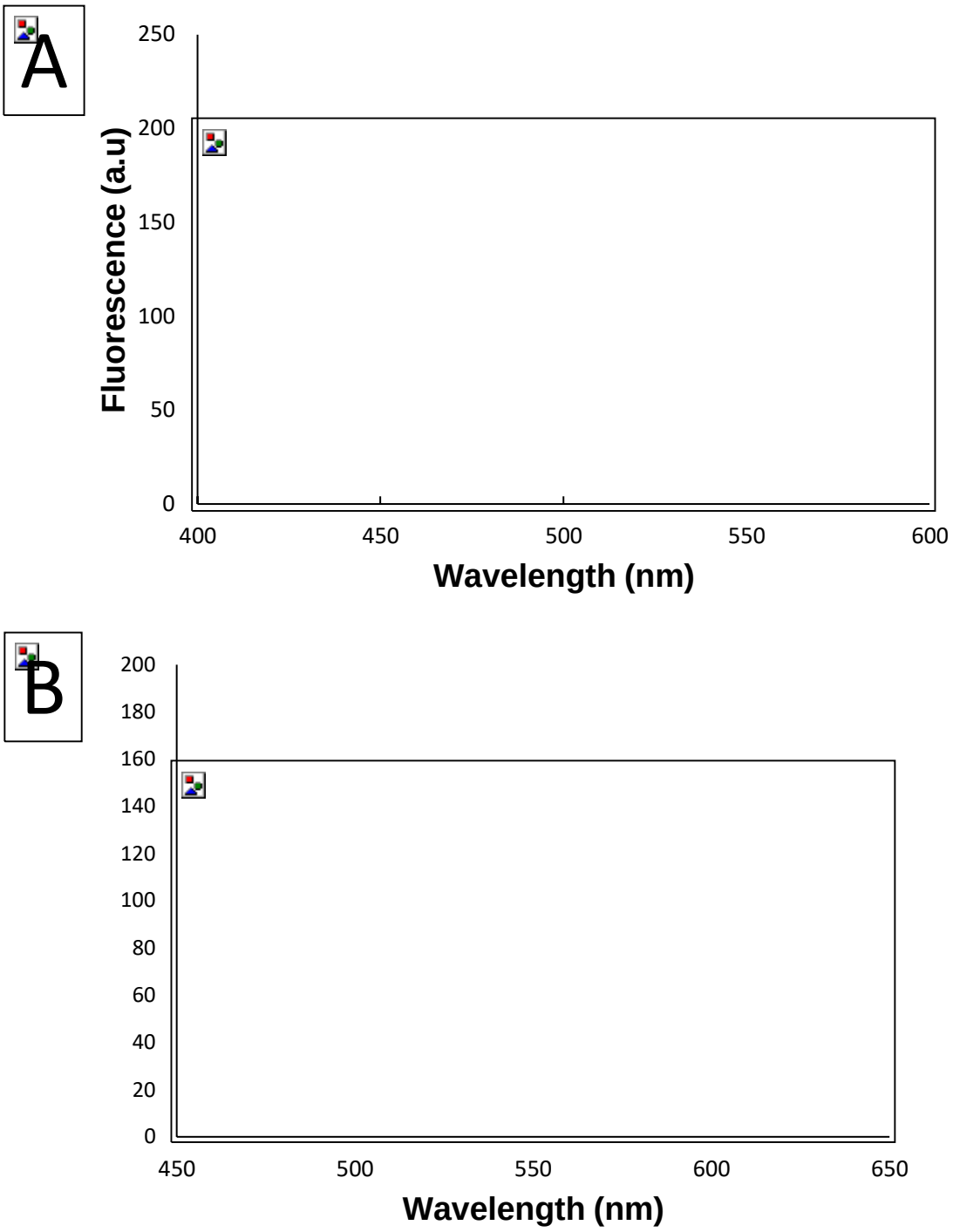


Figure 26: Tertiary structure analysis of ClpK using extrinsic fluorescence spectroscopy. **(A)** Fluorescence emission spectra of mant-ATP in the presence (red) or absence (blue) of 2.0 µM ClpK. The fluorescent nucleotides (10 µM) were excited at 355 nm and emission collected between 400 and 600 nm. Each spectrum represents an average of three accumulated spectra from three independent experiments. **(B)** ANS fluorescence spectroscopy of ClpK. ClpK (2.0 µM) with ANS alone (blue) and ANS in the presence of 0.2 mM ATP and ClpK (2.0 µM) (red). The fluorescence emission was recorded in wavelength region 400 – 650 nm after exciting the samples at 390 nm. The bandwidths were set at 5 nm and 10 nm for excitation and emission, respectively. The path length of the sample was 1 cm. Each spectrum represents an average of five accumulated spectra from three independent experiments. a.u is arbitrary unit.

3.8.3. Thermal stability

In order for pathogens to survive, it is important that they maintain protein homeostasis in extreme environments. Therefore, it is essential that pathogens contain proteins which can withstand extreme conditions and assist in the maintenance of protein homeostasis (Miotto *et al.*, 2018). As mentioned earlier, the *clpk* gene was identified to be present on a heat resistant locus (Bojer *et al.*, 2010). These authors suggested that the ClpK protein plays a role in providing thermostability to *K. pneumoniae*. In this regard, it is essential to investigate the thermostability of ClpK to gain an insight into the physical properties of the protein. The thermal properties of a protein can be investigated in one of two ways, firstly; determining the thermodynamic stability, and secondly; physical methods to investigate protein thermal resistance (Dehghan-Nayeri and Rezaei Tavirani, 2015, Luke *et al.*, 2007, Miotto *et al.*, 2018).

The thermodynamic stability assumes that protein folding is reversible and is calculated by subtracting the free energy between the unfolded and folded states (Luke *et al.*, 2007). The thermal resistance of the protein is determined by calculating the temperature at which the protein denatures, melting temperature (Tm) (Dehghan-Nayeri and Rezaei Tavirani, 2015). There are a number of methods to investigate the thermal properties of a protein, these include; fluorescence activity assays, differential scanning calorimetry, and CD spectroscopy (Greenfield, 2006). Therefore, this study investigated the effect of temperature on the structural stability of ClpK in one of two ways.

Firstly, the protein was incubated at varying temperatures over time and the effect of temperature variation on ClpK was visualized using non-reducing SDS-PAGE (Figure 27). A single band is observed from lane 2 (10^{o}C) to lane 9 (70^{o}C) suggesting that ClpK is structurally stable up to 70^{o}C. The protein appears to partially lose structural stability in Lane 10 (80^{o}C), Lane 11 (90^{o}C) and Lane 12 (100^{o}C) where multiple bands of high molecular weight are observed (between 130 to 250 kDa). This suggests that ClpK aggregates at higher temperatures and perhaps loses activity.

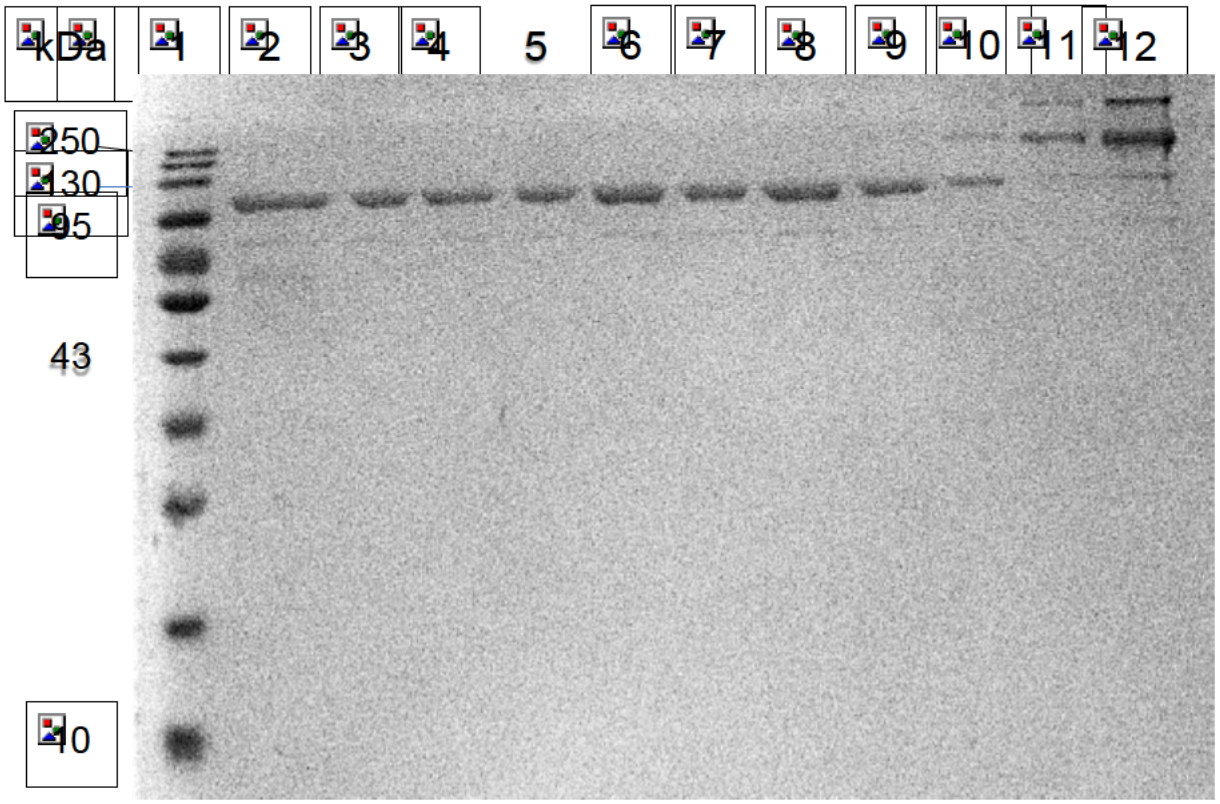


Figure 27: Effect of varying temperatures on protein structure as seen on a non-reducing SDS-PAGE gel. Purified ClpK was incubated at temperatures ranging from 10ºC to 100ºC for 1 h. Lane 1: MWM, Lane 2: protein at room temperature, Lane 3: protein incubated at 10ºC, Lane 4: protein incubated at 20ºC, Lane 5: protein incubated at 30ºC, Lane 6: protein incubated at 40ºC, Lane 7: protein incubated at 50ºC, Lane 8: protein incubated at 60ºC, Lane 9: protein incubated at 70ºC, Lane 10: protein incubated at 80ºC, Lane 11: protein incubated at 90ºC, and Lane 12: protein incubated at 100ºC.

ClpK displays an ellipticity trough at 222 nm (Figure 24) which is proportional to the α-helical protein content and is generally not observed in unfolded proteins. Therefore, the change in ellipticity at 222 nm was used to probe the stability of ClpK in response to increasing temperature. The change in ellipticity is indicative of protein unfolding as observed in Figure 28. Furthermore, a characteristic sigmoidal curve is obtained from the CD data suggesting that ClpK unfolds in a cooperative manner. The Tm for ClpK was estimated to be around 65ºC.

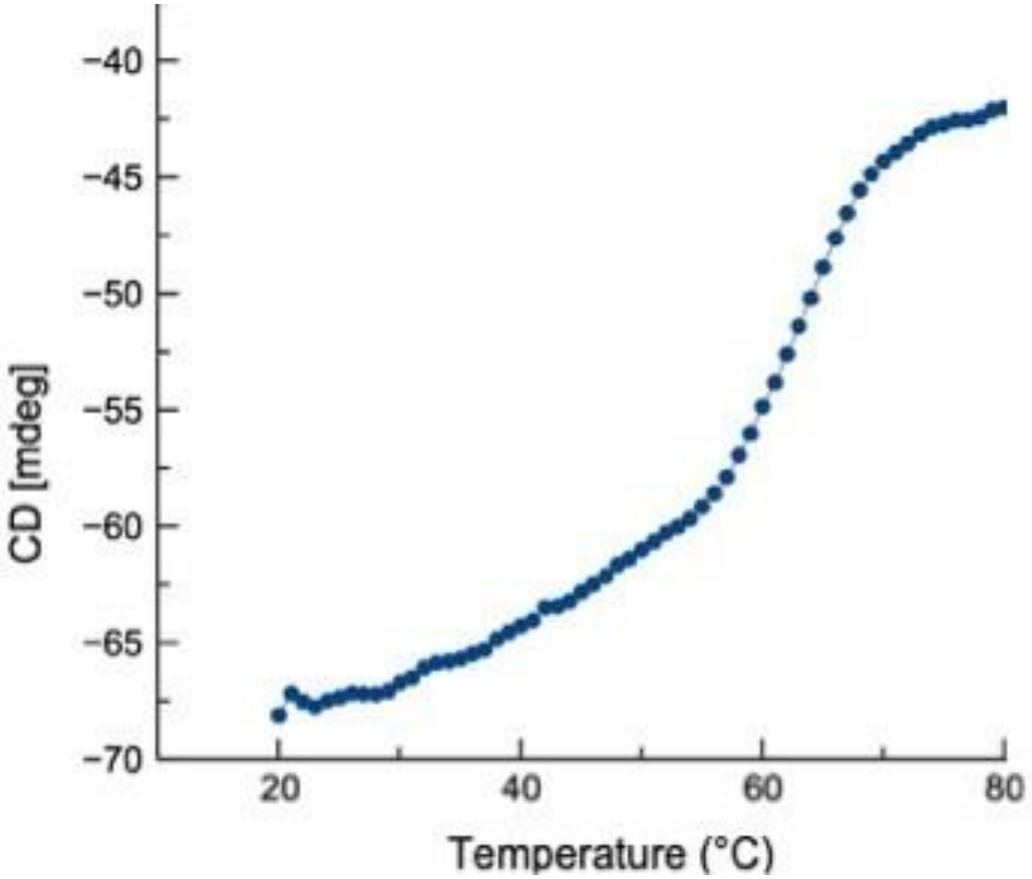


Figure 28: Circular Dichroism thermal melt from 20ºC to 80ºC for ClpK carried out at 222 nm. The protein profile was recorded as the temperature was increased from 20ºC to 80ºC. The graph obtained from the analysis was smoothed using MagicPlot 3.0.1.

Chapter 4: Discussion and Conclusion

Clp proteins play an important role in maintaining cell protein homoeostasis and are particularly of interest in pathogens as they enable their survival in harsh environmental conditions (Ahyoung *et al.*, 2015, Capestany *et al.*, 2008, Frees *et al.*, 2014). Therefore, Clp proteins could be potential targets to kill pathogens, especially those which have been classed as being antibiotic resistant. This study focused on the *in silico* and *in vitro* analysis of a novel Clp ATPase referred to as ClpK which was identified as being present in a heat resistant *K. pneumonia* clinical isolate.

4.1. *In silico* ClpK analysis

The presence and diversity of Clp ATPases is a continuously studied field in a number of pathogens however, the presence of Clp ATPases in the *Klebsiella* species has not been studied adequately (Thibault *et al.*, 2006, Wojtyra *et al.*, 2003, Maurizi and Xia, 2004). To address this knowledge gap, bioinformatic analysis was performed to gain an insight into the distribution and divergence of ClpK in the *Klebsiella* species. Investigation of the distribution of Clp ATPases showed that the distribution of *clpk* was considerably less compared to that of *clpa*, *clpb* and *clpx* (Figure 5, Figure 6). This was unexpected as ClpK is reported to be ubiquitous and is found in various species such as *Escherichia coli, Enterobacter cloacae,* and other *Klebsiella* strains other than *Klebsiella pneumoniae* (Bojer *et al.*, 2010). Also, taking into account that the *clpk* gene is hypothesized to be transferred through horizontal gene transfer, one would expect it to be present among a greater number of the investigated strains (Bojer *et al.*, 2010). However, the absence of ClpK in a majority of the studied strains may indicate the selective uptake of the *clpk* gene by pathogens to enable them to survive harsh environmental conditions. Our findings are similar to those obtained by Bojer *et al.* (2013) where they reported that only 31 out of the 105 clinical isolates contained the *clpk* gene suggesting that only certain strains acquired the plasmid through horizontal transfer to subsequently express the ClpK protein (Bojer *et al.*, 2010). Furthermore, it was observed that none of the investigated strains contained *clpc* (Figure 5, Figure 6). This observation correlates with the findings of Miller *et al.* (2018) who suggested that ClpC is a common ancestorial protein of ClpK and ClpA. Therefore inferring that the investigated strains may have contained ClpC at some point, which

have now mutated into either ClpA or ClpK depending on environmental conditions (Miller *et al.*, 2018).

The divergence of Clp ATPases within the *Klebsiella* species was investigated using phylogenetic tree analysis. Phylogenetic tree analysis is a useful technique as it allows for the representation of hierarchical biological data and shows evolutionary relationships between species and how they have evolved over time (Pavlopoulos *et al.*, 2010). The branches of ClpK, ClpB and ClpA connected before divergence (Figure 7), suggesting that these proteins all descended from a common ancestor and therefore are termed to be homologs (Berg Jm, 2002). The varying percentage identity (Table 4) and the visible divergence within ClpK indicates a divergence in the evolution of the protein (Miller *et al.*, 2018).

Taking into consideration that ClpK is a novel protein and its three-dimensional structure is yet to be determined, homology modelling was performed to gain an insight into the structural features of ClpK. ClpB was identified as being an appropriate template to model ClpK as it met the criteria set out for homology model templates. Additionally, the alignment of these two sequences showed good sequence coverage and conservation of the Clp ATPase domain (Figure 8). One of the striking differences between the two structures is the difference in length of the linker region connecting NBD1 and NBD2 (Figure 8, Figure 9A, Figure 9B). A linker region connects two adjacent domains and has been found to have a role in protein stability, protein folding and domain-domain interactions (Bae et al., 2005). As such, the ClpB linker region plays a role in protein stability and interdomain communication between NBD1 and NBD2 (Lee et al., 2003). However, the function of the ClpK linker domain is yet to be investigated. The difference in amino acid length of the ClpK and ClpB linker region indicates that the regions of both proteins may transport a different range of substrates. Additionally, non-identical residues were identified in the Walker B motif situated in NBD2, the lysine residue in ClpK is substituted with arginine in ClpB, both these amino acid residues are basic and therefore this is termed as a conserved substitution (French and Robson, 1983) (Figure 8). However, both the Walker B motifs contain arginine which is believed to coordinate the binding of magnesium ions for ATP hydrolysis (Chiraniya et al., 2013) (Figure 8). Therefore, this indicates that despite

the difference in identity between the Walker B motifs there may only be subtle differences in the metal ions bound by the Walker B motifs of ClpK and ClpB.

To confirm the stability of the modelled structure, various molecular dynamic (MD) simulation studies were performed. MD dynamic simulations are a popular tool in molecular biology used to investigate the behaviour of proteins and other biomolecules down to an atomic level (Hollingsworth and Dror, 2018). These simulations can be used out to investigate, i) atom motion (potential energy) or flexibility of a biomolecule, ii) the accuracy of a modelled structure, and iii) the effect of changes on a biomolecular system (i.e.; the effect the presence of ATP has on a particular protein) (Hollingsworth and Dror, 2018, Hospital et al., 2015). The potential energy profile of the modelled ClpK and template ClpB were investigated and indicated a slight, insignificant shift (Figure 14), this suggested that both proteins were relatively stable as there was no extreme force experienced by any atom due to the positioning of other atoms (Alagu Lakshmi et al., 2020, Hollingsworth and Dror, 2018). Proteins are dynamic in nature, these biomolecules are constantly in motion and the conformational change that results from this movement plays an essential role in determining protein function (Yang et al., 2014). It was observed the both ClpK and ClpB are highly dynamic in nature through two analyses. Firstly; the increase in the RMSD values observed over 100 ns indicated confirmational changes and this observation was consistent with the dynamic nature of proteins (Tiwari and Mohanty, 2013, Buchner, 2019) (Figure 15, Figure 16). Secondly; the features of the radius of gyration profiles of ClpK and ClpB (Figure 17) indicated structural transformation, suggesting that the proteins were constantly transforming during simulation. Both ClpK and ClpB are chaperones and it is therefore expected that both proteins would be highly dynamic in nature. Chaperones are important for cell protein homeostasis and their highly dynamic nature plays a key role in facilitating the formation of a complex between the chaperone and partnered protein (Sučec et al., 2021).

Seeing that the ClpK protein has not been previously expressed and purified, the protein disorder and binding disorder of ClpK was investigated through the IUPred2A server (Mészáros et al., 2018). Disordered protein regions make protein purification and biophysical characterisation studies difficult as they do not adopt a stable conformation (Babu, 2016). Therefore, investigating the protein disorder before performing protein

expression and purification allows for the conservation of time and resources. It was observed that a considerably low percentage of ClpK residues were predicted to be disordered (Figure 18), therefore it was concluded that the protein can be expressed, purified and used for subsequent biophysical characterisation. It is however interesting to observe that the C-terminal domain of modelled ClpK was fairly disordered (Figure 18). The C-terminal domain is generally referred to as the D2-small domain and forms a tight interface with the D2-large domain of an adjacent subunit, therefore providing sufficient binding energy to stabilise functional assembly (Zolkiewski, 2006). Subsequently, studies could focus on investigating compounds which bind to the C-terminal domain to disrupt the interaction the C-terminal domain and the D2-large domain.

4.2. *In vitro* analysis of ClpK

The ClpK gene was synthesized and cloned into pColdI for protein expression. Protein expression can be carried out using a number of vectors and host cells. Choosing a vector and getting the desired DNA cloned is the first step in recombinant protein expression. There are a number of vectors (such as the pET series, pUC series and pCold system) that can be used for expression. Each vector contains promoters, multiple cloning sites and replicons with their respective advantages and disadvantages (Rosano and Ceccarelli, 2014). pColdI is a vector which forms a part of the cold-shock expression system (pCold system). This system is used to express desired recombinant proteins at temperatures as low as 15^{o}C to suppress the production of undesirable background proteins (Sugiki *et al.*, 2017). The second step in recombinant protein expression is to choose suitable host cells for protein expression. There are a number of host cells cultures available for protein expression, these include; bacteria, yeast, mammal and plant cells (Demain and Vaishnav, 2009). The factors to consider when choosing host cells include the production speed, protein yield, protein functionality and protein quality. The use of *E. coli* for the expression of recombinant proteins is common and advantageous for the following reasons; *E. coli* cells have fast growth kinetics and the media for cell growth can be made from inexpensive and readily available components (Rosano and Ceccarelli, 2014). Competent *E. coli* BL21 (DE3) cells were transformed with the pColdI-ClpK construct for ClpK expression. Expression was performed at 15^{o}C for 24 h in the presence of varying concentrations of isopropyl β-d-1-thiogalactopyranoside (IPTG). IPTG is a

metabolite which induces the promoter to promote the expression of the recombinant protein (Briand *et al.*, 2016). It was observed that the optimum IPTG concentration for expression was 0.25 mM (Figure 22).

Following successful expression of soluble ClpK, the next step was to purify the expressed protein. It is important to purify proteins to study protein function, determine protein structure, and for applying the protein of interest in an industrial or pharmaceutical environment (Smith, 2005). The expressed, soluble ClpK protein fraction was purified using both anion exchange chromatography and affinity chromatography (Figure 23). Ion exchange chromatography separates molecules based on their overall charge which is dependent on the isoelectric point (pI) of the protein and the overall pH of the system (Benedini *et al.*, 2020). There are two types of ion exchange chromatography: anion and cation. Anion exchange chromatography involves the use of a positively charged resin which binds to negatively charged proteins. Cation exchange chromatography involves the use of a negatively charged resin which binds positively charged proteins (Benedini *et al.*, 2020, Duong-Ly and Gabelli, 2014). Affinity chromatography is a technique that exploits the biological property of proteins to specifically and reversibly bind to ligands (Cuatrecasas, 1970). The purification of proteins containing the His tag is based on the affinity of the histidine residues for metal ions such as nickel or copper (Spriestersbach *et al.*, 2015). ClpK has a pI of 5.61 and is therefore negatively charged at a pH of 7.4. Taking this into consideration, anion exchange chromatography was used for ClpK purification, and the protein was eluted using an increasing salt concentration. Anion exchange chromatography assists in removing background proteins with a different pI to the protein of interest. The protein eluted from the anion exchange column was then pooled and passed through an affinity chromatography column, taking advantage of the fact that the ClpK protein contained a His tag. Affinity chromatography separates proteins which do not have an affinity for the nickel resin from the target protein which does contain a His tag and therefore does have affinity for the nickel resin.

The biological activity of the purified protein was then investigated. It is important to investigate biological activity to determine whether the purified protein is active. Although all proteins are naturally active, it is important to assess whether a recombinant protein is truly active as the choice of the vector or host cell during

expression may affect the biological activity of the protein (Rosano and Ceccarelli, 2014). Additionally, the conditions used during purification may influence cell activity, for example if a protein is exposed to high salt concentrations during purification it may denature and not be biologically active (Rosano and Ceccarelli, 2014). The activity of ClpK was assayed using the ATPase assay because ClpK is part of the AAA+ superfamily, and therefore it is expected that this protein would hydrolyse ATP (Ahyoung *et al.*, 2015, Capestany *et al.*, 2008). The ATPase assay measures the release of free phosphate as adenosine triphosphate (ATP) is converted into adenosine diphosphate (ADP) and free phosphate is released (Palmgren, 1990). ClpK ATPase activity results (Figure 24) show that the purified protein was biology active. Furthermore the ATPase activity of ClpK increased as protein concentration increased (Figure 24). This observed result was expected as ClpK is an enzyme and it is known that increasing enzyme concentration speeds up enzyme reactions until a point of saturation is reached. An alternative assay which could be used to investigate biological activity of ClpK is the Luciferase assay. This assay would focus on the proteins ability to unfold luciferase. Additional studies could then be performed using the luciferase assay to then investigate if the protein was able to refold the luciferase enzyme following incubation for a few hours or a few days (Herbst *et al.*, 1998).

To determine the structural properties of ClpK, CD and fluorescence spectroscopy was used. CD is an optical spectroscopic method that is used to gain an insight into the secondary structure of the protein (Houde and Berkowitz, 2014, Miles and Wallace, 2015). The method uses the fact that optically active molecules absorb left- and right-handed circularly polarized light differently (Miles and Wallace, 2015). Fluorescent probes can be used to investigate the structure and interactions of proteins (Stryer, 1968). There are two types of fluorescent probes, namely; intrinsic and extrinsic chromophores (Stryer, 1968). Intrinsic fluorophores are the aromatic side chains of tryptophan, tyrosine and phenylalanine (Stryer, 1968). Extrinsic fluorophores are attached to the proteins at certain sites (Stryer, 1968). An example of an extrinsic fluorophore is 8-anilino-1-naphthalenesulfonic acid (ANS) (Gasymov and Glasgow, 2007). ANS is commonly used in protein folding studies and is less commonly used to detect misfolded protein conformations (Latypov *et al.*, 2008, Zuo *et al.*, 2012). ANS binds to hydrophobic residues on proteins, this results in an increase in the quantum fluorescence yield. However, these regions are assumed to be rare in native and

absent in denatured proteins (Ali *et al.*, 1999). Therefore, ideally an increase in quantum yield should not be observed if the protein under investigation is denatured or in its native, functioning state.

The ClpK CD spectra obtained in this work (Figure 25) agrees with the homology model of ClpK which showed that ClpK is a high α-helical and lower β-structural content. Also, from the ClpK CD spectra in the presence of ATP, it evident that the binding of ATP does induce some conformation changes on the structure of ClpK (Figure 25). This conformational change was indicated by a slight shift in maxima, however further ATP binding studies will have to be performed to conclude the extent to which ATP binding impacts ClpK structure.

The fluorescence studies of the purified protein were performed using two fluorescent probes, namely; mant-ATP and ANS (Figure 26A, Figure 26B). An increase in the fluorescence quantum yield was observed in the presence of mant-ATP and protein (Figure 26A), this indicates that the probe was able to bind to the nucleotide binding domain of ClpK (Aranovich *et al.*, 2006). Molecular docking studies could be performed to investigate whether the probe binds to both the nucleotide binding domains and the area of the domain to which it binds. Furthermore, a blue shift was observed during ANS binding analysis in the presence of ATP (Figure 26B), this shift indicates an increase in energy released (Wypych, 2015). Potentially, this increase could be attributed to ATP hydrolysis, however this possibility would have to be further investigated. The increase in fluorescence intensity in the presence of ANS indicate that the probe binds to hydrophobic residues on the protein. Once again, molecular docking studies could be performed to investigate how ANS binds to ClpK in the presence and absence of ATP, other nucleotides and metal ions.

With the knowledge that ClpK was implicated in the thermostable properties of ClpK, baseline thermal analysis studies were performed through CD analysis and protein incubation. CD analysis can be used to investigate the effect of temperature on protein structure (Greenfield, 2006). This is done by increasing temperature and measuring protein absorbance at a specific absorbance to monitor the change in protein structure as temperature increases (Greenfield, 2006). CD temperature analysis and protein incubation showed that the protein structure was stable up to approximately 65^{o}C (Figure 27, Figure 28). Although structural change is observed, it now becomes

important to test the activity of the protein at higher temperatures to investigate the bioactivity of the protein at those temperatures. Additionally, further studies could focus on investigating if ATP and metal ions stabilise the protein structure of ClpK at higher temperatures.

In conclusion, the observations made in this study allow us to form a baseline for the study of novel ClpK. Further investigations could focus on the crystallization of the ClpK structure, and drug binding studies could be performed to investigate the effect of potential drugs on the protein. Furthermore, various domains of the protein could be studied individually to determine their function in protein stability and function.

Chapter 5: References

ABAD, J. P. 2011. ATPase. *In:* GARGAUD, M., AMILS, R., QUINTANILLA, J. C., CLEAVES, H. J., IRVINE, W. M., PINTI, D. L. & VISO, M. (eds.) *Encyclopedia of Astrobiology.* Berlin, Heidelberg: Springer Berlin Heidelberg.

AHYOUNG, A. P., KOEHL, A., CASCIO, D. & EGEA, P. F. 2015. Structural mapping of the C lp B ATP ases of Plasmodium falciparum: Targeting protein folding and secretion for antimalarial drug design. *Protein Science,* 24**,** 1508-1520.

AINODA, Y., AOKI, K., ISHII, Y., OKUDA, K., FURUKAWA, H., MANABE, R., SAHARA, T., NAKAMURA-UCHIYAMA, F., KUROSU, H., ANDO, Y., FUJISAWA, M., HOSHINO, H., ARIMA, H. & OHNISHI, K. 2019. Klebsiella pneumoniae carbapenemase (KPC)-producing Klebsiella pneumoniae ST258 isolated from a Japanese patient without a history of foreign travel - a new public health concern in Japan: a case report. *BMC Infectious Diseases,* 19**,** 20.

ALAGU LAKSHMI, S., SHAFREEN, R., PRIYA, A. & SHUNMUGIAH, K. 2020. Ethnomedicines of Indian origin for combating COVID-19 infection by hampering the viral replication: using structure-based drug discovery approach. *Journal of Biomolecular Structure and Dynamics,* 39**,** 1-16.

ALEXANDER, E. M., KREITLER, D. F., GUIDOLIN, V., HURBEN, A. K., DRAKE, E., VILLALTA, P. W., BALBO, S., GULICK, A. M. & ALDRICH, C. C. 2020. Biosynthesis, Mechanism of Action, and Inhibition of the Enterotoxin Tilimycin Produced by the Opportunistic Pathogen Klebsiella oxytoca. *ACS Infectious Diseases,* 6**,** 1976-1997.

ALI, M. S. & BAEK, K.-H. 2020. Protective Roles of Cytosolic and Plastidal Proteasomes on Abiotic Stress and Pathogen Invasion. *Plants,* 9**,** 832.

ALI, V., PRAKASH, K., KULKARNI, S., AHMAD, A., MADHUSUDAN, K. P. & BHAKUNI, V. 1999. 8-Anilino-1-naphthalene Sulfonic Acid (ANS) Induces Folding of Acid Unfolded Cytochrome c to Molten Globule State as a Result of Electrostatic Interactions. *Biochemistry,* 38**,** 13635-13642.

AMOR, A. J., SCHMITZ, K. R., BAKER, T. A. & SAUER, R. T. 2019. Roles of the ClpX IGF loops in ClpP association, dissociation, and protein degradation. *Protein Science,* 28**,** 756-765.

ARANOVICH, A., GDALEVSKY, G., COHEN-LURIA, R., FISHOV, I. & PAROLA, A. 2006. Membrane-catalyzed nucleotide exchange on DnaA - Effect of surface molecular crowding. *The Journal of biological chemistry,* 281**,** 12526-34.

BABOUEE, B., WIDMER, A., DUBUIS, O., CIARDO, D., DROZ, S., BETSCH, B. Y., GARZONI, C., FÜHRER, U., BATTEGAY, M. & FREI, R. 2011. Emergence of four cases of KPC-2 and KPC-3-carrying Klebsiella pneumoniae introduced to Switzerland, 2009–10. *Eurosurveillance,* 16**,** 19817.

BABU, M. M. 2016. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochem Soc Trans,* 44**,** 1185-1200.

BAE, K., MALLICK, B. K. & ELSIK, C. G. 2005. Prediction of protein interdomain linker regions by a hidden Markov model. *Bioinformatics,* 21**,** 2264-2270.

BAJAJ, D. & BATRA, J. K. 2012. The C-Terminus of ClpC1 of Mycobacterium tuberculosis Is Crucial for Its Oligomerization and Function. *PLOS ONE,* 7**,** e51261.

BAKER, T. A. & SAUER, R. T. 2012. ClpXP, an ATP-powered unfolding and protein-degradation machine. *Biochim Biophys Acta,* 1823**,** 15-28.

BARRIOS-CAMACHO, H., AGUILAR-VERA, A., BELTRAN-ROJEL, M., AGUILAR-VERA, E., DURAN-BEDOLLA, J., RODRIGUEZ-MEDINA, N., LOZANO-AGUIRRE, L., PEREZ-CARRASCAL, O. M., ROJAS, J. & GARZA-RAMOS, U. 2019. Molecular epidemiology of Klebsiella variicola obtained from different sources. *Scientific Reports,* 9**,** 10610.

BEAUGERIE, L., METZ, M., BARBUT, F., BELLAICHE, G., BOUHNIK, Y., RASKINE, L., NICOLAS, J. C., CHATELET, F. P., LEHN, N. & PETIT, J. C. 2003. Klebsiella oxytoca as an agent of antibiotic-associated hemorrhagic colitis. *Clin Gastroenterol Hepatol,* 1**,** 370-6.

BECKER, L., BUNK, B., ELLER, C., STEGLICH, M., PFEIFER, Y., WERNER, G. & NÜBEL, U. 2015. Complete genome sequence of a CTX-M-15-producing Klebsiella pneumoniae outbreak strain from multilocus sequence type 514. *Genome announcements,* 3.

BENEDINI, L. J., FIGUEIREDO, D., CABRERA-CRESPO, J., GONÇALVES, V. M., SILVA, G. G., CAMPANI, G., ZANGIROLAMI, T. C. & FURLAN, F. F. 2020. Modeling and simulation of anion exchange chromatography for purification of proteins in complex mixtures. *J Chromatogr A,* 1613**,** 460685.

BENGOECHEA, J. A. & SA PESSOA, J. 2018. Klebsiella pneumoniae infection biology: living to counteract host defences. *FEMS Microbiology Reviews,* 43**,** 123-144.

BERG JM, T. J., STRYER L. 2002. Section 7.1, Homologs Are Descended from a Common Ancestor. *Biochemistry.* 5th edition ed. New York: W H Freeman.

BHANDARI, V., WONG, K. S., ZHOU, J. L., MABANGLO, M. F., BATEY, R. A. & HOURY, W. A. 2018. The Role of ClpP Protease in Bacterial Pathogenesis and Human Diseases. *ACS Chem Biol,* 13**,** 1413-1425.

BIDEWELL, C. A., WILLIAMSON, S. M., ROGERS, J., TANG, Y., ELLIS, R. J., PETROVSKA, L. & ABUOUN, M. 2018. Emergence of Klebsiella pneumoniae subspecies pneumoniae as a cause of septicaemia in pigs in England. *PLoS One,* 13**,** e0191958.

BJORNSDOTTIR-BUTLER, K., MCCARTHY, S. A., DUNLAP, P. V., TIMME, R. E. & BENNER, R. A. 2015. Draft genome sequences of histamine-producing Photobacterium kishitanii and Photobacterium angustum, isolated from albacore (Thunnus alalunga) and yellowfin (Thunnus albacares) tuna. *Genome announcements,* 3.

BOC, A., DIALLO, A. B. & MAKARENKOV, V. 2012. T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic acids research,* 40**,** W573-W579.

BOJER, M. S., STRUVE, C., INGMER, H., HANSEN, D. S. & KROGFELT, K. A. 2010. Heat resistance mediated by a new plasmid encoded Clp ATPase, ClpK, as a possible novel mechanism for nosocomial persistence of Klebsiella pneumoniae. *PloS one,* 5**,** e15467.

BOJER, M. S., STRUVE, C., INGMER, H. & KROGFELT, K. A. 2013. ClpP-dependent and-independent activities encoded by the polycistronic clpK-encoding locus contribute to heat shock survival in Klebsiella pneumoniae. *Research in microbiology,* 164**,** 205-210.

BRADFORD, M. M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry,* 72**,** 248-254.

BRENNER, S. E. 2001. A tour of structural genomics. *Nat Rev Genet,* 2**,** 801-9.

BRIAND, L., MARCION, G., KRIZNIK, A., HEYDEL, J.-M., ARTUR, Y., GARRIDO, C., SEIGNEURIC, R. & NEIERS, F. 2016. A self-inducible heterologous protein expression system in Escherichia coli. *Scientific reports,* 6**,** 33037.

BRÖTZ-OESTERHELT, H. & SASS, P. 2014. Bacterial caseinolytic proteases as novel targets for antibacterial treatment. *International Journal of Medical Microbiology,* 304**,** 23-30.

BUCHNER, J. 2019. Molecular chaperones and protein quality control: an introduction to the JBC Reviews thematic series. *J Biol Chem,* 294**,** 2074-2075.

BURALL, L. S., GRIM, C., GOPINATH, G., LAKSANALAMAI, P. & DATTA, A. R. 2014. Whole-genome sequencing identifies an atypical Listeria monocytogenes strain isolated from pet foods. *Genome announcements,* 2.

BUYS, H., MULOIWA, R., BAMFORD, C. & ELEY, B. 2016. Klebsiella pneumoniae bloodstream infections at a South African children's hospital 2006–2011, a cross-sectional study. *BMC Infectious Diseases,* 16**,** 570.

CALFEE, D. P. 2017. Recent advances in the understanding and management of Klebsiella pneumoniae. *F1000Research,* 6.

CAMACHO, C., COULOURIS, G., AVAGYAN, V., MA, N., PAPADOPOULOS, J., BEALER, K. & MADDEN, T. L. 2009. BLAST+: architecture and applications. *BMC bioinformatics,* 10**,** 1-9.

CAPESTANY, C. A., TRIBBLE, G. D., MAEDA, K., DEMUTH, D. R. & LAMONT, R. J. 2008. Role of the Clp system in stress tolerance, biofilm formation, and intracellular invasion in Porphyromonas gingivalis. *Journal of bacteriology,* 190**,** 1436-1446.

CARUGO, O. & PONGOR, S. 2001. A normalized root-mean-square distance for comparing protein three-dimensional structures. *Protein Sci,* 10**,** 1470-3.

CASHIKAR, A. G., SCHIRMER, E. C., HATTENDORF, D. A., GLOVER, J. R., RAMAKRISHNAN, M. S., WARE, D. M. & LINDQUIST, S. L. 2002. Defining a Pathway of Communication from the C-Terminal Peptide Binding Domain to the N-Terminal ATPase Domain in a AAA Protein. *Molecular Cell,* 9**,** 751-760.

CHAPMAN, P., FORDE, B. M., ROBERTS, L. W., BERGH, H., VESEY, D., JENNISON, A. V., MOSS, S., PATERSON, D. L., BEATSON, S. A., HARRIS, P. N. A. & MELLMANN, A. 2020. Genomic Investigation Reveals Contaminated Detergent as the Source of an Extended-Spectrum-β-Lactamase-Producing Klebsiella michiganensis Outbreak in a Neonatal Unit. *Journal of Clinical Microbiology,* 58**,** e01980-19.

CHEN, Y., BROOK, T. C., SOE, C. Z., APOS, NEILL, I., ALCON-GINER, C., LEELASTWATTANAGUL, O., PHILLIPS, S., CAIM, S., CLARKE, P., HALL, L. J. & HOYLES, L. 2020. Preterm infants harbour diverse Klebsiella populations, including atypical species that encode and produce an array of antimicrobial resistance- and virulence-associated factors. *Microbial Genomics,* 6.

CHIRANIYA, A., FINKELSTEIN, J., O'DONNELL, M. & BLOOM, L. B. 2013. A novel function for the conserved glutamate residue in the walker B motif of replication factor C. *Genes (Basel),* 4**,** 134-51.

CONLAN, S., PARK, M., DEMING, C., THOMAS, P. J., YOUNG, A. C., COLEMAN, H., SISON, C., WEINGARTEN, R. A., LAU, A. F., DEKKER, J. P., PALMORE, T. N., FRANK, K. M. & SEGRE, J. A. 2016. Plasmid Dynamics in KPC-Positive Klebsiella pneumoniae during Long-Term Patient Colonization. *mBio,* 7.

CONLAN, S., THOMAS, P. J., DEMING, C., PARK, M., LAU, A. F., DEKKER, J. P., SNITKIN, E. S., CLARK, T. A., LUONG, K. & SONG, Y. 2014. Single-molecule sequencing to track plasmid diversity of hospital-associated carbapenemase-producing Enterobacteriaceae. *Science translational medicine,* 6**,** 254ra126-254ra126.

COŞKUN, O. 2016. Separation techniques: Chromatography. *Northern Clinics of Istanbul,* 3.

CUATRECASAS, P. 1970. Protein purification by affinity chromatography. Derivatizations of agarose and polyacrylamide beads. *J Biol Chem,* 245**,** 3059-65.

DARBY, A., LERTPIRIYAPONG, K., SARKAR, U., SENEVIRATNE, U., PARK, D. S., GAMAZON, E. R., BATCHELDER, C., CHEUNG, C., BUCKLEY, E. M., TAYLOR, N. S., SHEN, Z., TANNENBAUM, S. R., WISHNOK, J. S. & FOX, J. G. 2014. Cytotoxic and pathogenic properties of Klebsiella oxytoca isolated from laboratory animals. *PLoS One,* 9**,** e100542.

DEHGHAN-NAYERI, N. & REZAEI TAVIRANI, M. 2015. THE INTERPRETATION OF PROTEIN STRUCTURE THROUGH RELATIONSHIP OF MELTING POINT (TM) AND ENTHALPY OF UNFOLDING (ΔHU). 41**,** 2278-246.

DEMAIN, A. L. & VAISHNAV, P. 2009. Production of recombinant proteins by microbes and higher organisms. *Biotechnol Adv,* 27**,** 297-306.

DENG, X., EICKHOLT, J. & CHENG, J. 2012. A comprehensive overview of computational protein disorder prediction methods. *Molecular BioSystems,* 8**,** 114-121.

DI, D. Y., JANG, J., UNNO, T. & HUR, H.-G. 2017. Emergence of Klebsiella variicola positive for NDM-9, a variant of New Delhi metallo-β-lactamase, in an urban river in South Korea. *Journal of Antimicrobial Chemotherapy,* 72**,** 1063-1067.

DI MARTINO, P., LIVRELLI, V., SIROT, D., JOLY, B. & DARFEUILLE-MICHAUD, A. 1996. A new fimbrial antigen harbored by CAZ-5/SHV-4-producing Klebsiella pneumoniae strains involved in nosocomial infections. *Infect Immun,* 64**,** 2266-73.

DI TOMMASO, P., MORETTI, S., XENARIOS, I., OROBITG, M., MONTANYOLA, A., CHANG, J.-M., TALY, J.-F. & NOTREDAME, C. 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic acids research,* 39**,** W13-W17.

DODS, R. 2020. *Concepts in Bioscience Engineering*, Springer International Publishing.

DOORDUIJN, D. J., ROOIJAKKERS, S. H., VAN SCHAIK, W. & BARDOEL, B. W. 2016. Complement resistance mechanisms of Klebsiella pneumoniae. *Immunobiology,* 221**,** 1102-1109.

DUONG-LY, K. C. & GABELLI, S. B. 2014. Using ion exchange chromatography to purify a recombinantly expressed protein. *Methods Enzymol,* 541**,** 95-103.

ELLIOTT, A. G., GANESAMOORTHY, D., COIN, L., COOPER, M. A. & CAO, M. D. 2016. Complete genome sequence of Klebsiella quasipneumoniae subsp. similipneumoniae strain ATCC 700603. *Genome announcements,* 4.

FETZER, C., KOROTKOV, V. & SIEBER, S. 2019. Hydantoin analogs inhibit the fully assembled ClpXP protease without affecting the individual peptidase and chaperone domains. *Organic & Biomolecular Chemistry,* 17.

FETZER, C., KOROTKOV, V. S., THÄNERT, R., LEE, K. M., NEUENSCHWANDER, M., VON KRIES, J. P., MEDINA, E. & SIEBER, S. A. 2017. A Chemical Disruptor of the ClpX Chaperone Complex Attenuates the Virulence of Multidrug-Resistant Staphylococcus aureus. *Angew Chem Int Ed Engl,* 56**,** 15746-15750.

FINLAY, B. B. & MCFADDEN, G. 2006. Anti-immunology: evasion of the host immune system by bacterial and viral pathogens. *Cell,* 124**,** 767-82.

FOUTS, D. E., TYLER, H. L., DEBOY, R. T., DAUGHERTY, S., REN, Q., BADGER, J. H., DURKIN, A. S., HUOT, H., SHRIVASTAVA, S. & KOTHARI, S. 2008. Complete genome sequence of the N 2-fixing broad host range endophyte Klebsiella pneumoniae 342 and virulence predictions verified in mice. *PLoS Genet,* 4**,** e1000141.

FREES, D., GERTH, U. & INGMER, H. 2014. Clp chaperones and proteases are central in stress survival, virulence and antibiotic resistance of Staphylococcus aureus. *Int J Med Microbiol,* 304**,** 142-9.

FREES, D., SAVIJOKI, K., VARMANEN, P. & INGMER, H. 2007. Clp ATPases and ClpP proteolytic complexes regulate vital biological processes in low GC, Gram-positive bacteria. *Mol Microbiol,* 63**,** 1285-95.

FRENCH, S. & ROBSON, B. 1983. What is a conservative substitution? *Journal of Molecular Evolution,* 19**,** 171-175.
GAO, H., LIU, Y., WANG, R., WANG, Q., JIN, L. & WANG, H. 2020. The transferability and evolution of NDM-1 and KPC-2 co-producing Klebsiella pneumoniae from clinical settings. *EBioMedicine,* 51**,** 102599.
GAO, W., KIM, J. Y., ANDERSON, J. R., AKOPIAN, T., HONG, S., JIN, Y. Y., KANDROR, O., KIM, J. W., LEE, I. A., LEE, S. Y., MCALPINE, J. B., MULUGETA, S., SUNOQROT, S., WANG, Y., YANG, S. H., YOON, T. M., GOLDBERG, A. L., PAULI, G. F., SUH, J. W., FRANZBLAU, S. G. & CHO, S. 2015. The cyclic peptide ecumicin targeting ClpC1 is active against Mycobacterium tuberculosis in vivo. *Antimicrob Agents Chemother,* 59**,** 880-9.
GASYMOV, O. & GLASGOW, B. 2007. ANS Fluorescence: Potential to Augment the Identification of the External Binding Sites of Proteins. *Biochimica et biophysica acta,* 1774**,** 403-11.
GAVRISH, E., SIT, C. S., CAO, S., KANDROR, O., SPOERING, A., PEOPLES, A., LING, L., FETTERMAN, A., HUGHES, D., BISSELL, A., TORREY, H., AKOPIAN, T., MUELLER, A., EPSTEIN, S., GOLDBERG, A., CLARDY, J. & LEWIS, K. 2014. Lassomycin, a ribosomally synthesized cyclic peptide, kills mycobacterium tuberculosis by targeting the ATP-dependent protease ClpC1P1P2. *Chem Biol,* 21**,** 509-518.
GONZALEZ-ESCALONA, N., THIRUNAVUKKARASU, N., SINGH, A., TORO, M., BROWN, E. W., ZINK, D., RUMMEL, A. & SHARMA, S. K. 2014. Draft genome sequence of bivalent Clostridium botulinum strain IBCA10-7060, encoding botulinum neurotoxin B and a new FA mosaic type. *Genome announcements,* 2.
GREENFIELD, N. J. 2006. Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols,* 1**,** 2876-2890.
HARK GAN, H., PERLOW, R. A., ROY, S., KO, J., WU, M., HUANG, J., YAN, S., NICOLETTA, A., VAFAI, J., SUN, D., WANG, L., NOAH, J. E., PASQUALI, S. & SCHLICK, T. 2002. Analysis of Protein Sequence/Structure Similarity Relationships. *Biophysical Journal,* 83**,** 2781-2791.
HERBST, R., GAST, K. & SECKLER, R. 1998. Folding of Firefly (Photinus pyralis) Luciferase: Aggregation and Reactivation of Unfolding Intermediates. *Biochemistry,* 37**,** 6586-6597.
HOLLINGSWORTH, S. A. & DROR, R. O. 2018. Molecular Dynamics Simulation for All. *Neuron,* 99**,** 1129-1143.
HOSPITAL, A., GOÑI, J. R., OROZCO, M. & GELPÍ, J. L. 2015. Molecular dynamics simulations: advances and applications. *Adv Appl Bioinform Chem,* 8**,** 37-47.
HOUDE, D. J. & BERKOWITZ, S. A. 2014. Biophysical Characterization of Proteins in Developing Biopharmaceuticals. 1-404.
HU, Y., WEI, L., FENG, Y., XIE, Y. & ZONG, Z. 2019. Klebsiella huaxiensis sp. nov., recovered from human urine. *International journal of systematic and evolutionary microbiology,* 69**,** 333-336.
HUA, X., CHEN, Q., LI, X., FENG, Y., RUAN, Z. & YU, Y. 2014. Complete genome sequence of Klebsiella pneumoniae sequence type 17, a multidrug-resistant strain isolated during tigecycline treatment. *Genome announcements,* 2.
HUANG, Y.-H., CHOU, S.-H., LIANG, S.-W., NI, C.-E., LIN, Y.-T., HUANG, Y.-W. & YANG, T.-C. 2018. Emergence of an XDR and carbapenemase-producing hypervirulent Klebsiella pneumoniae strain in Taiwan. *Journal of Antimicrobial Chemotherapy,* 73**,** 2039-2046.
INGMER, H., VOGENSEN, F. K., HAMMER, K. & KILSTRUP, M. 1999. Disruption and analysis of the clpB, clpC, and clpE genes in Lactococcus lactis: ClpE, a new Clp family in gram-positive bacteria. *J Bacteriol,* 181**,** 2075-83.
JOLLY, C. & MORIMOTO, R. I. 2000. Role of the Heat Shock Response and Molecular Chaperones in Oncogenesis and Cell Death. *JNCI: Journal of the National Cancer Institute,* 92**,** 1564-1572.
JØRGENSEN, S. B., BOJER, M. S., BOLL, E. J., MARTIN, Y., HELMERSEN, K., SKOGSTAD, M. & STRUVE, C. 2016. Heat-resistant, extended-spectrum β-lactamase-producing Klebsiella pneumoniae in endoscope-mediated outbreak. *Journal of Hospital Infection,* 93**,** 57-62.
KAR, N. P., SIKRIWAL, D., RATH, P., CHOUDHARY, R. K. & BATRA, J. K. 2008. Mycobacterium tuberculosis ClpC1: characterization and role of the N-terminal domain in its function. *Febs j,* 275**,** 6149-58.
KAZMAIER, U. & JUNK, L. 2021. Recent Developments on the Synthesis and Bioactivity of Ilamycins/Rufomycins and Cyclomarins, Marine Cyclopeptides That Demonstrate Anti-Malaria and Anti-Tuberculosis Activity. *Marine Drugs,* 19**,** 446.
KEDZIERSKA, S., AKOEV, V., BARNETT, M. E. & ZOLKIEWSKI, M. 2003. Structure and function of the middle domain of ClpB from Escherichia coli. *Biochemistry,* 42**,** 14242-8.
KELLEY, L. A., MEZULIS, S., YATES, C. M., WASS, M. N. & STERNBERG, M. J. E. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols,* 10**,** 845-858.

KING, T. L., SCHMIDT, S., THAKUR, S., FEDORKA-CRAY, P., KEELARA, S., HARDEN, L. & ESSACK, S. Y. 2021. Resistome of a carbapenemase-producing novel ST232 Klebsiella michiganensis isolate from urban hospital effluent in South Africa. *Journal of Global Antimicrobial Resistance,* 24**,** 321-324.

KO, J., PARK, H., HEO, L. & SEOK, C. 2012. GalaxyWEB server for protein structure prediction and refinement. *Nucleic acids research,* 40**,** W294-W297.

KO, W. C., PATERSON, D. L., SAGNIMENI, A. J., HANSEN, D. S., VON GOTTBERG, A., MOHAPATRA, S., CASELLAS, J. M., GOOSSENS, H., MULAZIMOGLU, L., TRENHOLME, G., KLUGMAN, K. P., MCCORMACK, J. G. & YU, V. L. 2002. Community-acquired Klebsiella pneumoniae bacteremia: global differences in clinical patterns. *Emerg Infect Dis,* 8**,** 160-6.

KRESS, W., MAGLICA, Ž. & WEBER-BAN, E. 2009. Clp chaperone–proteases: structure and function. *Research in Microbiology,* 160**,** 618-628.

KRUGER, N. 1994. The Bradford Method for Protein Quantitation. *Methods in molecular biology (Clifton, N.J.),* 32**,** 9-15.

KWON, T., YANG, J. W., LEE, S., YUN, M.-R., YOO, W. G., KIM, H. S., CHA, J.-O. & KIM, D.-W. 2016. Complete genome sequence of Klebsiella pneumoniae subsp. pneumoniae KP617, coproducing OXA-232 and NDM-1 carbapenemases, isolated in South Korea. *Genome announcements,* 4.

LABANA, P., DORNAN, M. H., LAFRENIÈRE, M., CZARNY, T. L., BROWN, E. D., PEZACKI, J. P. & BODDY, C. N. 2021. Armeniaspirols inhibit the AAA+ proteases ClpXP and ClpYQ leading to cell division arrest in Gram-positive bacteria. *Cell Chemical Biology,* 28**,** 1703-1715.e11.

LAEMMLI, U. K. 1970. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature,* 227**,** 680-685.

LASKOWSKI, R. A., MACARTHUR, M. W., MOSS, D. S. & THORNTON, J. M. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography,* 26**,** 283-291.

LATYPOV, R. F., LIU, D., GUNASEKARAN, K., HARVEY, T. S., RAZINKOV, V. I. & RAIBEKAS, A. A. 2008. Structural and thermodynamic effects of ANS binding to human interleukin-1 receptor antagonist. *Protein Sci,* 17**,** 652-63.

LÁZARO-PERONA, F., SOTILLO, A., TROYANO-HERNÁEZ, P., GÓMEZ-GIL, R., DE LA VEGA-BUENO, Á. & MINGORANCE, J. 2018. Genomic path to pandrug resistance in a clinical isolate of Klebsiella pneumoniae. *International journal of antimicrobial agents,* 52**,** 713-718.

LEE, S., SOWA, M. E., WATANABE, Y.-H., SIGLER, P. B., CHIU, W., YOSHIDA, M. & TSAI, F. T. 2003. The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state. *Cell,* 115**,** 229-240.

LETUNIC, I. & BORK, P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res,* 47**,** W256-w259.

LI, B., ZHAO, Y., LIU, C., CHEN, Z. & ZHOU, D. 2014. Molecular pathogenesis of Klebsiella pneumoniae. *Future Microbiol,* 9**,** 1071-81.

LI, L., YU, T., MA, Y., YANG, Z., WANG, W., SONG, X., SHEN, Y., GUO, T., KONG, J. & WANG, M. 2019. The genetic structures of an extensively drug resistant (XDR) Klebsiella pneumoniae and its plasmids. *Frontiers in cellular and infection microbiology,* 8**,** 446.

LIN, L., WEI, C., CHEN, M., WANG, H., LI, Y., LI, Y., YANG, L. & AN, Q. 2015. Complete genome sequence of endophytic nitrogen-fixing Klebsiella variicola strain DX120E. *Standards in genomic sciences,* 10**,** 1-7.

LO, J. H., BAKER, T. A. & SAUER, R. T. 2001. Characterization of the N-terminal repeat domain of Escherichia coli ClpA-A class I Clp/HSP100 ATPase. *Protein Sci,* 10**,** 551-9.

LOBANOV, M. Y., BOGATYREVA, N. S. & GALZITSKAYA, O. V. 2008. Radius of gyration as an indicator of protein structure compactness. *Molecular Biology,* 42**,** 623-628.

LOBLEY, A., WHITMORE, L. & WALLACE, B. A. 2002. DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics,* 18**,** 211-212.

LONG, S. W., LINSON, S. E., SAAVEDRA, M. O., CANTU, C., DAVIS, J. J., BRETTIN, T. & OLSEN, R. J. 2017a. Whole-genome sequencing of human clinical Klebsiella pneumoniae isolates reveals misidentification and misunderstandings of Klebsiella pneumoniae, Klebsiella variicola, and Klebsiella quasipneumoniae. *Msphere,* 2.

LONG, S. W., OLSEN, R. J., EAGAR, T. N., BERES, S. B., ZHAO, P., DAVIS, J. J., BRETTIN, T., XIA, F. & MUSSER, J. M. 2017b. Population genomic analysis of 1,777 extended-spectrum beta-lactamase-producing Klebsiella pneumoniae isolates, Houston, Texas: unexpected abundance of clonal group 307. *MBio,* 8.

LU, M. G.-X., JIANG, J., LIU, L., MA, A. P.-Y. & LEUNG, F. C.-C. 2015. Complete genome sequence of Klebsiella variicola strain HKUOPLA, a cellulose-degrading bacterium isolated from giant panda feces. *Genome announcements,* 3.

LUKE, K. A., HIGGINS, C. L. & WITTUNG-STAFSHEDE, P. 2007. Thermodynamic stability and folding of proteins from hyperthermophilic organisms. *The FEBS Journal,* 274**,** 4023-4033.

LYNCH, T., CHEN, L., PEIRANO, G., GREGSON, D. B., CHURCH, D. L., CONLY, J., KREISWIRTH, B. N. & PITOUT, J. D. 2016. Molecular evolution of a Klebsiella pneumoniae ST278 isolate harboring bla NDM-7 and involved in nosocomial transmission. *The Journal of infectious diseases,* 214**,** 798-806.

MA, D. Y., HUANG, H. Y., ZOU, H., WU, M. L., LIN, Q. X., LIU, B. & HUANG, S. F. 2020. Carbapenem-Resistant Klebsiella aerogenes Clinical Isolates from a Teaching Hospital in Southwestern China: Detailed Molecular Epidemiology, Resistance Determinants, Risk Factors and Clinical Outcomes. *Infect Drug Resist,* 13**,** 577-585.

MAILLARD, R. A., CHISTOL, G., SEN, M., RIGHINI, M., TAN, J., KAISER, C. M., HODGES, C., MARTIN, A. & BUSTAMANTE, C. 2011. ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell,* 145**,** 459-69.

MALHOTRA-KUMAR, S., XAVIER, B. B., DAS, A. J., LAMMENS, C., BUTAYE, P. & GOOSSENS, H. 2016. Colistin resistance gene mcr-1 harboured on a multidrug resistant plasmid. *The Lancet infectious diseases,* 16**,** 283-284.

MARSEE, J. D., RIDINGS, A., YU, T. & MILLER, J. M. 2018. Mycobacterium tuberculosis ClpC1 N-Terminal Domain Is Dispensable for Adaptor Protein-Dependent Allosteric Regulation. *Int J Mol Sci,* 19.

MARTIN, R. M. & BACHMAN, M. A. 2018. Colonization, infection, and the accessory genome of Klebsiella pneumoniae. *Frontiers in cellular and infection microbiology,* 8**,** 4.

MATHERS, A. J., CROOK, D., VAUGHAN, A., BARRY, K. E., VEGESANA, K., STOESSER, N., PARIKH, H. I., SEBRA, R., KOTAY, S., WALKER, A. S. & SHEPPARD, A. E. 2019. Klebsiella quasipneumoniae Provides a Window into Carbapenemase Gene Transfer, Plasmid Rearrangements, and Patient Interactions with the Hospital Environment. *Antimicrob Agents Chemother,* 63.

MATHERS, A. J., STOESSER, N., SHEPPARD, A. E., PANKHURST, L., GIESS, A., YEH, A. J., DIDELOT, X., TURNER, S. D., SEBRA, R. & KASARSKIS, A. 2015. Klebsiella pneumoniae carbapenemase (KPC)-producing K. pneumoniae at a single institution: insights into endemicity from whole-genome sequencing. *Antimicrobial agents and chemotherapy,* 59**,** 1656-1663.

MAURIZI, M. R. & XIA, D. 2004. Protein binding and disruption by Clp/Hsp100 chaperones. *Structure,* 12**,** 175-183.

MEDRANO, E. G., FORRAY, M. M. & BELL, A. A. 2014. Complete genome sequence of a Klebsiella pneumoniae strain isolated from a known cotton insect boll vector. *Genome announcements,* 2.

MESSNE, M., MANDL, M. M., HACKL, M. W., REINHARDT, T., ARDELT, M. A., SZCZEPANOWSKA, K., FRÄDRICH, J. E., WASCHKE, J., JEREMIAS, I., FUX, A., STAHL, M., VOLLMAR, A. M., SIEBER, S. A. & PACHMAYR, J. 2021. Small molecule inhibitors of the mitochondrial ClpXP protease possess cytostatic potential and re-sensitize chemo-resistant cancers. *Scientific Reports,* 11**,** 11185.

MÉSZÁROS, B., ERDŐS, G. & DOSZTÁNYI, Z. 2018. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Research,* 46**,** W329-W337.

MILES, A. J. & WALLACE, B. A. 2015. Chapter 6 - Circular Dichroism Spectroscopy for Protein Characterization: Biopharmaceutical Applications. *In:* HOUDE, D. J. & BERKOWITZ, S. A. (eds.) *Biophysical Characterization of Proteins in Developing Biopharmaceuticals.* Amsterdam: Elsevier.

MILLER, J. M., CHAUDHARY, H. & MARSEE, J. D. 2018. Phylogenetic analysis predicts structural divergence for proteobacterial ClpC proteins. *Journal of structural biology,* 201**,** 52-62.

MIOTTO, M., OLIMPIERI, P. P., DI RIENZO, L., AMBROSETTI, F., CORSI, P., LEPORE, R., TARTAGLIA, G. G. & MILANETTI, E. 2018. Insights on protein thermal stability: a graph representation of molecular interactions. *Bioinformatics,* 35**,** 2569-2577.

MOGK, A., SCHLIEKER, C., STRUB, C., RIST, W., WEIBEZAHN, J. & BUKAU, B. 2003. Roles of individual domains and conserved motifs of the AAA+ chaperone ClpB in oligomerization, ATP hydrolysis, and chaperone activity. *J Biol Chem,* 278**,** 17615-24.

MONTGOMERIE, J. Z. 1979. Epidemiology of Klebsiella and hospital-associated infections. *Rev Infect Dis,* 1**,** 736-53.

MULANI, M. S., KAMBLE, E. E., KUMKAR, S. N., TAWRE, M. S. & PARDESI, K. R. 2019. Emerging Strategies to Combat ESKAPE Pathogens in the Era of Antimicrobial Resistance: A Review. *Frontiers in Microbiology,* 10.

MULLINS, J. G. 2012. Structural modelling pipelines in next generation sequencing projects. *Advances in protein chemistry and structural biology,* 89**,** 117-167.

NAIR, S., MILOHANIC, E. & BERCHE, P. 2001. ClpC ATPase Is Required for Cell Adhesion and Invasion of Listeria monocytogenes. *Infection and immunity,* 68**,** 7061-8.

NEET, K. E. & LEE, J. C. 2002. Biophysical Characterization of Proteins in the Post-genomic Era of Proteomics *. *Molecular & Cellular Proteomics,* 1**,** 415-420.

NEUWALD, A. F., ARAVIND, L., SPOUGE, J. L. & KOONIN, E. V. 1999. AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. *Genome Res,* 9**,** 27-43.

NEWELL, P., FRICKER, A., ROCO, C., CHANDRANGSU, P. & MERKEL, S. 2013. A Small-Group Activity Introducing the Use and Interpretation of BLAST. *Journal of microbiology & biology education : JMBE,* 14**,** 238-43.

NGCOBO, N. S., CHILIZA, Z. E., CHEN, W., YU, J.-H., NELSON, D. R., TUSZYNSKI, J. A., PRETO, J. & SYED, K. 2020. Comparative Analysis, Structural Insights, and Substrate/Drug Interaction of CYP128A1 in Mycobacterium tuberculosis. *International journal of molecular sciences,* 21**,** 4816.

OSEI SEKYERE, J. & AMOAKO, D. G. 2017. Genomic and phenotypic characterisation of fluoroquinolone resistance mechanisms in Enterobacteriaceae in Durban, South Africa. *PLoS One,* 12**,** e0178888.

PACZOSA, M. K. & MECSAS, J. 2016. Klebsiella pneumoniae: Going on the Offense with a Strong Defense. *Microbiology and Molecular Biology Reviews,* 80**,** 629.

PALMGREN, M. G. 1990. An H+-ATPase assay: proton pumping and ATPase activity determined simultaneously in the same sample. *Plant physiology,* 94**,** 882-886.

PARK, S. S., KWON, H. Y., TRAN, T. D., CHOI, M. H., JUNG, S. H., LEE, S., BRILES, D. E. & RHEE, D. K. 2015. ClpL is a chaperone without auxiliary factors. *Febs j,* 282**,** 1352-67.

PATERSON, D. L. 2006. Resistance in gram-negative bacteria: Enterobacteriaceae. *Am J Infect Control,* 34**,** S20-8; discussion S64-73.

PATHAK, R., GUPTA, A., SHUKLA, R. & BAUNTHIYAL, M. 2018. Identification of new drug-like compounds from millets as Xanthine oxidoreductase inhibitors for treatment of Hyperuricemia: A molecular docking and simulation study. *Computational Biology and Chemistry,* 76.

PAUL, M., NARENDRAKUMAR, L., VASANTHAKUMARY, A. R., JOSEPH, I. & THOMAS, S. 2019. Genome sequence of a multidrug-resistant Klebsiella pneumoniae ST78 with high colistin resistance isolated from a patient in India. *Journal of global antimicrobial resistance,* 17**,** 187-188.

PAVLOPOULOS, G. A., SOLDATOS, T. G., BARBOSA-SILVA, A. & SCHNEIDER, R. 2010. A reference guide for tree analysis and visualization. *BioData Min,* 3**,** 1.

PENDLETON, J. N., GORMAN, S. P. & GILMORE, B. F. 2013. Clinical relevance of the ESKAPE pathogens. *Expert Rev Anti Infect Ther,* 11**,** 297-308.

PETTENGILL, E. A., HOFFMANN, M., BINET, R., ROBERTS, R. J., PAYNE, J., ALLARD, M., MICHELACCI, V., MINELLI, F. & MORABITO, S. 2015. Complete genome sequence of enteroinvasive Escherichia coli O96: H19 associated with a severe foodborne outbreak. *Genome announcements,* 3.

PHILIPPE, N., MAIGRE, L., SANTINI, S., PINET, E., CLAVERIE, J.-M., DAVIN-RÉGLI, A.-V., PAGÈS, J.-M. & MASI, M. 2015. In vivo evolution of bacterial resistance in two cases of Enterobacter aerogenes infections during treatment with imipenem. *PloS one,* 10**,** e0138828.

PIETROSIUK, A., LENHERR, E. D., FALK, S., BÖNEMANN, G., KOPP, J., ZENTGRAF, H., SINNING, I. & MOGK, A. 2011. Molecular basis for the unique role of the AAA+ chaperone ClpV in type VI protein secretion. *Journal of Biological Chemistry,* 286**,** 30010-30021.

PIRONE-DAVIES, C., CHEN, Y., PIGHTLING, A., RYAN, G., WANG, Y., YAO, K., HOFFMANN, M. & ALLARD, M. W. 2018. Genes significantly associated with lineage II food isolates of Listeria monocytogenes. *BMC genomics,* 19**,** 1-11.

PITT, M. E., ELLIOTT, A. G., CAO, M. D., GANESAMOORTHY, D., KARAISKOS, I., GIAMARELLOU, H., ABBOUD, C. S., BLASKOVICH, M. A., COOPER, M. A. & COIN, L. J. 2018. Multifactorial chromosomal variants regulate polymyxin resistance in extensively drug-resistant Klebsiella pneumoniae. *Microbial genomics,* 4.

PODSCHUN, R. & ULLMANN, U. 1998. Klebsiella spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. *Clin Microbiol Rev,* 11**,** 589-603.

RAJU, R. M., UNNIKRISHNAN, M., RUBIN, D. H., KRISHNAMOORTHY, V., KANDROR, O., AKOPIAN, T. N., GOLDBERG, A. L. & RUBIN, E. J. 2012. Mycobacterium tuberculosis ClpP1 and ClpP2 function together in protein degradation and are required for viability in vitro and during infection. *PLoS Pathog,* 8**,** e1002511.

RAMIREZ, M. S., XIE, G., TRAGLIA, G. M., JOHNSON, S. L., DAVENPORT, K. W., VAN DUIN, D., RAMAZANI, A., PEREZ, F., JACOBS, M. R. & SHERRATT, D. J. 2016. Whole-genome comparative analysis of two carbapenem-resistant ST-258 Klebsiella pneumoniae strains Isolated during a North-eastern Ohio outbreak: Differences within the high heterogeneity zones. Oxford University Press.

RAMSAMY, Y., ESSACK, S. Y., SARTORIUS, B., PATEL, M. & MLISANA, K. P. 2018. Antibiotic resistance trends of ESKAPE pathogens in Kwazulu-Natal, South Africa: A five-year retrospective analysis. *Afr J Lab Med,* 7**,** 887.

RIBET, D. & COSSART, P. 2015. How bacterial pathogens colonize their hosts and invade deeper tissues. *Microbes Infect,* 17**,** 173-83.

RODGER, A. 2013. Far UV Protein Circular Dichroism. *In:* ROBERTS, G. C. K. (ed.) *Encyclopedia of Biophysics.* Berlin, Heidelberg: Springer Berlin Heidelberg.

RODRÍGUEZ-MEDINA, N., BARRIOS-CAMACHO, H., DURAN-BEDOLLA, J. & GARZA-RAMOS, U. 2019. Klebsiella variicola: an emerging pathogen in humans. *Emerg Microbes Infect,* 8**,** 973-988.

ROSANO, G. L. & CECCARELLI, E. A. 2014. Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology,* 5.

ROSENZWEIG, R., FARBER, P., VELYVIS, A., RENNELLA, E., LATHAM, M. P. & KAY, L. E. 2015. ClpB N-terminal domain plays a regulatory role in protein disaggregation. *Proceedings of the National Academy of Sciences,* 112**,** E6872.

ROY, A., KUCUKURAL, A. & ZHANG, Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols,* 5**,** 725-738.

RUAN, Z., SUN, Q., JIA, H., HUANG, C., ZHOU, W., XIE, X. & ZHANG, J. 2019. Emergence of a ST2570 Klebsiella pneumoniae isolate carrying mcr-1 and blaCTX-M-14 recovered from a bloodstream infection in China. *Clinical Microbiology and Infection,* 25**,** 916-918.

SAFAVI, M., BOSTANSHIRIN, N., HAJIKHANI, B., YASLIANIFARD, S., VAN BELKUM, A., GOUDARZI, M., HASHEMI, A., DARBAN-SAROKHALIL, D. & DADASHI, M. 2020. Global genotype distribution of human clinical isolates of New Delhi metallo-β-lactamase-producing Klebsiella pneumoniae; A systematic review. *Journal of Global Antimicrobial Resistance,* 23**,** 420-429.

SAKAMOTO, N., AKEDA, Y., SUGAWARA, Y., TAKEUCHI, D., MOTOOKA, D., YAMAMOTO, N., LAOLERD, W., SANTANIRAND, P. & HAMADA, S. 2018. Genomic characterization of carbapenemase-producing Klebsiella pneumoniae with chromosomally carried blaNDM-1. *Antimicrobial agents and chemotherapy,* 62.

SANTAJIT, S. & INDRAWATTANA, N. 2016. Mechanisms of Antimicrobial Resistance in ESKAPE Pathogens. *Biomed Res Int,* 2016**,** 2475067.

SATLIN, M. J., CHEN, L., PATEL, G., GOMEZ-SIMMONDS, A., WESTON, G., KIM, A. C., SEO, S. K., ROSENTHAL, M. E., SPERBER, S. J. & JENKINS, S. G. 2017. Multicenter clinical and molecular epidemiological analysis of bacteremia due to carbapenem-resistant Enterobacteriaceae (CRE) in the CRE epicenter of the United States. *Antimicrobial agents and chemotherapy,* 61.

SAUER, R. T. & BAKER, T. A. 2011. AAA+ proteases: ATP-fueled machines of protein destruction. *Annu Rev Biochem,* 80**,** 587-612.

SCHIRMER, E. C., GLOVER, J. R., SINGER, M. A. & LINDQUIST, S. 1996. HSP100/Clp proteins: a common mechanism explains diverse functions. *Trends Biochem Sci,* 21**,** 289-96.

SCHÖNING-STIERAND, K., DIEDRICH, K., FÄHRROLFES, R., FLACHSENBERG, F., MEYDER, A., NITTINGER, E., STEINEGGER, R. & RAREY, M. 2020. ProteinsPlus: interactive analysis of protein–ligand binding interfaces. *Nucleic Acids Research,* 48**,** W48-W53.

SCHRÖDINGER 2021. Release 2021-2: Maestro. New York.

SCHRODINGER, L. Version 2010. The PyMOL molecular graphics system., 1, (5), 0.

SCHROLL, C., BARKEN, K. B., KROGFELT, K. A. & STRUVE, C. 2010. Role of type 1 and type 3 fimbriae in Klebsiella pneumoniae biofilm formation. *BMC Microbiology,* 10**,** 179.

SCHWEDE, T., KOPP, J., GUEX, N. & PEITSCH, M. C. 2003. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res,* 31**,** 3381-5.

SEKIZUKA, T., INAMINE, Y., SEGAWA, T., HASHINO, M., YATSU, K. & KURODA, M. 2019a. Potential KPC-2 carbapenemase reservoir of environmental Aeromonas hydrophila and Aeromonas caviae isolates from the effluent of an urban wastewater treatment plant in Japan. *Environmental microbiology reports,* 11**,** 589-597.

SEKIZUKA, T., INAMINE, Y., SEGAWA, T. & KURODA, M. 2019b. Characterization of NDM-5-and CTX-M-55-coproducing Escherichia coli GSH8M-2 isolated from the effluent of a wastewater treatment plant in Tokyo Bay. *Infection and drug resistance,* 12**,** 2243.

SEKIZUKA, T., YATSU, K., INAMINE, Y., SEGAWA, T., NISHIO, M., KISHI, N. & KURODA, M. 2018. Complete genome sequence of a blaKPC-2-positive Klebsiella pneumoniae strain isolated from the effluent of an urban sewage treatment plant in Japan. *Msphere,* 3.

SHAO, S., SUN, X., CHEN, Y., ZHAN, B. & ZHU, X. 2019. Complement Evasion: An Effective Strategy That Parasites Utilize to Survive in the Host. *Frontiers in Microbiology,* 10.

SHARMA, A. K., DHASMANA, N., DUBEY, N., KUMAR, N., GANGWAL, A., GUPTA, M. & SINGH, Y. 2017. Bacterial Virulence Factors: Secreted for Survival. *Indian J Microbiol,* 57**,** 1-10.

SHEPPARD, A. E., STOESSER, N., SEBRA, R., KASARSKIS, A., DEIKUS, G., ANSON, L., WALKER, A. S., PETO, T. E., CROOK, D. W. & MATHERS, A. J. 2016. Complete genome sequence of KPC-producing Klebsiella pneumoniae strain CAV1193. *Genome announcements,* 4.

SICHTIG, H., MINOGUE, T., YAN, Y., STEFAN, C., HALL, A., TALLON, L., SADZEWICZ, L., NADENDLA, S., KLIMKE, W. & HATCHER, E. 2019. FDA-ARGOS is a database with public quality-controlled reference genomes for diagnostic use and regulatory science. *Nature communications,* 10**,** 1-13.

SIEVERS, F., WILM, A., DINEEN, D., GIBSON, T. J., KARPLUS, K., LI, W., LOPEZ, R., MCWILLIAM, H., REMMERT, M. & SÖDING, J. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology,* 7**,** 539.

SINGER, J. R., BLOSSER, E. G., ZINDL, C. L., SILBERGER, D. J., CONLAN, S., LAUFER, V. A., DITORO, D., DEMING, C., KUMAR, R. & MORROW, C. D. 2019. Preventing dysbiosis of the neonatal mouse intestinal microbiome protects against late-onset sepsis. *Nature medicine,* 25**,** 1772-1782.

SMITH, C. 2005. Striving for purity: advances in protein purification. *Nature methods,* 2**,** 71-77.

SNIDER, J., THIBAULT, G. & HOURY, W. A. 2008. The AAA+ superfamily of functionally diverse proteins. *Genome Biol,* 9**,** 216.

SPRIESTERSBACH, A., KUBICEK, J., SCHÄFER, F., BLOCK, H. & MAERTENS, B. 2015. Purification of His-Tagged Proteins. *Methods Enzymol,* 559**,** 1-15.

STOESSER, N., GIESS, A., BATTY, E., SHEPPARD, A., WALKER, A., WILSON, D., DIDELOT, X., BASHIR, A., SEBRA, R. & KASARSKIS, A. 2014. Genome sequencing of an extended series of NDM-producing Klebsiella pneumoniae isolates from neonatal infections in a Nepali hospital characterizes the extent of community-versus hospital-associated transmission in an endemic setting. *Antimicrobial agents and chemotherapy,* 58**,** 7347-7357.

STRYER, L. 1968. Fluorescence Spectroscopy of Proteins. *Science,* 162**,** 526-533.

SUČEC, I., BERSCH, B. & SCHANDA, P. 2021. How do Chaperones Bind (Partly) Unfolded Client Proteins? *Front Mol Biosci,* 8**,** 762005.

SUGIKI, T., FUJIWARA, T. & KOJIMA, C. 2017. Cold-Shock Expression System in E. coli for Protein NMR Studies. *Methods Mol Biol,* 1586**,** 345-357.

THIBAULT, G., TSITRIN, Y., DAVIDSON, T., GRIBUN, A. & HOURY, W. A. 2006. Large nucleotide-dependent movement of the N-terminal domain of the ClpX chaperone. *Embo j,* 25**,** 3367-76.

TIWARI, G. & MOHANTY, D. 2013. An In Silico Analysis of the Binding Modes and Binding Affinities of Small Molecule Modulators of PDZ-Peptide Interactions. *PLOS ONE,* 8**,** e71340.

VAN DER WEIDE, H., TEN KATE, M. T., VERMEULEN-DE JONGH, D. M. C., VAN DER MEIJDEN, A., WIJMA, R. A., BOERS, S. A., VAN WESTREENEN, M., HAYS, J. P., GOESSENS, W. H. F. & BAKKER-WOUDENBERG, I. 2020. Successful High-Dosage Monotherapy of Tigecycline in a Multidrug-Resistant Klebsiella pneumoniae Pneumonia-Septicemia Model in Rats. *Antibiotics (Basel),* 9.

VAN DUIN, D., PEREZ, F., RUDIN, S. D., COBER, E., HANRAHAN, J., ZIEGLER, J., WEBBER, R., FOX, J., MASON, P. & RICHTER, S. S. 2014. Surveillance of carbapenem-resistant Klebsiella pneumoniae: tracking molecular epidemiology and outcomes through a regional network. *Antimicrobial agents and chemotherapy,* 58**,** 4035-4041.

VEDADI, M., ARROWSMITH, C. H., ALLALI-HASSANI, A., SENISTERRA, G. & WASNEY, G. A. 2010. Biophysical characterization of recombinant proteins: a key to higher structural genomics success. *Journal of structural biology,* 172**,** 107-119.

VERRIPS, C. T., GLAS, R. & KWAST, R. H. 1979. Heat resistance of Klebsiella pneumoniae in media with various sucrose concentrations. *European journal of applied microbiology and biotechnology,* 8**,** 299-308.

WALKER, J. E., SARASTE, M., RUNSWICK, M. J. & GAY, N. J. 1982. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *Embo j,* 1**,** 945-51.

WARETH, G. & NEUBAUER, H. 2021. The Animal-foods-environment interface of Klebsiella pneumoniae in Germany: an observational study on pathogenicity, resistance development and the current situation. *Veterinary Research,* 52**,** 16.

WEI, W., LI, A., YANG, J., MA, F., WU, D., XING, J., ZHOU, X. & ZHAO, D. 2015. Synergetic effects and flocculation behavior of anionic polyacrylamide and extracellular polymeric substrates extracted from Klebsiella sp. J1 on improving soluble cadmium removal. *Bioresource technology,* 175**,** 34-41.

WEINGARTEN, R. A., JOHNSON, R. C., CONLAN, S., RAMSBURG, A. M., DEKKER, J. P., LAU, A. F., KHIL, P., ODOM, R. T., DEMING, C. & PARK, M. 2018. Genomic analysis of hospital plumbing reveals diverse reservoir of bacterial plasmids conferring carbapenem resistance. *MBio,* 9.

WESEVICH, A., SUTTON, G., RUFFIN, F., PARK, L. P., FOUTS, D. E., FOWLER, V. G., THADEN, J. T. & LEDEBOER, N. A. 2020. Newly Named Klebsiella aerogenes (formerly Enterobacter aerogenes) Is Associated with Poor Clinical Outcomes Relative to Other Enterobacter Species in Patients with Bloodstream Infection. *Journal of Clinical Microbiology,* 58**,** e00582-20.

WHITFIELD, C. D., STEERS, E. J., JR. & WEISSBACH, H. 1970. Purification and properties of 5-methyltetrahydropteroyltriglutamate-homocysteine transmethylase. *J Biol Chem,* 245**,** 390-401.

WILLIAMS, C. J., HEADD, J. J., MORIARTY, N. W., PRISANT, M. G., VIDEAU, L. L., DEIS, L. N., VERMA, V., KEEDY, D. A., HINTZE, B. J., CHEN, V. B., JAIN, S., LEWIS, S. M., ARENDALL, W. B., 3RD, SNOEYINK, J., ADAMS, P. D., LOVELL, S. C., RICHARDSON, J. S. & RICHARDSON, D. C. 2018. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci,* 27**,** 293-315.

WLODAWER, A., MINOR, W., DAUTER, Z. & JASKOLSKI, M. 2008. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *Febs j,* 275**,** 1-21.

WOJTYRA, U. A., THIBAULT, G., TUITE, A. & HOURY, W. A. 2003. The N-terminal zinc binding domain of ClpX is a dimerization domain that modulates the chaperone function. *J Biol Chem,* 278**,** 48981-90.

WOLF, N. M., LEE, H., CHOULES, M. P., PAULI, G. F., PHANSALKAR, R., ANDERSON, J. R., GAO, W., REN, J., SANTARSIERO, B. D., LEE, H., CHENG, J., JIN, Y. Y., HO, N. A., DUC, N. M., SUH, J. W., ABAD-ZAPATERO, C. & CHO, S. 2019. High-Resolution Structure of ClpC1-Rufomycin and Ligand Binding Studies Provide a Framework to Design and Optimize Anti-Tuberculosis Leads. *ACS Infect Dis,* 5**,** 829-840.

WU, J. H., WU, A. M., TSAI, C. G., CHANG, X. Y., TSAI, S. F. & WU, T. S. 2008. Contribution of fucose-containing capsules in Klebsiella pneumoniae to bacterial virulence in mice. *Exp Biol Med (Maywood),* 233**,** 64-70.

WYPYCH, G. 2015. 11 - PRINCIPLES OF STABILIZATION. *In:* WYPYCH, G. (ed.) *PVC Degradation and Stabilization (Third Edition).* Boston: ChemTec Publishing.

WYRES, K. L., LAM, M. M. C. & HOLT, K. E. 2020. Population genomics of Klebsiella pneumoniae. *Nature Reviews Microbiology,* 18**,** 344-359.

XAVIER, B., LAMMENS, C., BUTAYE, P., GOOSSENS, H. & MALHOTRA-KUMAR, S. 2016. Complete sequence of an IncFII plasmid harbouring the colistin resistance gene mcr-1 isolated from Belgian pig farms. *Journal of Antimicrobial Chemotherapy,* 71**,** 2342-2344.

YANG, J., WEI, W., PI, S., MA, F., LI, A., WU, D. & XING, J. 2015a. Competitive adsorption of heavy metals by extracellular polymeric substances extracted from Klebsiella sp. J1. *Bioresource technology,* 196**,** 533-539.

YANG, J., YAN, R., ROY, A., XU, D., POISSON, J. & ZHANG, Y. 2015b. The I-TASSER Suite: protein structure and function prediction. *Nat Methods,* 12**,** 7-8.

YANG, L. Q., SANG, P., TAO, Y., FU, Y. X., ZHANG, K. Q., XIE, Y. H. & LIU, S. Q. 2014. Protein dynamics and motions in relation to their functions: several case studies and the underlying mechanisms. *J Biomol Struct Dyn,* 32**,** 372-93.

YE, J., OSBORNE, A. R., GROLL, M. & RAPOPORT, T. A. 2004. RecA-like motor ATPases--lessons from structures. *Biochim Biophys Acta,* 1659**,** 1-18.

ZHANG, Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics,* 9**,** 40.
ZHANG, Y., JIN, L., OUYANG, P., WANG, Q., WANG, R., WANG, J., GAO, H., WANG, X. & WANG, H. 2020. Evolution of hypervirulence in carbapenem-resistant Klebsiella pneumoniae in China: a multicentre, molecular epidemiological analysis. *Journal of Antimicrobial Chemotherapy,* 75**,** 327-336.
ZHAO, Y., ZENG, C. & MASSIAH, M. A. 2015. Molecular Dynamics Simulation Reveals Insights into the Mechanism of Unfolding by the A130T/V Mutations within the MID1 Zinc-Binding Bbox1 Domain. *PLOS ONE,* 10**,** e0124377.
ZHENG, B., HALPERIN, T., HRUSKOVA-HEIDINGSFELDOVA, O., ADAM, Z. & CLARKE, A. K. 2002. Characterization of chloroplast Clp proteins in Arabidopsis: localization, tissue specificity and stress responses. *Physiologia plantarum,* 114**,** 92-101.
ZOLKIEWSKI, M. 2006. A camel passes through the eye of a needle: protein unfolding activity of Clp ATPases. *Molecular Microbiology,* 61**,** 1094-1100.
ZOLKIEWSKI, M., ZHANG, T. & NAGY, M. 2012. Aggregate reactivation mediated by the Hsp100 chaperones. *Arch Biochem Biophys,* 520**,** 1-6.
ZUO, Z., FAN, H., GUO, J., ZHOU, W. & LI, L. 2012. Kinetics and Thermodynamics of 1-anilino-8-naphthalene Sulfonate Interactions with Urinary Trypsin Inhibitor. *The Protein Journal,* 31**,** 585-591.

Appendix A

Table A1: Klebsiella species and strains used for the genomic analysis of Clp ATPases. The species name, web-link and references were obtained from the NCBI genomes database for prokaryotes.

Strain	Web-link	References
K. aerogenes		
CAV1320	https://www.ncbi.nlm.nih.gov/genome/3417?genome_assembly_id=232549	(Mathers *et al.*, 2015, Sheppard *et al.*, 2016)
FDAARGOS_139	https://www.ncbi.nlm.nih.gov/genome/327?genome_assembly_id=363711	(Sichtig *et al.*, 2019)
G7	https://www.ncbi.nlm.nih.gov/genome/327?genome_assembly_id=271718	(Philippe *et al.*, 2015)
FDAARGOS_327	https://www.ncbi.nlm.nih.gov/genome/327?genome_assembly_id=1552	(Sichtig *et al.*, 2019)
FDAARGOS_152	https://www.ncbi.nlm.nih.gov/genome/327?genome_assembly_id=363028	(Sichtig *et al.*, 2019)
FDAARGOS_513	https://www.ncbi.nlm.nih.gov/genome/327?genome_assembly_id=3379	(Sichtig *et al.*, 2019)
FDAARGOS_363	https://www.ncbi.nlm.nih.gov/genome/327?genome_assembly_id=34818	(Sichtig *et al.*, 2019)
FDAARGOS_641	https://www.ncbi.nlm.nih.gov/genome/3417?genome_assembly_id=695478	(Sichtig *et al.*, 2019)
K. michiganensis		
FDAARGOS_647	https://www.ncbi.nlm.nih.gov/genome/33882?genome_assembly_id=695495	(Sichtig *et al.*, 2019)
K. oxytoca		
CAV1015	https://www.ncbi.nlm.nih.gov/genome/1165?genome_assembly_id=291700	(Mathers *et al.*, 2015, Sheppard *et al.*, 2016)
KONIH5	https://www.ncbi.nlm.nih.gov/genome/1165?genome_assembly_id=361932	(Weingarten *et al.*, 2018)
CAV1374	https://www.ncbi.nlm.nih.gov/genome/1165?genome_assembly_id=232487	(Mathers *et al.*, 2015, Sheppard *et al.*, 2016)
CAV1335	https://www.ncbi.nlm.nih.gov/genome/1165?genome_assembly_id=232486	(Mathers *et al.*, 2015, Sheppard *et al.*, 2016)
FDAARGOS_335	https://www.ncbi.nlm.nih.gov/genome/1165?genome_assembly_id=366590	(Sichtig *et al.*, 2019)
CAV1752	https://www.ncbi.nlm.nih.gov/genome/1165?genome_assembly_id=303753	(Mathers *et al.*, 2015, Sheppard *et al.*, 2016)
FDAARGOS_500	https://www.ncbi.nlm.nih.gov/genome/1165?genome_assembly_id=3369	(Sichtig *et al.*, 2019)

Table A1 continued...

K. pneumoniae		
2_GR_12	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=56801	(Elliott *et al.*, 2016, Pitt *et al.*, 2018)
WCHKP649	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=375331	(Hu *et al.*, 2019)
FDAARGOS_775	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=589914	(Sichtig *et al.*, 2019)
TVGHCRE225	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=380774	(Huang *et al.*, 2018)
BK13043	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=319029	(Long *et al.*, 2017b, Long *et al.*, 2017a)
KP64 DNA	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=16976	(Sakamoto *et al.*, 2018)
KPNIH48[#]	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=363608	(Weingarten *et al.*, 2018)
C2660	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=77054	(Gao *et al.*, 2020)
WCHKP020037	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=77722	(Hu *et al.*, 2019)
GSU10-3 DNA	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=17291	(Sekizuka *et al.*, 2019a, Sekizuka *et al.*, 2019b, Sekizuka *et al.*, 2018)
WCHKP2	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=382097	(Hu *et al.*, 2019)
C789	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=74186	(Zhang *et al.*, 2020)
Isolate blood sample 2	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=280776	(Babouee *et al.*, 2011)
WCHKP115068	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=65659	(Hu *et al.*, 2019)
WCHKP2080	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=65653	(Hu *et al.*, 2019)
WCHKP36	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=373304	(Hu *et al.*, 2019)
WCHKP020098	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=65657	(Hu *et al.*, 2019)
Xen39	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=786792	(Singer *et al.*, 2019)
WCHKP115069	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly id=27777	(Hu *et al.*, 2019)

Table A1 continued...

K. pneumoniae		
16_GR_13	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=4137	(Elliott *et al.*, 2016, Pitt *et al.*, 2018)
FDAARGOS_531	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=3364	(Sichtig *et al.*, 2019)
SCKP020003	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=10198	(Hu *et al.*, 2019)
KP617	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=249502	(Kwon *et al.*, 2016)
XH209[#]	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=212989	(Hua *et al.*, 2014)
CAV1217	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=296907	(Mathers *et al.*, 2015, Sheppard *et al.*, 2016)
1_GR_13	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=1578	(Elliott *et al.*, 2016, Pitt *et al.*, 2018)
121	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=493866	(Ruan *et al.*, 2019)
SCKP020135	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=476671	(Hu *et al.*, 2019)
WCHKP095845	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=13002	(Hu *et al.*, 2019)
2-1	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=19	(Li *et al.*, 2019)
KP33 DNA	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=16975	(Sakamoto *et al.*, 2018)
SCKP020049	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=20206	(Hu *et al.*, 2019)
FDAARGOS_630	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=6954	(Sichtig *et al.*, 2019)
FDAARGOS_156	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=36549	(Sichtig *et al.*, 2019)
CFSAN054111	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=370067	(Bjornsdottir-Butler *et al.*, 2015, Pirone-Davies *et al.*, 2018, Pettengill *et al.*, 2015, Gonzalez-Escalona *et al.*, 2014, Burall *et al.*, 2014)
J1	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=260523	(Wei *et al.*, 2015, Yang *et al.*, 2015a)
KpN01	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=27223	(Lynch *et al.*, 2016)

Table A1 continued...

K. pneumoniae		
FDAARGOS_566	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=3363	(Sichtig *et al.*, 2019)
Carbapenem-resistant blaNDM-1	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=205032	(Van Duin *et al.*, 2014)
WCHKP34	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=382098	(Hu *et al.*, 2019)
FDAARGOS_631	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=6953	(Sichtig *et al.*, 2019)
TA6363	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=771384	(Sichtig *et al.*, 2019)
CRKP I	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=521650	(Paul *et al.*, 2019)
PMK1	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=211364	(Stoesser *et al.*, 2014)
1050	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=10329	(Lázaro-Perona *et al.*, 2018)
KPNIH39	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=277271	(Conlan *et al.*, 2016)
KPN1482	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=319030	(Long *et al.*, 2017b, Long *et al.*, 2017a)
CAV1392	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=232380	(Sheppard *et al.*, 2016)
KPNIH45	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=468513	(Weingarten *et al.*, 2018)
CAV1344	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=232381	(Mathers *et al.*, 2015, Sheppard *et al.*, 2016)
CAV1042	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=296901	(Mathers *et al.*, 2015, Sheppard *et al.*, 2016)
C2414	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=770443	(Gao *et al.*, 2020)
39427	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=908959	(Satlin *et al.*, 2017)
CAV1217	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=296238	(Mathers *et al.*, 2015, Sheppard *et al.*, 2016)
CAV1453	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=296230	(Mathers *et al.*, 2015, Sheppard *et al.*, 2016)
FDAARGOS_439	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=343753	(Sichtig *et al.*, 2019)

Table A1 continued...

K. pneumoniae		
FDAARGOS_439	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=343751	(Sichtig *et al.*, 2019)
K pneumoniae* subsp. *pneumoniae		
KPNIH32	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=21262	(Conlan *et al.*, 2014)
WCHKP015093	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=65655	(Hu *et al.*, 2019)
TGH8	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=272815	(Malhotra-Kumar *et al.*, 2016, Ramirez *et al.*, 2016, Xavier *et al.*, 2016)
KPNIH31	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=213199	(Conlan *et al.*, 2014)
234-12	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=230300	(Becker *et al.*, 2015)
KPNIH30	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=213198	(Conlan *et al.*, 2014)
KPNIH29	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=213197	(Conlan *et al.*, 2014)
KPNIH33	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=212640	(Conlan *et al.*, 2014)
SCKP020079	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=377125	(Hu *et al.*, 2019)
SCKP020046	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=373305	(Hu *et al.*, 2019)
SCKP040074	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=377126	(Hu *et al.*, 2019)
ST:101960100	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=354684	(Osei Sekyere and Amoako, 2017)
TGH10[1]	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly_id=272816	(Malhotra-Kumar *et al.*, 2016, Ramirez *et al.*, 2016, Xavier *et al.*, 2016)
TGH13	https://www.ncbi.nlm.nih.gov/genome/815?genome_assembly id=283648	(Malhotra-Kumar *et al.*, 2016, Ramirez *et al.*, 2016, Xavier *et al.*, 2016)
K. quasipneumoniae		
ATCC 700603	https://www.ncbi.nlm.nih.gov/genome/38419?genome_assembly_id=294621	(Elliott *et al.*, 2016, Pitt *et al.*, 2018)
HKUOPLA	https://www.ncbi.nlm.nih.gov/genome/3829?genome_assembly_id=20687	(Lu *et al.*, 2015)
FDAARGOS_93	https://www.ncbi.nlm.nih.gov/genome/3829?genome_assembly id=36525	(Sichtig *et al.*, 2019)

Table A1 continued...

K. variicola		
WCHKP19	https://www.ncbi.nlm.nih.gov/genome/2121?genome_assembly_id=666503	(Lu *et al.*, 2015)
DX120E	https://www.ncbi.nlm.nih.gov/genome/2121?genome_assembly_id=280598	(Lin *et al.*, 2015)
FDAARGOS_627	https://www.ncbi.nlm.nih.gov/genome/2121?genome_assembly_id=695475	(Sichtig *et al.*, 2019)
GJ1	https://www.ncbi.nlm.nih.gov/genome/2121?genome_assembly_id=30584	(Di *et al.*, 2017)
GJ3	https://www.ncbi.nlm.nih.gov/genome/2121?genome_assembly_id=30586	(Di *et al.*, 2017)
KP5-1	https://www.ncbi.nlm.nih.gov/genome/2121?genome_assembly_id=20979	(Medrano *et al.*, 2014)
KPN1481	https://www.ncbi.nlm.nih.gov/genome/2121?genome_assembly_id=20630	(Long *et al.*, 2017b, Long *et al.*, 2017a)
342	https://www.ncbi.nlm.nih.gov/genome/2121?genome_assembly_id=410978	(Fouts *et al.*, 2008)
WCHKV030666	https://www.ncbi.nlm.nih.gov/genome/2121?genome_assembly_id=666504	(Hu *et al.*, 2019)

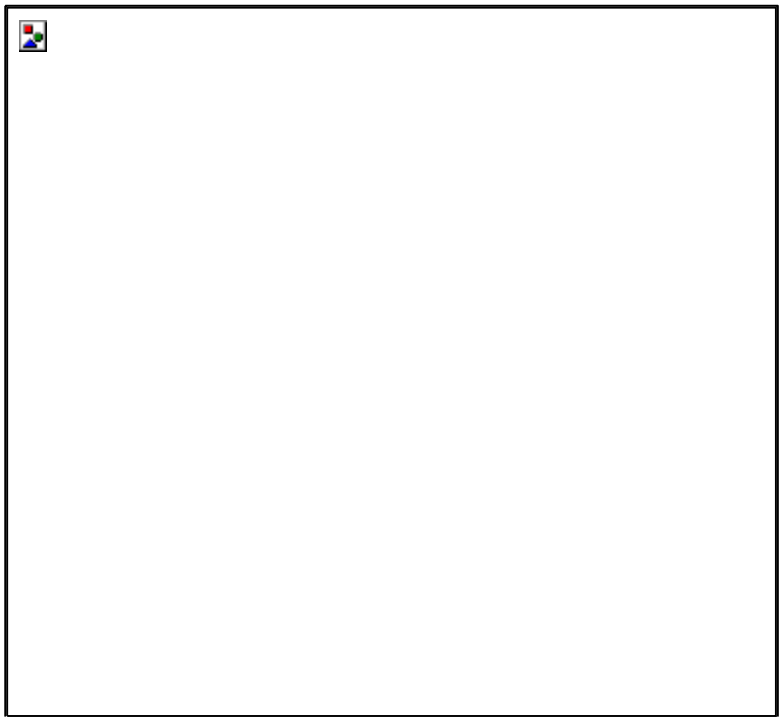


Figure A1: Primary amino acid sequence of ClpK. The sequence was obtained from GenBank (Uniprot: E0W6V3).