Identification and analysis of *Cryptosporidium* Glutathione Transferase

By

Sizamile Mbalenhle Mfeka

BSc. (Hons) Biochemistry

Submitted in fulfilment of the academic requirements for the degree of Master of Science in Biochemistry,

School of Life Sciences,

University of KwaZulu-Natal,

Pietermaritzburg.

2021

Supervised by Dr. Thandeka Khoza

Preface

The experimental work described in this dissertation was carried out in the School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, from January 2018 to December 2020, 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: Sizamile Mbalenhle Mfeka (213511702)

Signature: __  _____ Date: _____04 February 2021_____

As the candidate's supervisor I agree to the submission of this dissertation.

Supervisor: Dr. Thandeka Khoza

Signature: ____  _ Date: _____04 February 2021_____

Declaration - Plagiarism

I, Sizamile Mbalenhle Mfeka, declare that:

1. The research reported in this dissertation, except where otherwise indicated, is my original research.
2. This dissertation has not been submitted for any degree or examination at any other university.
3. The research reported does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
4. The dissertation contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
    a. Their words have been re-written but the general information attributed to them has been referenced.
    b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.
5. Lastly the contents of this dissertation do 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 references sections.

Student Name: Sizamile Mbalenhle Mfeka

Signature: _  _____ Date: _____04 February 2021_____

Declaration - Publications

Scientific Reports, Published: 23 November 2020

Comparative analyses and structural insights of new class glutathione transferases in *Cryptosporidium* species

Mbalenhle Sizamile Mfeka [1], José Martínez-Oyanedel [2], Wanping Chen [3], Ikechukwu Achilonu [4], Syed Khajamohiddin [5]*, Thandeka Khoza [1]*

[1] *Department of Biochemistry, School of Life Sciences, University of KwaZulu-Natal (Pietermaritzburg campus), Scottsville, 3209, KwaZulu-Natal, South Africa.*

[2] *Laboratorio de Biofísica Molecular, Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Biológicas, Universidad de Concepción, Barrio Universitario S/N, Casilla 160_C, Concepción, Chile.*

[3] *Department of Molecular Microbiology and Genetics, University of Göttingen, Göttingen 37077, Germany.*

[4] *Protein Structure-Function Research Unit, School of Molecular and Cell Biology, University of the Witwatersrand, Braamfontein, Johannesburg, South Africa.*

[5] *Department of Biochemistry and Microbiology, Faculty of Science and Agriculture, University of Zululand, KwaDlangezwa 3886, South Africa.*

* **Corresponding authors' email:** khajamohiddinsyed@gmail.com and khozat1@ukzn.ac.za

Author contributions:

K.S., I.A. and T.K. designed and conceptualized the study. K.S. and T.K. provided funding for the study. M.S.M conducted the study, analyzed results and wrote the manuscript. All authors reviewed and approved the manuscript. Manuscript has been adjusted for dissertation format.

Student: _____  ____ Date: _____04 February 2021_____

Supervisor: ____  __ Date: _____04 February 2021_____

International Journal of Molecular Sciences, prepared for review

Expression and Purification of Recombinant *Cryptosporidium parvum* Glutathione Transferase

Mbalenhle Sizamile Mfeka[1] and Thandeka Khoza [1*]

Department of Biochemistry, School of Life Sciences, University of KwaZulu-Natal (Pietermaritzburg campus), Scottsville, 3209, KwaZulu-Natal, South Africa; smmfeka850@outlook.com (M.S.M); khozat1@ukzn.ac.za (T.K.)

[*]**Corresponding authors email:** khozat1@ukzn.ac.za (T.K.)

Author contributions

T.K designed and conceptualized the study. T.K planned and funded the experiments and analysis. M.S.M conducted the study, analyzed results and wrote the manuscript.

Student: _____[redacted]_ Date: _____04 February 2021_____

Supervisor: ___[redacted] Date: _____04 February 2021_____

Abstract

Cryptosporidiosis, caused by *Cryptosporidium* spp. is a gastrointestinal disease which gives rise to severe life threatening complications in immunecompromised patients. The disease causing parasite has a proficient defense system against xenobiotic compounds and substances that renders the only drug designed to treat the gastroenteritis infection inefficient in immune compromised patients. This defense system includes a phase II enzyme called Glutathione Transferase (GST) which detoxifies a wide range of oxidant based substrates. The overexpression of this protein in multi drug resistant cases and its presence in multiple stages of the parasites life cycle highlights the parasites dependence and utilization of the GST protein thus making it a suitable therapeutic target. This study was then set out to determine characteristic features of *Cryptosporidium* GSTs in comparison to well studied GSTs using molecular biology and bioinformatics tools. A genome wide search was performed across multiple protein databases to mine the Cryptosporidium GST. The 15 *Cryptosporidium* spp. found to possess full length proteins were compared amongst themselves within the species and against other species using phylogenetic analyses. This led to the discovery of three novel classes of *Cryptosporidium* GST based on amino acid sequence identity. The classes were named Gamma, Psi and Vega GSTs. The GSTs varied in amino acid length, and secondary structure characteristics determined through homology modeling. In comparison to preexisting GSTs, the Psi and Vega class GSTs did not have the typical active site Tyr^7 found in most cytosolic GST, furthermore the Vega class GST also did not have the typical thioredoxin like fold conserved in the N-terminal region of all GSTs. The Gamma class GSTs were found to most resemble pre existing GSTs consisting of the typical thioredoxin fold and the active site Tyr^7 and thus selected for expression and purification studies. pET, pCOLD1 and pCOLDTF vectors were used to determine a suitable vector to facilitate the expression of a soluble gamma class GST in *Escherichia coli*. pCOLDTF which utilizes cold shock proteins at low temperatures and a chaperone called trigger factor assisted in the recombinant expression of the gamma class GST resulting in a protein with the monomer size of ~50 kDa, which is double that of existing GSTs. This is owed to by the N-terminal and C-terminal extensions that the protein possesses. The protein was purified to homogeneity using affinity chromatography and size exclusion chromatography. The resulting protein was found to be dimeric under native conditions.

Acknowledgements

First and foremost, I would like to thank my loving and caring family. My dad for teaching me to be curious and sincere about life. My mother, for the unflagging and unconditional support throughout studies and immeasurable patience throughout. I would like to thank my youngest brother Bandile Mfeka for his encouragement throughout my academic pursuits and tremendous help on the long and rather bumpy road of research.

I express my sincere gratitude to the National Research Foundation for funding this project.

My deepest gratitude to my supervisor, Dr Thandeka Khoza, for having allowed me to pursue my postgraduate studies in the Discipline of Biochemistry. Thank you for the excellent guidance throughout my studies and giving me the opportunity to work in an ever-growing magnificent laboratory. I would also like to thank her for the never-ending positive attitude, for patience, and for helping me to overcome many crises throughout my studies and the completion of this dissertation

I would also like to thank Professor Syed Khajamohiddin, Dr Ikechukwu Achilonu and all co-authors for sharing their expertise and affording me the chance 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. Ntobeko Zwane, Ntuthuko Zungu and Nomfundo Ntombela thank you for the hours we spent in the laboratory together. Our time spent together made my working environment extremely pleasant with our stimulating discussions and lightening the work load with much needed laughter.

To my loving and caring friends, Nosipho Nkwanyana, Samkelisiwe Ndlela and Thando Maseko, thank you is the last word I would say to express my appreciation and unbelievable support throughout this time.

Lastly I know all this was possible through Jesus Christ my Lord and Savior. The opportunity and unexplainable peace, patience and drive granted to me to finish this project is all attributable to him.

Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Appendix A

Appendix C

List of Tables

Chapter 1

Chapter 2

Chapter 3

Appendix A

Abbreviations

2 x YT	2 x yeast tryptone
°C	degrees Celsius
3D	three dimensional
µg	Microgram
µg/mL	microgram per milliliter
µL	microliter
µM	micromolar
x g	relative centrifugal force
α	alpha
A_{280}	absorbance at 280 nm
ASK1	apoptosis signal-regulated kinase
β	beta
BLAST	Basic Local Alignment Search Tool
bp	base pair
BUSCA	Bologna United subcellular component annotator
CA	California
CDC	Centres for Disease Control and Prevention
CDNB	1-chloro-2, 4-dinitrobenzene
CLIC	chloride intracellular channel
CmGST	*Cryptosporidium meleagridis* UKMEL1 GST3
*Cp*GST	*Cryptosporidium parvum* glutathione transferase
CspA	cold shock protein A
C-terminal	carboxyl terminal
DNA	deoxyribonuecleic acid
EDTA	ethylenediaminetetra-acetic acid
EMBL	European Molecular Biology Laboratory
FT	flow through
GPX	Glutathione peroxidase
GS	Oxididized glutathione
GSH	reduced glutathione

G-site	glutathione binding site
GST	glutathione transferase
h	hour/s
H-site	hydrophobic electrophilic binding site
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl-β-D-thio-galactoside
I-TASSER	iterative threading assembly refinement
JNK	c-Jun N-terminal kinase
JTT	Jones Taylor Thornton
kDa	kilodaltons
M	molar
MAP	mitogen activated protein
MAPEG	membrane associated proteins involved in eicosanoid and glutathione metabolism
min	minute/s
mg/mL	milligram per milliliter
mL	milliliter
mM	millimolar
MWM	molecular weight marker
ng/µL	nanograms per microliter
nm	nanomolar
NMR	nuclear magnetic resonance
NCBI	National Centre of Biotechnology Information
N-terminal	amino terminal
NTZ	nitazoxanide
p	pellet
PAGE	polyacrylamide gel electrophoresis
PAR-1	proteinase-activated receptor 1
PCR	polymerase chain reaction
PfGST	*Plasmodium falciparum* glutathione transferase
PGD_2	prostaglandin D2

PGH_2	prostaglandin H2
PHYRE	protein homology/analogy recognition engine
rpm	rotations per minute
RSA	Republic of South Africa
s	supernatant
SDS	sodium dodecyl sulphate
sec	seconds
soc	super optimal broth
spp	species
TEMED	N',N',N',N',- tetramethylethylenediamine
TMHMM	cellular localization data transmembrane helices hidden markov models
TNF	tumour necrosis factor
™	trade mark
TRAF	Tumor necrosis factor receptor associated factor
Tris	2-amino-2-(hydroxymethyl)-1.3-propandiol
UK	United Kingdom
USA	United States of America
UV	ultraviolet
v/v	volume per volume
WHO	World Health Organization
w/v	weight per volume

CHAPTER 1:

LITERATURE REVIEW

1.1. Global Burden of Diseases

Cryptosporidiosis is a zoonotic parasitic disease that is considered by the Centres for Disease Control and Prevention (CDC) as one of the most prevalent of its kind (Chen *et al*., 2002). Though it is not as salient as diseases of the "Big Three" namely Malaria, HIV/AIDS and tuberculosis, cryptosporidiosis is still regarded as one of the top five parasitic zoonoses which are a burden globally (Hotez *et al*., 2007, Hotez *et al*., 2014).

Cryptosporidiosis is considered a disease of the bottom billion, which affects the worlds most disadvantaged population (Hotez *et al*., 2009). However, it has often been neglected and overlooked by public health officials, researchers and funders (Bamaiyi and Redhuan, 2017, Hotez *et al*., 2009). The neglect of this disease has allowed for its propagation in rural areas and urban slums and is slowly allowing the disease to pose a possible pandemic risk (Feasey *et al*., 2009, King and Bertino, 2008). Furthermore, the zoonotic nature of the disease increases the threat it poses, as 60 % of infectious diseases that have been identified since 1940 have also been zoonotic (Morse *et al*., 2012). What additionally makes *Cryptosporidium* a great health burden is; (i) Cryptosporidiosis contributing alarming numbers to the global burden of disease with 8.37 million disability adjusted life years according to 2010 data (Hotez *et al*., 2014, Pisarski, 2019). (ii) The vast global distribution of the *Cryptosporidium* genus (Certad *et al*., 2017, Pisarski, 2019). (iii) The thick-walled oocysts ease of spread and propagation through rainwater run offs, floods and other water bodies (Jagai, 2009, Pisarski, 2019). (iv) Their evasiveness to water treatment through their advanced defense mechanism (Certad *et al*., 2017, Jagai *et al*., 2009). *Cryptosporidium* spp. had joined the neglected diseases initiative in 2004 to alleviate this burden. However, the World Health Organization (WHO) has been greatly critiqued for not including the parasite in the neglected tropical disease category despite the crippling effects it has on child development, pregnancy as well as agricultural productivity and the common links the disease has to poverty (Hotez *et al*., 2014, Pisarski, 2019, Thompson *et al*., 2016).

1.1.1. Gastroenteritis as a Burden of Disease

Cryptosporidiosis was initially heavily prevalent in the early 1970's infecting humans and various vertebrates including birds, reptiles, fish and mammals (Bird and Smith, 1980, Checkley *et al*., 2015, O'Donoghue, 1995). The sudden sprouting of infections during that time incited surveillance into the cause of this disease, and efforts into rapidly developing molecular tools to try treat this disease (Harhay *et al*., 2010). The inability to treat these cryptosporidiosis infections early and the deliberate disregard decades up to now allowed cryptosporidiosis to emerge as a life-threatening disease (Harhay *et al*., 2010, Checkley *et al*., 2015). Not particularly in developed countries but in underprivileged areas. Realizing the burden of disease brought about by *Cryptosporidium* spp. and the implications linking it to poverty lead to its inclusion in the Neglected Disease Initiative by the WHO (Savioli *et al*., 2006). Since then, the disease has been excluded from this list despite the growing evidence to its case. This exclusion was received negatively as cryptosporidiosis was stated to be the most insidious neglected disease urgently deserving of policy prioritization as a global burden. It is believed that this prioritization would aid in cryptosporidiosis intervention being recognized as a necessity (Pisarski, 2019).

Cryptosporidiosis is a gastrointestinal disease which is globally known to be responsible for inflammation of the digestive tract caused by parasitic infection (Merriman, 2013). This disease primarily infects the small intestines whilst also colonizing the lumen and the epithelial surface (Certad *et al*., 2017, Thompson *et al*., 2016). The pathogenesis occurs from interactions between the parasite byproducts which are serine and cysteine proteinases (Flynn and Buret, 2004). The membrane bound proteinases cause caspase 3-dependent killing of host cells leading to the breakage of the epithelial barrier as well as host inflammatory and immunological responses (Guk *et al*.,2003, Savioli *et al*., 2006). A cryptosporidiosis infection generally leads to enterocyte apoptosis associated with the disruption of tight junctional proteins (Certad *et al*., 2017). Additionally a *Cryptosporidium* infection induces the proteinases activation of the hosts gastrointestinal Proteinase-activated receptor 1 (PAR-1). Though the mechanism of activation is not understood, the implications of this is the modulation of apoptosis and the increased enterocyte permeability by PAR-1 (Chin *et al*., 2003, Yang *et al*., 2009).

The mode of *Cryptosporidium* infection is particularly vast, ranging from person to person/animal contact, through feces and through ingestion of contaminated water and food (Mead *et al*., 1999, Ryan *et al*., 2018). Upon infection, symptom manifestation takes about 2-14 days and is typically characterized with perfuse watery diarrhea, abdominal cramps, vomiting, nausea, and low-grade fever which all vary in severity according to the patients immunity (Merriman, 2013). The diarrhea from cryptosporidiosis can be identified by being yellow in colour with a soft to liquid consistency accompanied with a strong unfamiliar odour. In children however, the symptoms observed are mild to severe diarrhea which leads to dehydration, growth retardation, cognitive deficit and possible death (Squire and Ryan, 2017).

The symptoms of this disease are not life threatening in immunocompetent patients as they are self-limiting lasting up to three weeks (Certad *et al*., 2017, Feasey *et al*., 2012). In immunocompromised patients however, a cryptosporidiosis infection can devastate further as it may give way to other unwanted downstream diseases. An example of this is wasting syndrome, often observed in AIDS patients (Cama *et al*., 2007). The loss of epithelial barrier function induced by a gastroenteritis infection among other failures causes decreased absorptive surfaces, leading to malabsorption accompanied by Crohns disease and irritable bowel syndrome (Irvine and Marshall, 2000). Other immunocompromised patients could also present with jaundice, bile duct infections and even pancreatic infections (Kotloff *et al*., 2013).

Despite the relatively disturbing consequences cryptosporidiosis has on its patients, the highest incidences of this disease, is found in children younger than the age of five (Delahoy *et al*., 2018, Kotloff *et al*., 2013). The prevalence of the disease is considerably high in groups of all ages, however the levels found in immunocompromised younger children has been striking (Certad *et al*., 2017). Diarrhea however accounts for 10.5% (Striepen, 2013). A report from (Kotloff *et al*., 2013) undertaken from three sub-Saharan African sites and three south Asian sites, showed that *Cryptosporidium* was the second most prevalent parasite responsible for severe diarrhea resulting in cases of morbidity in children between 12-24 months old (Khan *et al*., 2019). It was also discovered that often the children who survived the infection beyond 12-23 months suffered reduced cognition, impaired immune responses and growth retardation in later years regardless of adequate nutrition after treatment (Delahoy *et al*., 2018, Guerrant *et al*., 1999). These statistics reveal that gastroenteritis by *Cryptosporidium* is endemic in the developing world and urgently needs to be addressed. Currently, of the four

diarrheal pathogens that contribute to child morbidity, namely *rotavirus, Shigella* bacteria, *enterotoxigenic Escherichia coli* and *Cryptosporidium*, *Cryptosporidium* is the only one that does not have fully effective drug treatment or vaccines developed (Guerrant *et al*., 1990, Walker *et al*., 2010). Whilst treatments and vaccines are available or being developed for the diarrheal disease caused by the before mentioned *rotavirus, Shigella* bacteria and *enterotoxigenic Escherichia coli* pathogens (Kotloff *et al*., 2013, Striepen, 2013).

1.2. *Cryptosporidium* spp.

There are multiple ways in which gastroenteritis is caused, however the most common form of human infection is vectored by the coccoidal *Cryptosporidium* spp. These species account for more than 90% of gastroenteritis contractions (Rossle and Latif, 2013). Organisms in this genus are classified in the phylum Apicomplexa under the class Conoidasida (Certad *et al*., 2017). Although classified under this phylum, Cryptosporidium lacks the phylum defining plasmid known as apicoplast as well as a mitochondria, crippling the haloparasite (Ryan and Hijjawi, 2015). For this reason, *Cryptosporidium* is heavily dependent on the invasion of the host, not only for the completion of its life cycle, but virtually all nutrients for a number of their metabolic functions (Certad *et al*., 2017, Miyamoto and Eckmann, 2015, Thompson *et al*., 2016).

Of the 31 *Cryptosporidium* spp, *C. parvum* and *C. hominis* are responsible for the highest level of clinically relevant infections worldwide (Ryan *et al*., 2016, Squire and Ryan, 2017). Between these two species *C. parvum* is also known for being zoonotic (Xiao and Feng, 2008). A threatening source for human infection is seen to come from animal husbandry through the release of large amounts of resistant oocysts in surface water and the environment (Ghoneim *et al*., 2017). This was observed in cases of cryptosporidiosis outbreaks in veterinary students, children who attended agricultural fairs and research technicians (Xiao and Feng, 2008). This not only causes a strain on human patients but animals as well.

1.2.1. Life Cycle

The life cycle of *Cryptosporidium* is monoxenous, involving both asexual and sexual multiplication (Squire and Ryan, 2017). The *Cryptosporidium* life cycle is divided into six developmental stages which are excystation, merogony, gametogony, fertilization and zygote development, formation of environmentally resistant oocyst walls and sporogony (Miyamoto and Eckmann, 2015, Rossle and Latif, 2013). The onset of a *Cryptosporidium* spp. infection is

through the ingestion of oocysts. These contain four sporozoites within a tough two layered wall which are released into the ilium and mark the beginning of the excystation stage (Rossle and Latif, 2013). Once released, the sporozoites infect the tissue on the superficial surface of the intestinal epithelium causing the destruction of the epithelial layer (Certad *et al*., 2017, Thompson *et al*., 2016).

The sporozoites attach to the luminal surface of the epithelial cells and through merogony differentiate asexually into spherical tropozoites producing two different types of meronts (Thompson *et al*., 2005, Miyamoto and Eckmann, 2015). Type I meronts form up to eight merozoites which then enter neighboring epithelial cells (Striepen, 2013, Hijjawi *et al*., 2004). These go on to either develop into type II meronts or complete another cycle of type I meronts undergoing another round of asexual multiplication as depicted in Figure 1.1. The type II meronts differentiate into either microgamonts or macrogamonts initiating sexual reproduction (Certad *et al*., 2017, Squire and Ryan, 2017). The fertilization between these gamonts results in the formation of a zygote which develops into oocytes containing four sporozoites. The two types of oocytes produced are thin walled oocytes which are excreted in faeces and infectious to the new host and thin walled oocytes which then recirculate in the intestinal tract causing auto infection (Certad *et al*., 2017, Miyamoto and Eckmann, 2015). Through this process the functioning of the intestinal barrier is disruted, weakening its absorption functionality and promoting the secretion of fluids, electrolytes and nutrients leading to malnutrition and the obvious watery diarrhea (Ryan and Hijjawi, 2015, Striepen, 2013).

Cryptosporidium parvum has proven tricky to treat due to the fact that the parasite has a solid defense mechanism in place. Firstly, when the sporozoites differentiate into tropozoites the parasite resides within a parasitophorous vacuole protecting the parasite from harsh environment of the hosts gastrointestinal tract while also allowing acquisition of nutrients from the host cell (Miyamoto and Eckmann, 2015, Ryan and Hijjawi, 2015, Xiao and Feng, 2008). To achieve this, the parasite rests on the apical end of the enterocyte separated from the host cytoplasm by the parasite and host membranes and the feeder organelle. The exact transportation of ions and nutrients to the parasite is not fully known. However, the existence and localization of ATP-binding cassette transporters suggest the existence of a portal of entry (LaGier *et al*., 2001). Secondly, the host cell membrane completely covers the parasite during

its epithelial growth phase. This makes the efficient use of drugs slightly difficult as they have to cross this at that stage of the growth cycle (Hijjawi *et al*., 2004, Miyamoto and Eckmann, 2015). Thirdly, the *C. parvum* oocytes are quite resilient when exposed to changes in the environment. They are resistant to chemical disinfection and not affected by chlorine, chloramines and chlorine dioxides permitted in drinking water treatments (Thompson *et al*., 2005). A potent chemical disinfectant for *C. parvum* is ozone however it runs the risk of being hazardous. The overexposure of ozone gives rise to the formation of high concentrations of by-products, some of which could be genotoxic (Certad *et al*., 2017, Ryan and Hijjawi, 2015).

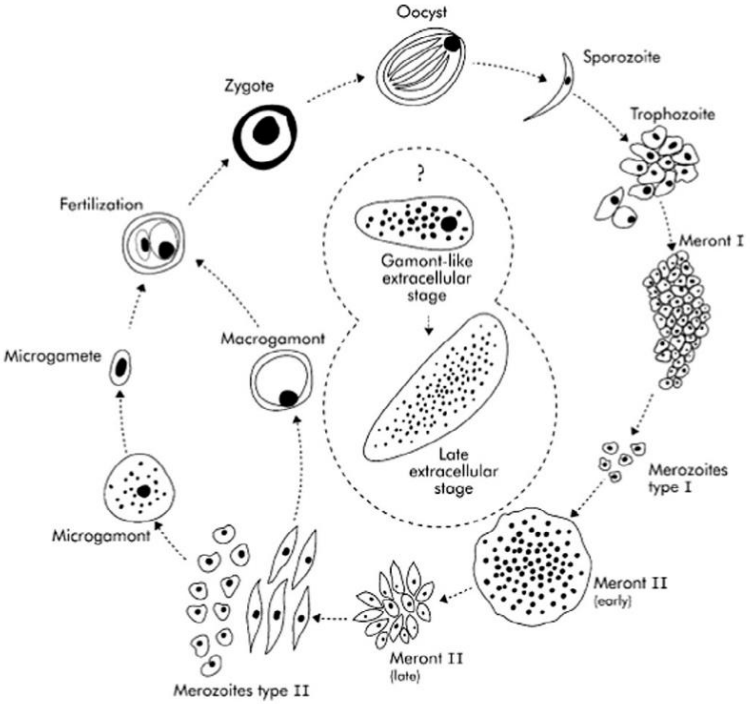


Figure 1.1: Life cycle of *Cryptosporidium* spp. taken from Hijjawi *et al* (2004).

The ingestion of thick-walled oocysts give rise to sporozoites, which initiate the infection. These invade epithelial cells to form trophozoites (Hijjawi *et al*., 2004). Growth occurs by asexual multiplication, leading to further cycles of infection and growth, or sexual multiplication involving gamonts and gametes, and leading to fertilized zygotes. These can differentiate into thin-walled oocysts that can initiate further rounds of autoinfection, or to thick-walled oocysts that are shed in the feces (Ryan and Hijjawi, 2015).

Due to the devastating and reoccurring nature of this disease, the treatment for cryptosporidiosis is highly necessary. However, the only efficient treatment available against this diseases is nitazoxanide (NTZ) (Amadi *et al*., 2002, Rossle and Latif, 2013). This drug is

relatively effective in immunocompetent patients but does not yield the same effects for immunocompromised patients (Gargala, 2008). Additionally, NTZ has not yet become generally available worldwide (Bamaiyi and Redhuan, 2017). Many alternative avenues are continuously being explored to neutralize the parasite. Unfortunately, these efforts have not yielded any successful results as yet, as other drugs being synthesized have shown to have the same effectiveness as NTZ (Amadi *et al*., 2002, Gargala, 2008, Sparks *et al*., 2015).

1.3. Oxidative Stress - Proteins

Cryptosporidium manages to evade a variety of internal and external stresses such as UV radiation, drugs, free radicals as well as the hosts immune response at various stages of its life cycle (Bajszár and Dekonenko, 2010, Certad *et al*., 2017, Ryan and Hijjawi, 2015). This is done similarly to how humans and other mammals use oxidative stress for the destruction of many xenobiotics and intracellular parasites (Xu *et al*., 2005). In parasitic protozoa, the antioxidant system is used for the prevention of oxidative stress and protection against oxidative killing by the hosts immune effector cells (Kang *et al*., 2013, Mannervik *et al*., 1985). In doing so, the proteins in this system additionally then serve to aid survival from the threat posed by other endogenously produced or xenobiotic compounds. This is achieved by the metabolism, detoxification and the expelling of xenobiotic compounds and drugs in a highly sophisticated process divided into three major parts (Croom, 2012, Xu *et al*., 2005).

The process is initiated by phase I enzymes listed in Table 1.1 which function to make compounds more hydrophilic while creating sites to allow phase II conjugation reactions to occur (Croom, 2012). This makes the xenobiotic compound a less toxic metabolite which is more water soluble (Sheehan *et al*., 2001). This is followed by the binding of phase III enzymes to the conjugated compound transporting them out (Xu *et al*., 2005). Through genome analysis, *C. parvum* has been revealed to code for a various number of these enzymes such as Glutathione Peroxidase (GPX), Glutathione-S-Transferase (GST) and superoxide dismutase. The existence of these enzymes better prepares *Cryptosporidium* for its parasitic life, enabling it to continuously stain its host (Miyamoto and Eckmann, 2015).

The presence of these detoxification enzymes coupled to *C. parvum* resistance to water chlorine disinfection raises *C. parvum* to be a health concern and a potential water borne bioterrorism agent (Khan *et al*., 2001). This highlights an urgency to control *C. parvum*. Alternative avenues are continuously being explored to neutralize the parasite. Multidrug-

resistance in *Cryptosporidium* spp. has been most associated with the overexpression of GST compared to the other detoxification enzymes that the parasite possesses (Tsavaris and Skopelitis, 2007). It is believed to be important through all stages of the parasite's life cycle. The GST enzyme is also vital in the parasites detoxification pathway and thus considered an appropriate molecular drug development target (Mauzy *et al*., 2012).

1.3.1. Glutathione Transferase Superfamily

Glutathione transferases (GST) are a family of multifunction proteins typically serving as binding proteins in different detoxification processes and cell metabolism (Frova, 2006). They belong to phase II group of enzymes along with GPX and thioredoxin as explained in Table 1.1 (Sheehan *et al*., 2001). GST catalyze reactions forming glutathione (GSH, γ-Glu-Cys-Gly) conjugates and the reduction of hydroperoxides forming oxidized glutathione's. These are formed through nucleophilic aromatic substitutions, Michael additions to α,β-unsaturated ketones and epoxide ring-opening reactions (Dixon and Edwards, 2010, Sheehan *et al*., 2001). Typically, the nonpolar compounds conjugated to GSH have electrophilic centers (carbon, nitrogen and sulfur) (Oakley, 2011). This enzyme superfamily is divided into three subfamilies based on their cellular locations (Oakley, 2005, Sheehan *et al*., 2001). This diversity makes the GST family well suited for cellular detoxification.

Table 1.1: Summary of drug metabolizing enzymes

Categories	Enzymes	Reactions	Functions	Reference
Phase I	Aldo-Keto reductase, Carboxylesterases, Cytochrome P450 monooxygenase, Epoxide hydrolase	oxidation, reduction, hydrolysis	Modify functional groups	Casida, 2018
Phase II	Glutathione-S-transferase, Glutathione peroxidase, UDP-glucuronosyltransferase, N-acetyltransferase, Sulfotransferase	conjugation	Increases hydrophilicity through conjugation with various peptides	Casida, 2018, Xu *et al*., 2005
Phase III	Multidrug associated protein, Organic anion transporting polypeptide 2, P-glycoprotein transporters	antiporter system/ process	Transport hydrophilic metabolites out of cells to eliminate biotransformed toxins.	Sheehan *et al*., 2001, Xu *et al*., 2005

The three subfamilies are mitochondrial GSTs, microsomal GSTs (or membrane bound) and cytosolic GSTs which are generally soluble (Frova, 2006). Other GSTs exists such as fosfomycin-resistance proteins FosA and dehydroascorbate reductases in plants which do not share the regular conjugative activities previously mentioned and are thus not classified into these subfamilies (Torres-Rivera and Landa, 2008). Within the subfamilies, GSTs are further divided into classes according to their amino acid sequence identities, with GSTs within each class sharing the same immunological cross-reactivity and specificity towards the electrophilic substrate and sensitivity to inhibitors (Sheehan *et al*., 2001, Salinas and Wong, 1999). Mitochondrial GSTs and were thought to be ancestors of cytosolic GSTs until it was proved that they instead evolved together and then diverged (Robinson *et al*., 2004). Microsomal GSTs contain four classes I, II, III and IV but are very poorly defined and classified (Knight *et*

al., 2008, Nebert and Vasiliou, 2004). Cytosolic GSTs are the most abundant and very well described. The classes are described as ubiquitous and organism specific with their functions described in Table 1.2 and Figure 1.2. In addition to these classes, a non-catalytic group of GSTs exist called Chloride intracellular channel (CLIC) proteins (Littler *et al*., 2010, Nebert and Vasiliou, 2004). Not much else can be added about these proteins but the fact that they exist as biologically active monomers, with poor conjugation activity but rather promotes the formation of mixed disulphides with glutathione (Frova, 2006). Although they are soluble, they can auto-insert into cellular membranes to form ion channels and undergo structural changes when transitioning from the cytosol to the membrane (Board and Menon, 2013, Goodchild *et al*., 2009).

Table 1.2: The functions of GST classes in addition to glutathione conjugation

GST Class	Cellular localization	General Information	Reference
Alpha	Cytosol	Found in a broad range of species. Involved in the biosynthesis of sex steroids and keto-steroid isomerase activity.	Deponte and Becker, 2005
Beta	Cytosol	Typically found in bacterial species. Known for conjugating antibiotics, assisting in antibiotic resistance to other organisms.	Shehu *et al*., 2019
CLIC	Cytosol	Found in a broad range of species. Enter intracellular membranes and form membrane channels.	Board and Menon, 2013
Delta and Epsilon	Cytosol	Typically found in insects. Thought to contribute to detoxication or antioxidative stress during development. Delta GSTs are also involved in oogenesis.	Scian *et al*., 2015, Udomsinprasert *et al*., 2005
Kappa	Mitochondrial	To date, found in primates and mice. Oligomerization of adiponectin.	Robinson *et al*., 2004
Lambda	Cytosol	Typically found in plants. Function is not yet known as they have no detectable GSH-conjugating activity.	Chronopoulou *et al*., 2017
MAPEG	Microsomal	Found in a broad range of species. Involved in production of leukotrienes and prostaglandin E and are mediators of inflammation.	Akil *et al*., 2012
Mu	Cytosol	Found in a broad range of species. Forms inhibitory complexes with ASK1, another member of the MAP kinase pathway.	Torres-Rivera and Landa, 2008
Omega	Cytosol	Found in a broad range of species. Catalyzes reduction and thioltransferase reactions.	Wu and Dong, 2012
Phi	Cytosol	Typically found in plants. Inhibits oxidative damage through the removal of endogenous cytotoxic hydroperoxides.	Munyampundu *et al*., 2016
Pi	Cytosol	Found in a broad range of species. Regulates JNK and TRAF signaling and catalyzes the S-glutathionylation reactions.	Prade *et al*., 1997
Sigma	Cytosol	Found in a broad range of species. Involved in prostaglandin synthesis by isomerization of PGH_2 – PGD_2.	Board and Menon, 2013
Tau	Cytosol	Typically found in plants. Involved in reactive oxygen species scavenging and improves plant chilling tolerance	Yang *et al*., 2016
Theta	Cytosol	Found in a broad range of species.	Shehu *et al*., 2019

		Has dichloromethane dehalogenase activity for the degradation of dichloromethane to obtain energy	
Xi	Cytosol	Typically found in bacteria, fungi, and archaea. Aids in extreme haloalkaphilic conditions.	Di Matteo *et al*., 2019
Zeta	Cytosol	Found in a broad range of species. Involved in isomerization of maleyacetoacetate to fumaracetoacetate in tyrosine degradation pathway and biotransformation of dichloroacetic acid to glyoxylate	Board *et al*., 1997

Symbol: *, Based on in silico prediction. Abbreviations: GSH, Glutathione; ASK1, Apoptosis signal-regulated kinase 1; MAP, Mitogen activated protein; JNK, c-Jun N-Terminal Kinase; TRAF, Tumor necrosis factor receptor (TNF)-associated factor; PGH2, Prostaglandin H2; PGD2, Prostaglandin D2.

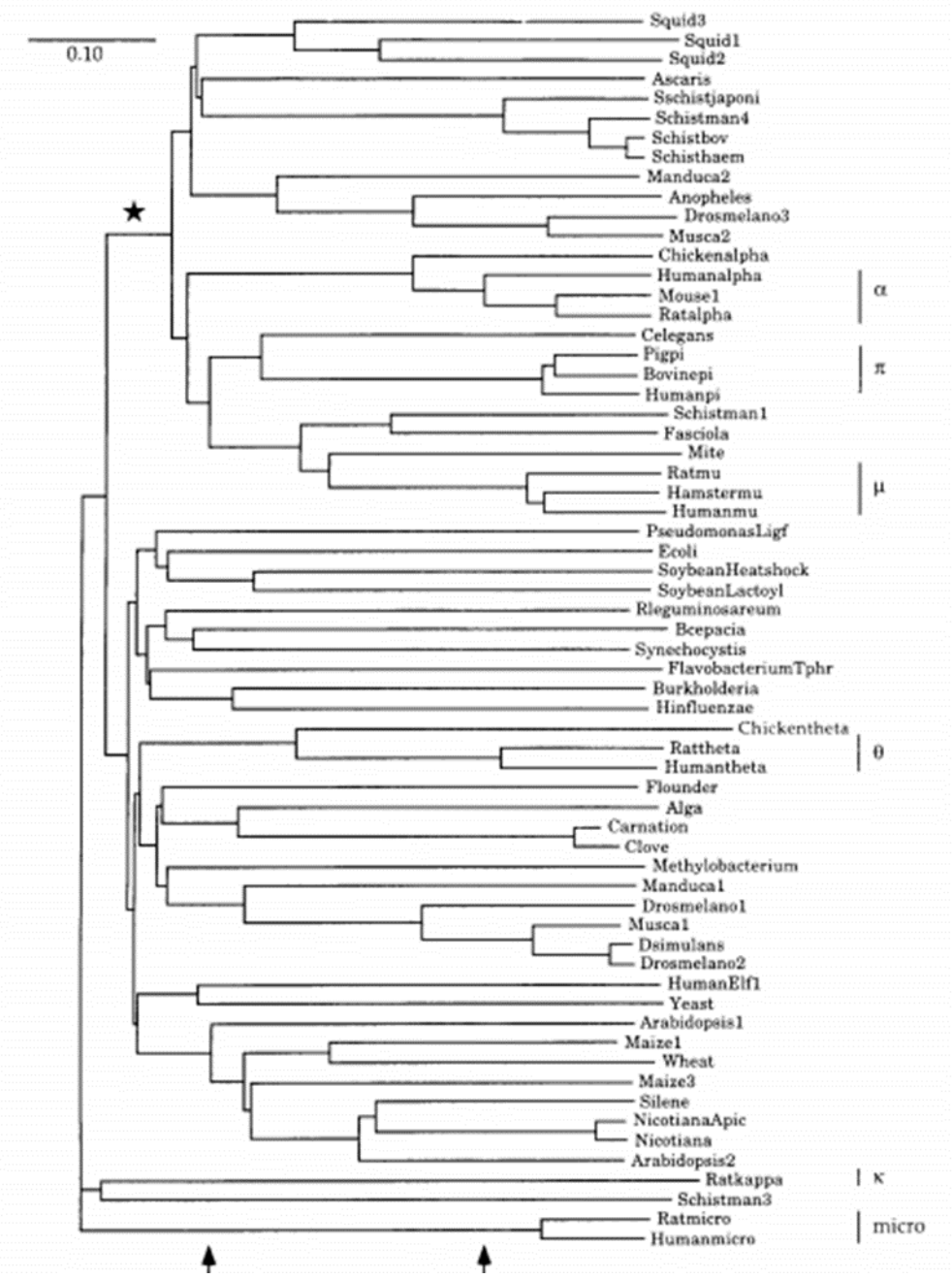


Figure 1.2: Phylogenetic tree of various GST classes compiled from different species taken from Snyder and Maddison (1997). Arrows at the bottom represent the boundaries of divergence that preserve the traditional classes of alpha, pi, mu, sigma, and microsomal GSTs. The star marks the strongly supported GST grouping.

1.3.2. Glutathione Transferase Structure

GSTs exist as homodimers or heterodimers. The composition of the dimer interface is different for each class with the dimerization process being highly specific occurring only between subunits within the same gene class (Torres-Rivera and Landa, 2008). GSTs do not form interclass heterodimers since the variant classes have different subunit interfaces

(Sheehan *et al*., 2001). Generally these dimer interfaces are stabilized by lock and key structures where the Met, Tyr or Phe side chain residue of one subunit packs into the hydrophobic pocket of the other subunit depicted in Figure 1.3 (Hegazy *et al*., 2004). In some cases, the "key" of one class does not line up with the "lock" of the other class resulting in the formation of unstable heterodimers (Board and Menon, 2013). Other stabilizing contacts observed between GST monomers are N-capping box which is specific patterns of hydrogen bonding and hydrophobic interactions situated in the ends of helices in the N-terminal domain (Aurora and Rosee, 1998).

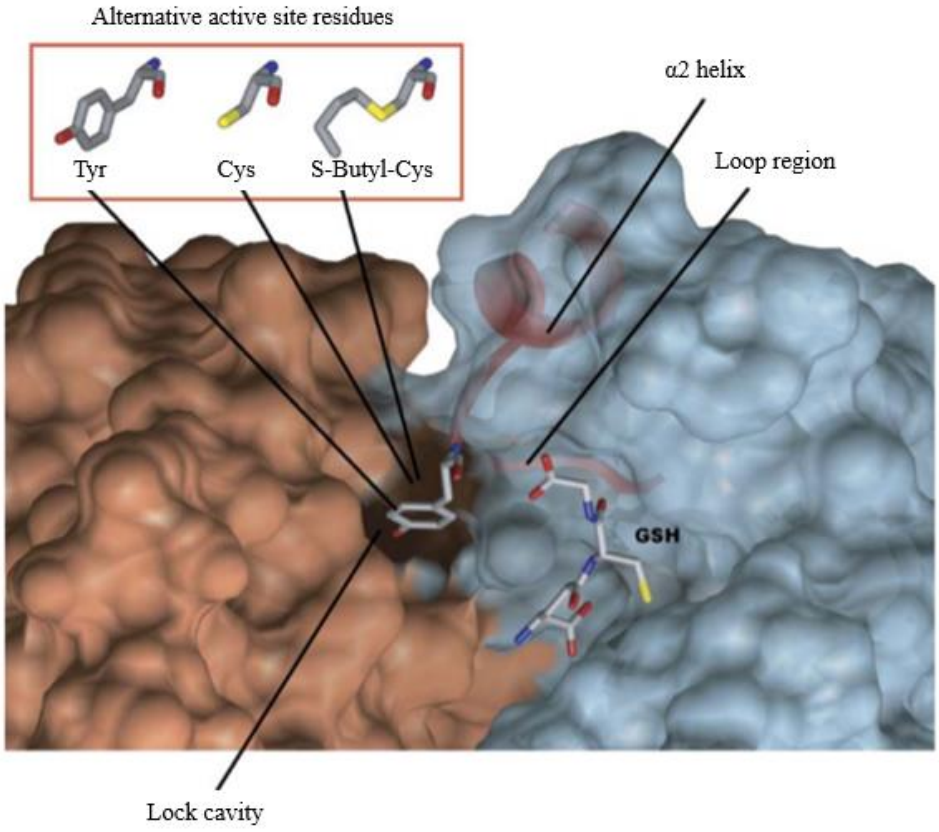


Figure 1.3: Glutathione binding site of homodimeric GST (6GSS) showing the lock and key cavity situated after α2 helix taken from Hegazy *et al* (2006). The active site tyrosine and GSH are shown in sticks.

Despite the proposed common ancestry, it has been found that sequence identity within classes is generally high at around 70% while the sequence identity between classes drops to as low as 10% (Sheehan *et al*., 2001). An example of this is the theta class of GSTs which contains a broad group of GST-like enzymes such as dehalogenases which results in low intra-class sequence identity (Rossjohn *et al*., 1997). In spite of the broad sequence identity their crystal structures still show a similar structural fold of the proteins, with differences observed

at the active site and the inter-subunit interface. The proteins in this family still have monomer sizes of 21-28 kDa averaging an amino acid sequence of ~ 220 (Frova, 2006, Torres-Rivera and Landa, 2008).

1.3.3. Conserved Thioredoxin like Fold

The GST monomers are divided into two domains namely an N-terminal domain and a C-terminal domain (Oakley, 2005). The N-terminal domain is responsible for GSH binding (Torres-Rivera and Landa, 2008, Sheehan *et al*., 2001). This domain assumes a topology resembling a thioredoxin fold which is generally conserved across all GST classes (Board and Menon, 2013, Sheehan *et al*., 2001). This conserved canonical fold is observed in several proteins despite the limited sequence identity (Oakley, 2011). The thioredoxin like fold is composed of three α helices flanking four β sheets as depicted in Figure 1.4 (Sheehan *et al*., 2001). The distinct N-terminal motif has a βαβ and ββα arrangement, which is linked by an α-helix (Atkinson and Babbitt, 2009). The N-terminus begins with a β_1 sheet followed by an α_1 helix leading to a second β_2 sheet parallel to β_1. An irregular loop region, known as the *cis*-Pro loop connects α_2 helix to β_3 (Sheehan *et al*., 2001). The ββα arrangement begins with two β sheets (β_3 and β_4) antiparallel to one another followed by an α_3 (Board and Menon, 2013). All four β sheets are in the same plane with α_1 and α_3 oriented below this plane and α_2 above it generally facing the solvent. A characteristic proline residue within the loop, connecting α_2 and β_3 is found in the least dominant cis conformation and is highly conserved in all GSTs (Atkinson and Babbitt, 2009, Wu and Dong, 2012). This cis-pro loop which is not involved in any catalysis is incredibly crucial for retaining the enzyme in a catalytically competent structure (Atkinson and Babbitt, 2009). The N-terminal domain is responsible for most of the glutathione binding site while also being connected to the second domain by a short linker sequence depicted in Figure 1.5.

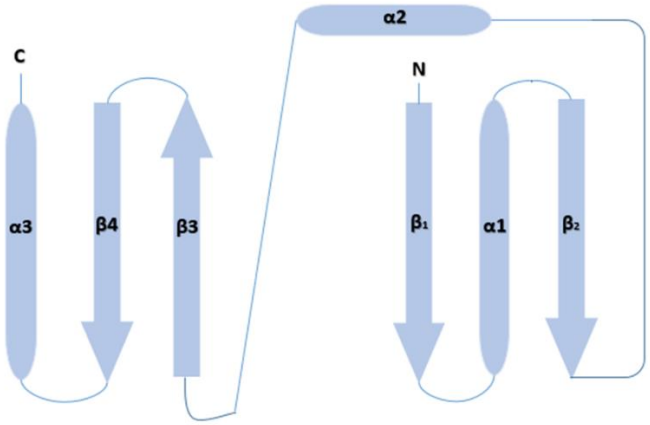


Figure 1.4: Schematic diagram representing the thioredoxin fold. The template structure to which the GST N-terminal domain follows.

The C-terminal domain is positioned downstream the thioredoxin like domain and is connected by a short liker sequence (Sheehan *et al.,* 2001). The C-terminal domain is composed exclusively of α helices. This domain binds the hydrophobic substrate (Atkinson and Babbitt, 2009, Wu and Dong, 2012). Depending on the class of transferases, the C-termianl domain can be made up of 4 to 8 α helices (Oakley, 2011, Sheehan *et al*., 2001). This end terminal is less conserved across all GST classes at a sequence and structural level with the helices varying in number, length, curvature and orientation (Nebert and Vasiliou, 2004, Sheehan *et al*., 2001, Strange *et al*., 2000). The amino acids of this domain largely contribute the residues that bind with the hydrophobic substrate in addition to providing an aspartic acid residue which is highly conserved, to the site where the tripeptide glutathione binds (Salinas and Wong, 1999). The variation in this region is said to account for the diverse range of substrates that are bound to be detoxified. For example the mu class of GST catalyse the detoxification of molecules containing oxiranes and α, β unsaturated carbonyl groups due to the variable C-terminal components while the alpha class GSTs act on 4-hydroxyalkenals and peroxides (Fritz-Wolf *et al*., 2003).

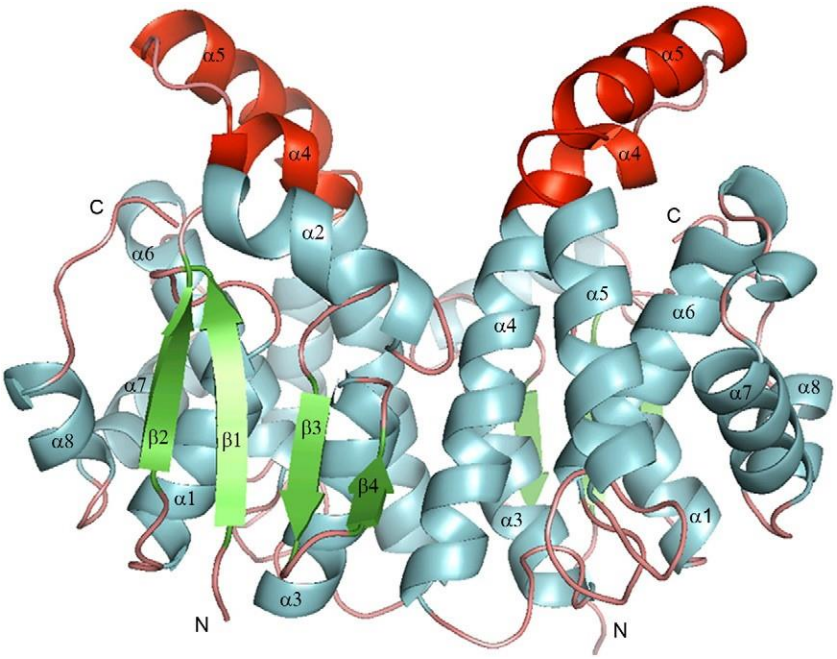


Figure 1.5: Homodimer of *Onchocerca volvulus* GST 1 showing secondary structure elements. Alpha helices are shown in blue and red, the beta sheets shown in green and the loops shown in pink taken from Perband *et al* (2008).

1.3.4. Active Site

GST subunits contain two ligand binding sites. These sites are named the glutathione binding site (G site) and the hydrophobic substrate binding site (H site) (Sheehan *et al*., 2001). The G site is highly conserved and is an essential feature of the enzymes catalytic mechanism (Dixon and Edwards, 2010, Sheehan *et al*., 2001). The H site is constructed of residues with non-polar site chains lying in the C-terminal domain (Mannervik *et al*., 1988). The two sites together constitute the catalytically active site which works autonomously to the other (Frova, 2006).

1.3.4.1. Glutathione Binding Site

The G site exclusively binds to tripeptide GSH and is thus highly conserved across the GST superfamily (Frova 2006). At the G site, the sulfur of GSH is activated for nucleophilic attack (Sheehan *et al*., 2001). The GSH, which runs antiparallel to the loop found after β_3 is bound in an extended conformation (Rossjohn *et al*., 1997). Here electrostatic interactions anchor the tripeptide to the domain typically through a network of hydrogen bonds running from β_3- β_4-α_2 (Oakley, 2011). GSH is bound and activated using specific residues of amino acids depending on the class of GST (Mannervik *et al*., 1988). The variation of residues responsible for GSH activation highlights the chemistry of reactions catalyzed by each isoenzyme (Frova,

2006). The amino acids that allow for conjugation or thiol transfer to occur are tyrosine (mu, pi, alpha, and sigma classes), serine (theta, zeta classes) or cysteine (omega, beta, lambda and CLIC class GSTs) (Prade *et al*., 1997,Torres-Rivera and Landa, 2008). The tyrosine and serine hydroxyl group acts as a hydrogen donors to the GSH thiol group. This leads to the formation and stabilization of a highly reactive thiolate anion, which is the target for nucleophilic attack of an electrophilic substrate (Frova, 2006). The cysteine however, promotes the formation of mixed disulfides GSH instead of forming thiolate anion. The contrast in these mechanisms results in the latter of the GST classes having poor GSH conjugation activities, but rather promoting redox reactions instead (Board and Menon, 2013).

1.3.4.2. Hydrophobic Substrate Binding Site

Positioned adjacent to the G site is the hydrophobic co-substrate binding site denoted the H site (Dixon and Edwards, 2010). The H site is mainly formed by residues from the C-terminal region and varies between classes in amino acid sequence and topology allowing for numerous substrate specificities (Oakley, 2005). Once GSH has bound to GST, the thiolate anion of the GSH is then stabilized enough to conjugate a wide range of electrophiles of different physicochemical properties, such as hydrophobicity, shape, and size (Board and Menon, 2013). The substrates bound in this site are typically held by residues within the H site through hydrophobic interactions.

1.3.5. Comparison Between Human and Parasitic Active Site

Analyzing GST structures belonging to parasites or rather species belonging to the apicomplexan phylum might give a closer estimation of what the *Cryptosporidium spp.* GST structures might resemble. Much like all other GST structures, parasitic GSTs are also enzymically active as homodimers (Hiller *et al*., 2006). Though packed differently, they too follow a thioredoxin like fold. Some GSTs such as *Plasmodium falciparum* GST (*Pf*GST) have a connecting loop between β_2 and α_2 (Fritz-Wolf *et al*., 2003). Each of these dimers are held together not only by hydrophobic forces but by four salt bridges and another four hydrogen bonded amino acid pairs (Fritz-Wolf *et al*., 2003, Hiller *et al*., 2006).

Focusing more into the G site region of the active site, it seems as though apicomplexa models share similar binding models to the well-known mu, pi and alpha classes of GST (Prade *et al*., 1997, Torres-Rivera and Landa, 2008). However, in *Pf*GST the GSH runs antiparallel to the conserved α_2-β_3 loop, adopting an extended conformation. In *Pf*GST models the G site is also

occupied by two formate molecules (Hiller *et al*., 2006). The formate molecules mimic the glutamyl carboxylate of GSH. These two formate molecules are stabilized by the before mentioned salt bridges. They are located in a position in the active center corresponding to the γ-carbonyl of GSH when compared to the μ, α, and π classes further illustrating the utilization of salt bridges by *Pf*GST. Additionally, in all classes, including *Pf*GST, the active site Tyr forms hydrogen bonds to the backbone N of Lys (Fritz-Wolf *et al*., 2003). At that position, only α class GST enzymes possess an Arg that stabilizes the Cys moiety of GSH with the help of the active site Try all using the N_e atom.

The H site is expected to be more variable than the G site to allow for diverse substrate binding (Torres-Rivera and Landa, 2008). In most GSTs, the loop joining β_1 and α_1, together with the C-terminal part of the helix α_4 and the residues after α_8 are what form the H site as illustrated in Figure 1.6 (Hiller *et al*., 2006). The H site is generally shielded by the C-terminal region from the surrounding solvent in various ways depending on the class (Frova, 2006). The α class uses its large α-9 present at the C-terminus while μ and π classes posses wall like structures. The μ class has an additional ~ 10 residues located between β_2 and α_1 called the μ loop which is also said to assist in reducing solvent accessibility (Board and Menon, 2013, Mannervik *et al*., 2005). *Pf*GST also has a μ loop but it is too short containing only five amino acids and thus cannot to form a wall, making the H site of *Pf*GST more solvent accessible (Torres-Rivera and Landa, 2008). This suggests that the H site of *Pf*GST allows for the binding of a broader range of substrates which include amphiphilic compounds as well (Sheehan *et al*., 2001).

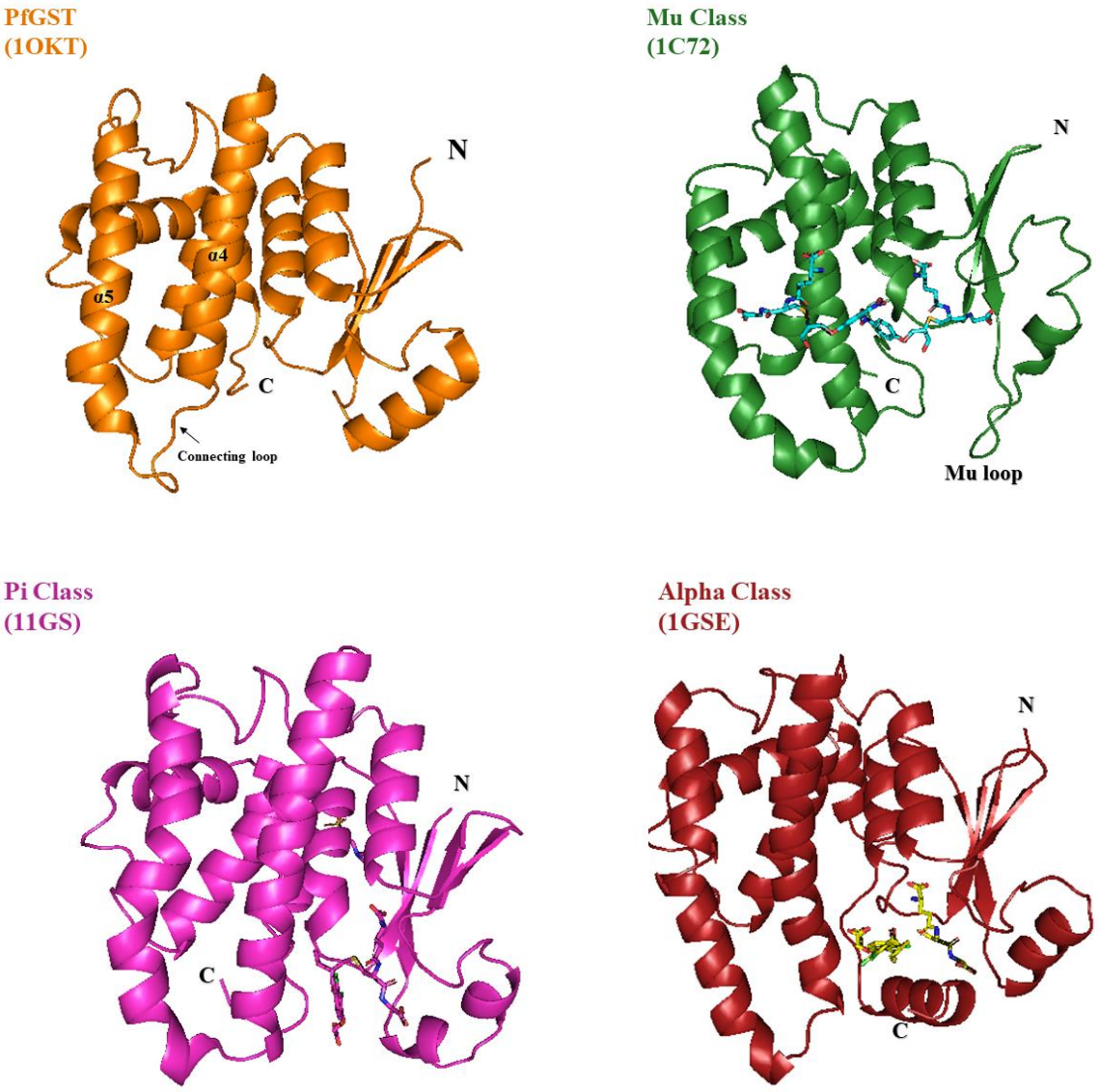


Figure 1.6: Structural comparison of GST enzymes. The mu class enzyme contains the GSH conjugate 1-hydroxy-2-S-glutathionyl-3-para-nitrophenoxy-propane. The alpha and pi class enzymes are shown in complex with the GSH conjugate of EA. In PfGST, the active-site region contains the connecting loop between α-4 and α-5 of the crystallographically independent monomer. All structures visualized using pymol.

1.4. GST Catalyzed Reaction

Although there is a generalized idea of the role of GST in *Cryptosporidium,* the extent at which it is dependent on its GST to avoid inactivation by endogenous and exogenous toxic compounds has not been well understood. For this reason, the biochemical mechanism cannot accurately be speculated without the structure of the enzyme, size and class of the enzyme being known. It is understood however that most xenobiotic compounds are

activated by cytochrome P450 resulting in the production of reactive metabolites which attack DNA and proteins, thus modifying them (Torres-Rivera and Landa, 2008). GSTs primarily function to catalyze reduced GSH conjugation of these toxic electrophilic compounds making them more soluble and easily expelled (Frova, 2006, Oakley, 2011). Glutathione conjugation reactions are the first step to the mercapturic acid pathway rated as an essential detoxification process producing the *N*-acetylcysteine derivative of its original xenobiotic substrate (Pickett and Lu, 1989, Sheehan *et al*., 2001, Wilce and Parker, 1994). The second step of this process is catalyzed by γ-glutamyltransferase. In addition to conjugation catalysis, some forms of this enzyme exhibit isomerase activity toward ketosteroids, glutathione peroxidase activity toward hydroperoxides from lipids and nucleic acids and they also act as carrier proteins (Atkinson and Babbitt, 2009, Sheehan *et al*., 2001).

Most GSTs exhibit conjugation activity toward 1-chloro-2, 4-dinitrobenzene (CDNB) and is used as a substrate for enzyme assays (Torres-Rivera and Landa, 2008). When conjugating CDNB most classes of GST unfold via a highly co-operative two-state pathway (Sheehan *et al*., 2001). Initially GST is activated at the G site where a GS anion is formed through the lowering of the G-site pKa allowing the deprotonation of the thiol group of cysteine in GSH (Wilce and Parker, 1994). This is accomplished by a range of amino acids either tyrosine, serine or cysteine governed by the class of GST. The deprotonation to the amino acids hydroxyl group lowers the pKa from 9.2 all the way down to 6.2 allowing the formation and stabilization of a nucleophilic thiolate anion of GSH facilitating the reaction depicted in Figure 1.7 (Atkins *et al*., 1993, Board and Menon, 2013). The Tyr and Ser hydroxyls are not the only components to stabilize the GSH thiolate as other studies have shown that the glutamyl α-carboxyl group of GSH may accept the thiol proton for stabilization purposes (Dourado *et al* 2008, Gustafsson *et al*., 2001). Additionally, when a Cys is used there is a formation of mixed disulphides with GSH due to the thiol transferase. Nucleophilic addition can then occur at the H site where the nucleophilic GS anion reacts with an aromatic electrophilic substrate (Tsuchida, 2000). The resulting substrate is more soluble prompting degradation of the product and excretion as mentioned (Wilce and Parker, 1994).

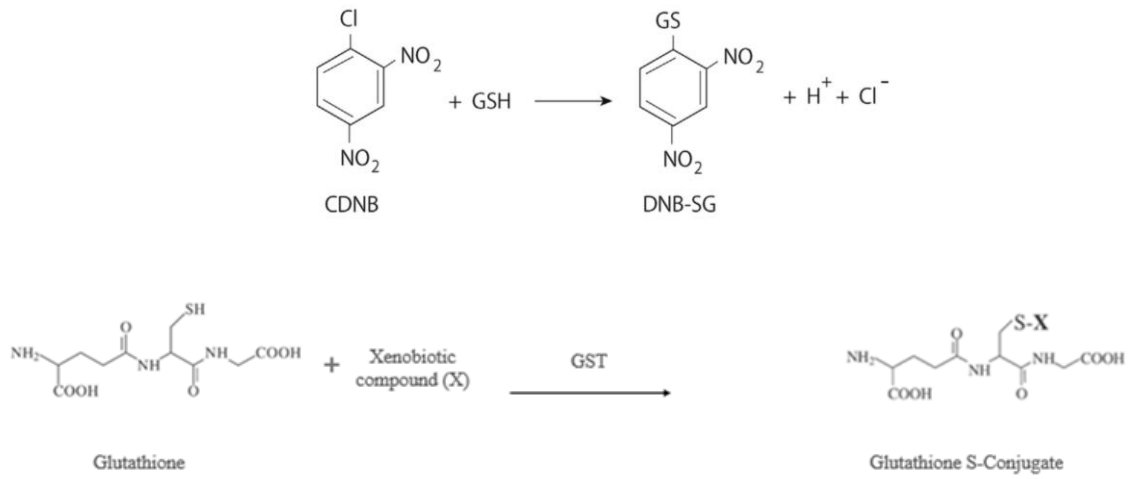


Figure 1.7: GSH conjugation to xenobiotic- and GST-catalysed reaction

It has been hypothesised that the GST catalysed reaction passes through a transition state which is an analogue of a Meisenheimer complex as seen in typical nucleophilic addition reactions (Prade *et al*., 1997). Kinetic studies have shown however that the formation of this transition step is the rate limiting step of GST catalysis (Prade *et al*., 1997).

In addition to the catalytic role that GSTs play in the conjugation and elimination of electrophilic compounds, these proteins have also been implicated in a range of alternative functions. In some cases, GSTs have been found to have the same catalytic activity as maleylacetoacetate isomerases. These enzymes catalyse the isomerization of maleylacetoacetate to fumarylacetoacetate which is an essential step in the tyrosine and phenylalanine degradation pathway (Gilge *et al*., 2008, Townsend *et al*., 2009). Other GSTs such as alpha class GSTs exhibit keto-steroid isomerase activity and glutathione peroxidase activity in addition to the wide range of glutathione conjugation reactions they catalyse (Benson *et al*., 1977). GSTs also play a considerable role in the regulation of cell signalling pathways (Board and Menon 2013). Pi class GSTs are powerful inhibitors of signalling molecules such as Jun Kinase, ASK1 and TRAF2 done so by conjugation independent protein-protein interactions (Adler *et al*., 1999, Elsby *et al*., 2003, Townsend *et al*., 2009). Other GSTs fall under the GST superfamily, but do not observe any significant enzymatic activity such as the CLIC proteins previously discussed in section 1.4.1. Another group of GSTs have also been recorded to being involved in the glutathionylation of protein cysteinyl thiols (Alder *et al*., 1999, Castro-Caldas *et al*., 2012). When these thiol groups are exposed on the surfaces of

proteins, they become susceptible to oxidative damage, thus being reversibly oxidized to sulfenic and sulfinic acid, targeting the protein for proteosomal degradation upon exposure to more stress (Cooper *et al*., 2011, Dalle-Donne *et al*., 2007). Glutathionylation, which is the formation of reversible disulphide bonds between protein thiols and glutathione, is considered a primary line of defence against oxidative stress (Cooper *et al*., 2011, Lock *et al*., 2011). As different GSTs emerge with varying structural motifs, sequence compositions and physiological properties, more functions of these enzymes emerge and are constantly being discovered, with GSTs no longer being limited to just electrophile conjugation alone.

1.5. GST Application

It is very apparent that GSTs bare vital significance in many parasites for detoxification purposes due to the lack of cytochrome P450 activity (Tsuchida, 2000). This statement is also supported by knockout studies in *P. falciparum* which showed the importance of functional GSTs in these organisms (Deponte and Becker, 2005). Additionally, GSTs and other phase II enzymes have been found to inhibit carcinogenesis as well as tumorigenesis through their detoxification activities in many studies (Devadoss *et al*., 2018, Sheehan *et al*., 2001, Zheng *et al*., 1992). These facts place them as targets for the development of vaccines or chemotherapeutic agents.

The biological functions that GSTs play in various roles are crucial and have a huge following in biomedical, biotechnological and nanotechnological sectors as displayed in Table 1.3. The biotechnological techniques already laid out have greatly increased our knowledge of 3D structures, functions and the evolutionary roles of the enzyme. This knowledge base can be used for the bioengineering and transgenesis of novel biomolecules of medical and industrial importance through forced evolution (Frova, 2006). The proteins are considered highly adaptable and thus can be catalytically enhanced for the creation of biocatalyst and drugs (Frova, 2006).

Table 1.3: Various glutathione transferase applications

Application	Description	Reference
Biosensor	In GSTs from maize fiber-optic portable biosensor were constructed to identify a pesticide-like atrazine.	Andreou and Clonis, 2002
	GSTs from soy were used to construct biosensor for the identify chloroacetanilide herbicide alachor.	Fragoulaki *et al*., 2007
	Optical biosensors which are sensitive for the detection of carcinogen captan in water supplies constructed from GST immobilized gel fillm.	Choi *et al*., 2003
Bioassay	Mosquito GSTs were used to construct specific enzyme assay for the determination of DDT.	Morou *et al*., 2008
Biomarker	Pi class GSTs were used to generate immunohistological markers for gastric cancer and gliomas.	Fan *et al*., 1995
Drug and Pro-drug Design	Alpha class GSTs were used for the production of drugs to treat steroid hormone dependent diseases such as cancer.	Johansson and Mannervik, 2001
	Pi class GSTs are used for Pro-drug activation such as the actiation of Telcyta.	Morgan *et al*., 1996
Bioremediation	Phi and Tau class GSTs from plats were used for the detoxification of herbicides such as chloroacetanilide, thiocarbamate and aryloxyphenoxypropionate.	Axarli *et al*., 2009
Transgenesis	GST genes from maize were used to transgenicially engineer tobacco with high tolerance to the herbicide alachlor.	Karavangeli *et al*., 2005

There is no doubt that the existence and contribution of GSTs to the metabolism and neutralization of drugs and other xenobiotics is an otherwise advantageous system in plants, animals and humans. It is however catastrophic when utilized by disease causing bacteria and parasites (Torres-Rivera and Landa, 2008). Not only do they cause the accumulation of multidrug resistant strains of bacteria but they are also responsible for chemotherapy resistance within some tumour cells (Oakley, 2005, Sheehan *et al*., 2001). For this reason, the urgency to develop inhibitors to enhance the therapeutic efficiency of Nitazoxanide or terminate cryptosporidiosis all together has never been more imperative.

Before inhibitors can be designed, the exact mechanism taken by the *Cryptosporidium*'s GST needs to be understood together with its biophysical characteristics. It is unfortunate however that there is a great deal unknown about the *Cryptosporidium* GSTs. Various studies would be needed to determine the GST classification, three-dimensional (3D) structure of the enzyme, general size, and oligomerization patterns of the protein. Through structure

determination, ligand-protein interactions can be maximised to design a chemical test suited to inhibit the enzyme.

Hindering the enzymes defence mechanism against the host and environmental reactive species could be the best possible option in reducing the catastrophic effects that have unfolded due to *Cryptosporidium* infections in immune-compromised patients. In an experiment conducted by Fritz-Wolf *et al* (2003). The isolation and characterisation of *Pf*GST which was present in all intra-erythrocyte stages of the parasite was sought. The crystal structure of the GST was determined so to design an inhibitor to reduce number of resistant parasites against the presently available drugs (Hiller *et al*., 2006). *Plasmodium falciparum* GST could not be assigned to any pre-existing class of GST indicating an ever growing class of GSTs (Deponte and Becker, 2005). The same mode of action can be taken for the determination of *Cryptosporidium* spp. GST so to determine the structure of the protein for possible drug design.

1.6. Aim of Study

Currently very little is known about the molecular and biophysical properties of *Cryptosporidium* GST. To this end the aim of this study was to use bioinformatic and molecular biology tools to fill this knowledge gap through elucidating a 3D structure of the parasitic protein for a comparative analysis with well studied GSTs. The secondary aim was to obtain a pure recombinant GST protein expressed using an appropriate expression vector to lead way for further characterization steps.

To achieve these aims the following objectives were set out:

- Genome data mining of all *Cryptosporidium* GSTs
- Phylogenetic analysis and identification of *Cryptosporidium* GST classes
- Homology modelling of *Cryptosporidium* GST
- Recombinant expression of soluble *Cryptosporidium* GST
- Purification of *Cryptosporidium* GST to homogeneity

1.7. References

ADLER, V., YIN, Z., FUCHS, S. Y., BENEZRA, M., ROSARIO, L., TEW, K. D., PINCUS, M. R., SARDANA, M., HENDERSON, C. J. & WOLF, C. R. 1999. Regulation of JNK signaling by GSTp. *The EMBO journal,* 18**,** 1321-1334.

AKIL, F., AKIL, H., LUTFIE, A., WIBOWO, W. S., MISKAD, U. & YUSUF, I. 2012. The role of xenobotic metabolism MGST1 gene polymorphism in colorectal cancer patients. *Acta Med Indones,* 44**,** 284-289.

AMADI, B., MWIYA, M., MUSUKU, J., WATUKA, A., SIANONGO, S., AYOUB, A. & KELLY, P. 2002. Effect of nitazoxanide on morbidity and mortality in Zambian children with cryptosporidiosis: a randomised controlled trial. *The Lancet,* 360**,** 1375-1380.

ANDREOU, V. G. & CLONIS, Y. D. 2002. Novel fiber-optic biosensor based on immobilized glutathione S-transferase and sol–gel entrapped bromcresol green for the determination of atrazine. *Analytica Chimica Acta,* 460**,** 151-161.

ATKINS, W., WANG, R., BIRD, A., NEWTON, D. & LU, A. 1993. The catalytic mechanism of glutathione S-transferase (GST). Spectroscopic determination of the pKa of Tyr-9 in rat alpha 1-1 GST. *Journal of Biological Chemistry,* 268**,** 19188-19191.

ATKINSON, H. J. & BABBITT, P. C. 2009. Glutathione transferases are structural and functional outliers in the thioredoxin fold. *Biochemistry,* 48**,** 11108-11116.

AURORA, R. & ROSEE, G. D. 1998. Helix capping. *Protein Science,* 7**,** 21-38.

AXARLI, I., LABROU, N., PETROU, C., RASSIAS, N., CORDOPATIS, P. & CLONIS, Y. 2009. Sulphonamide-based bombesin prodrug analogues for glutathione transferase, useful in targeted cancer chemotherapy. *European journal of medicinal chemistry,* 44.

BAJSZÁR, G. & DEKONENKO, A. 2010. Stress-induced Hsp70 gene expression and inactivation of *Cryptosporidium parvum* oocysts by chlorine-based oxidants. *Applied Environmental Microbiology,* 76**,** 1732-1739.

BAMAIYI, P. H. & REDHUAN, N. E. M. 2017. Prevalence and risk factors for cryptosporidiosis: a global, emerging, neglected zoonosis. *Asian Biomedicine,* 10**,** 309-325.

BENSON, A. M., TALALAY, P., KEEN, J. H. & JAKOBY, W. B. 1977. Relationship between the soluble glutathione-dependent delta 5-3-ketosteroid isomerase and the glutathione S-transferases of the liver. *Proceedings of the National Academy of Sciences,* 74**,** 158-162.

BIRD, R. & SMITH, M. 1980. Cryptosporidiosis in man: parasite life cycle and fine structural pathology. *The Journal of pathology,* 132**,** 217-233.

BOARD, G. P., BAKER, T. R., CHELVANAYAGAM, G. & JERMIIN, S. L. 1997. Zeta, a novel class of glutathione transferases in a range of species from plants to humans. *Biochemical Journal,* 328**,** 929-935.

BOARD, P. G. & MENON, D. 2013. Glutathione transferases, regulators of cellular metabolism and physiology. *Biochimica et biophysica acta (bba)-general subjects,* 1830**,** 3267-3288.

CAMA, V. A., ROSS, J. M., CRAWFORD, S., KAWAI, V., CHAVEZ-VALDEZ, R., VARGAS, D., VIVAR, A., TICONA, E., NAVINCOPA, M. & WILLIAMSON, J. 2007. Differences in clinical manifestations among *Cryptosporidium* species and subtypes in HIV-infected persons. *The Journal of Infectious Diseases,* 196**,** 684-691.

CASIDA, J. E. 2018. Pesticide detox by design. *Journal of Agricultural and Food Chemistry,* 66**,** 9379-9383.

CASTRO-CALDAS, M., CARVALHO, A. N., RODRIGUES, E., HENDERSON, C., WOLF, C. R. & GAMA, M. J. 2012. Glutathione S-transferase pi mediates MPTP-induced c-Jun N-terminal kinase activation in the nigrostriatal pathway. *Molecular Neurobiology,* 45**,** 466-477.

CERTAD, G., VISCOGLIOSI, E., CHABÉ, M. & CACCIÒ, S. M. 2017. Pathogenic mechanisms of *Cryptosporidium* and *Giardia*. *Trends in Parasitology,* 33**,** 561-576.

CHECKLEY, W., WHITE JR, A. C., JAGANATH, D., ARROWOOD, M. J., CHALMERS, R. M., CHEN, X.-M., FAYER, R., GRIFFITHS, J. K., GUERRANT, R. L. & HEDSTROM, L. 2015. A review of the global

burden, novel diagnostics, therapeutics, and vaccine targets for *Cryptosporidium*. *The Lancet Infectious Diseases,* 15**,** 85-94.

CHEN, B., PENG, W., XU, J., FENG, J., DONG, C. & XU, P. 2017. Genomic analysis of glutathione S-transferases (GST) family in common carp: identification, phylogeny and expression. *Pakistan Journal of Zoology,* 49.

CHOI, J.-W., KIM, Y.-K., SONG, S.-Y., LEE, I.-H. & LEE, W. H. 2003. Optical biosensor consisting of glutathione-S-transferase for detection of captan. *Biosensors and Bioelectronics,* 18**,** 1461-1466.

CHRONOPOULOU, E., ATAYA, F. S., POULIOU, F., PERPEROPOULOU, F., GEORGAKIS, N., NIANIOU-OBEIDAT, I., MADESIS, P., IOANNOU, E. & LABROU, N. E. 2017. Structure, evolution and functional roles of plant glutathione transferases. *Glutathione in Plant Growth, Development, and Stress Tolerance.* Springer, 195-213.

COOPER, A. J., PINTO, J. T. & CALLERY, P. S. 2011. Reversible and irreversible protein glutathionylation: biological and clinical aspects. *Expert opinion on drug metabolism & toxicology,* 7**,** 891-910.

CROOM, E. 2012. Metabolism of xenobiotics of human environments. *Progress in Molecular Biology and Translational Science.* Elsevier, 112. 31-88.

DELAHOY, M. J., OMORE, R., AYERS, T. L., SCHILLING, K. A., BLACKSTOCK, A. J., OCHIENG, J. B., MOKE, F., JARON, P., AWUOR, A. & OKONJI, C. 2018. Clinical, environmental, and behavioral characteristics associated with *Cryptosporidium* infection among children with moderate-to-severe diarrhea in rural western Kenya, 2008–2012: The Global Enteric Multicenter Study (GEMS). *PLoS Neglected Tropical Diseases,* 12**,** e0006640.

DALLE-DONNE, I., ROSSI, R., GIUSTARINI, D., COLOMBO, R. & MILZANI, A. 2007. S-glutathionylation in protein redox regulation. *Free Radical Biology and Medicine,* 43**,** 883-898.

DEPONTE, M. & BECKER, K. 2005. Glutathione S-transferase from malarial parasites: structural and functional aspects. *Methods in Enzymology,* 401**,** 241-253.

DEVADOSS, D., RAMAR, M. & CHINNASAMY, A. 2018. Galangin, a dietary flavonol inhibits tumor initiation during experimental pulmonary tumorigenesis by modulating xenobiotic enzymes and antioxidant status. *Archives of Pharmacal Research,* 41**,** 265-275.

DIXON, D. P. & EDWARDS, R. 2010. Glutathione transferases. *The Arabidopsis Book/American Society of Plant Biologists,* 8.

DOURADO, D., FERNANDES, P. A. & RAMOS, M. J. 2008. Mammalian cytosolic glutathione transferases. *Current Protein & Peptide Science,* 9**,** 325-337.

ELSBY, R., KITTERINGHAM, N. R., GOLDRING, C. E., LOVATT, C. A., CHAMBERLAIN, M., HENDERSON, C. J., WOLF, C. R. & PARK, B. K. 2003. Increased constitutive c-Jun N-terminal kinase signaling in mice lacking glutathione S-transferase Pi. *Journal of Biological Chemistry,* 278**,** 22243-22249

FAN, K. C., HUANG, Y. C. & LI, C. H. 1995. Radioimmunoassay for plasma glutathione S-transferase-pi and its clinical application in gastrointestinal cancer. *Cancer,* 76**,** 1363-1367.

FEASEY, N., WANSBROUGH-JONES, M., MABEY, D. C. & SOLOMON, A. W. 2009. Neglected tropical diseases. *British Medical Bulletin,* 93**,** 179-200.

FEASEY, N. A., DOUGAN, G., KINGSLEY, R. A., HEYDERMAN, R. S. & GORDON, M. A. 2012. Invasive non-typhoidal salmonella disease: an emerging and neglected tropical disease in Africa. *The Lancet,* 379**,** 2489-2499.

FRAGOULAKI, M., AXARLI, I., LABROU, N. & CLONIS, Y. 2007. Recombinant glutathione S-transferase for the determination of the herbicide alachlor: The foundations of an optical biosensor. 1st UK-US Conference on Chemical and Biological Sensors and Detectors. *London, UK*.

FRITZ-WOLF, K., BECKER, A., RAHLFS, S., HARWALDT, P., SCHIRMER, R. H., KABSCH, W. & BECKER, K. 2003. X-ray structure of glutathione S-transferase from the malarial parasite *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences,* 100**,** 13821-13826.

FROVA, C. 2006. Glutathione transferases in the genomics era: new insights and perspectives. *Biomolecular Engineering,* 23**,** 149-169.

GARGALA, G. 2008. Drug treatment and novel drug target against *Cryptosporidium*. *Parasite*, 15, 275-281.

GHONEIM, N. H., HASSANAIN, M. A., HAMZA, D. A., SHAAPAN, R. M. & DRAZ, S. H. 2017. Prevalence and molecular epidemiology of *Cryptosporidium* infection in calves and hospitalized children in Egypt. *Research Journal of Parasitology,* 12**,** 19-26.

GILGE, J. L., FISHER, M. & CHAI, Y.-C. 2008. The effect of oxidant and the non-oxidant alteration of cellular thiol concentration on the formation of protein mixed-disulfides in HEK 293 cells. *PloS One,* 3**,** e4015.

GOODCHILD, S. C., HOWELL, M. W., CORDINA, N. M., LITTLER, D. R., BREIT, S. N., CURMI, P. M. & BROWN, L. J. 2009. Oxidation promotes insertion of the CLIC1 chloride intracellular channel into the membrane. *European Biophysics Journal,* 39**,** 129.

GUERRANT, D. I., MOORE, S. R., LIMA, A. A., PATRICK, P. D., SCHORLING, J. B. & GUERRANT, R. L. 1999. Association of early childhood diarrhea and cryptosporidiosis with impaired physical fitness and cognitive function four-seven years later in a poor urban community in northeast Brazil. *The American Journal of Tropical Medicine and Hygiene,* 61**,** 707-713.

GUERRANT, R. L., HUGHES, J. M., LIMA, N. L. & CRANE, J. 1990. Diarrhea in developed and developing countries: magnitude, special settings, and etiologies. *Reviews of Infectious Diseases,* 12**,** S41-S50.

GUK, S.-M., YONG, T.-S. & CHAI, J.-Y. 2003. Role of murine intestinal intraepithelial lymphocytes and lamina propria lymphocytes against primary and challenge infections with *Cryptosporidium parvum. Journal of Parasitology,* 89**,** 270-275.

GUSTAFSSON, A., PETTERSSON, P. L., GREHN, L., JEMTH, P. & MANNERVIK, B. 2001. Role of the glutamyl α-carboxylate of the substrate glutathione in the catalytic mechanism of human glutathione transferase A1-1. *Biochemistry,* 40**,** 15835-15845.

HEGAZY, U. M., HELLMAN, U. & MANNERVIK, B. 2006. Replacement surgery with unnatural amino acids in the lock-and-key joint of glutathione transferase subunits. *Chemistry and Biology,* 13**,** 929-936.

HARHAY, M. O., HORTON, J. & OLLIARO, P. L. 2010. Epidemiology and control of human gastrointestinal parasites in children. *Expert review of anti-infective therapy,* 8**,** 219-234.

HEGAZY, U. M., MANNERVIK, B. & STENBERG, G. 2004. Functional role of the lock and key motif at the subunit interface of glutathione transferase p1-1. *Journal of Biological Chemistry,* 279**,** 9586-9596.

HIJJAWI, N., MELONI, B., NG'ANZO, M., RYAN, U., OLSON, M., COX, P., MONIS, P. & THOMPSON, R. 2004. Complete development of *Cryptosporidium parvum* in host cell-free culture. *International Journal for Parasitology,* 34**,** 769-777.

HILLER, N., FRITZ-WOLF, K., DEPONTE, M., WENDE, W., ZIMMERMANN, H. & BECKER, K. 2006. *Plasmodium falciparum* glutathione S-transferase—Structural and mechanistic studies on ligand binding and enzyme inhibition. *Protein Science,* 15**,** 281-289.

HOTEZ, P. J., ALVARADO, M., BASÁÑEZ, M.-G., BOLLIGER, I., BOURNE, R., BOUSSINESQ, M., BROOKER, S. J., BROWN, A. S., BUCKLE, G. & BUDKE, C. M. 2014. The global burden of disease study 2010: interpretation and implications for the neglected tropical diseases. *PLoS Neglected Tropical Diseases,* 8, 2865.

HOTEZ, P. J., FENWICK, A., SAVIOLI, L. & MOLYNEUX, D. H. 2009. Rescuing the bottom billion through control of neglected tropical diseases. *The Lancet,* 373**,** 1570-1575.

HOTEZ, P. J., MOLYNEUX, D. H., FENWICK, A., KUMARESAN, J., SACHS, S. E., SACHS, J. D. & SAVIOLI, L. 2007. Control of neglected tropical diseases. *New England Journal of Medicine,* 357**,** 1018-1027.

IRVINE, E. J. & MARSHALL, J. K. 2000. Increased intestinal permeability precedes the onset of Crohn's disease in a subject with familial risk. *Gastroenterology,* 119**,** 1740-1744.

JAGAI, J. S., CASTRONOVO, D. A., MONCHAK, J. & NAUMOVA, E. N. 2009. Seasonality of cryptosporidiosis: A meta-analysis approach. *Environmental Research,* 109**,** 465-478.

JOHANSSON, A.-S. & MANNERVIK, B. 2001. Human glutathione transferase A3-3, a highly efficient catalyst of double-bond isomerization in the biosynthetic pathway of steroid hormones. *Journal of Biological Chemistry,* 276**,** 33061-33065.

KANG, J., JU, H., SOHN, W. & NA, B. 2014. Characterization of biochemical properties of a selenium-independent glutathione peroxidase of *Cryptosporidium parvum*. *Parasitology,* 141**,** 570.

KARAVANGELI, M., LABROU, N. E., CLONIS, Y. D. & TSAFTARIS, A. 2005. Development of transgenic tobacco plants overexpressing maize glutathione S-transferase I for chloroacetanilide herbicides phytoremediation. *Biomolecular Engineering,* 22**,** 121-128.

KHAN, A., SHAMS, S., KHAN, S., KHAN, M. I., KHAN, S. & ALI, A. 2019. Evaluation of prevalence and risk factors associated with *Cryptosporidium* infection in rural population of district Buner, Pakistan. *PloS One,* 14**,** e0209188.

KHAN, A. S., SWERDLOW, D. L. & JURANEK, D. D. 2001. Precautions against biological and chemical terrorism directed at food and water supplies. *Public Health Reports,* 116**,** 3.

KING, C. H. & BERTINO, A.-M. 2008. Asymmetries of poverty: why global burden of disease valuations underestimate the burden of neglected tropical diseases. *PLoS Neglected Tropical Diseases,* 2, 209.

KNIGHT, T. R., CHOUDHURI, S. & KLAASSEN, C. D. 2008. Induction of hepatic glutathione S-transferases in male mice by prototypes of various classes of microsomal enzyme inducers. *Toxicological Sciences,* 106**,** 329-338.

KOTLOFF, K. L., NATARO, J. P., BLACKWELDER, W. C., NASRIN, D., FARAG, T. H., PANCHALINGAM, S., WU, Y., SOW, S. O., SUR, D. & BREIMAN, R. F. 2013. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *The Lancet,* 382**,** 209-222.

LAGIER, M. J., ZHU, G. & KEITHLY, J. S. 2001. Characterization of a heavy metal ATPase from the apicomplexan *Cryptosporidium parvum*. *Gene,* 266**,** 25-34.

LITTLER, D. R., HARROP, S. J., GOODCHILD, S. C., PHANG, J. M., MYNOTT, A. V., JIANG, L., VALENZUELA, S. M., MAZZANTI, M., BROWN, L. J. & BREIT, S. N. 2010. The enigma of the CLIC proteins: Ion channels, redox proteins, enzymes, scaffolding proteins? *FEBS letters,* 584**,** 2093-2101.

LOCK, J. T., SINKINS, W. G. & SCHILLING, W. P. 2011. Effect of protein S-glutathionylation on Ca2+ homeostasis in cultured aortic endothelial cells. *American Journal of Physiology-Heart and Circulatory Physiology,* 300**,** H493-H506.

MANNERVIK, B., ALIN, P., GUTHENBERG, C., JENSSON, H., TAHIR, M. K., WARHOLM, M. & JÖRNVALL, H. 1985. Identification of three classes of cytosolic glutathione transferase common to several mammalian species: correlation between structural data and enzymatic properties. *Proceedings of the National Academy of Sciences,* 82**,** 7202-7206.

MANNERVIK, B., BOARD, P. G., HAYES, J. D., LISTOWSKY, I. & PEARSON, W. R. 2005. Nomenclature for mammalian soluble glutathione transferases. *Methods in Enzymology,* 401**,** 1-8.

MANNERVIK, B., HELENA DANIELSON, U. & KETTERER, B. 1988. Glutathione transferases—structure and catalytic activit. *Critical Reviews in Biochemistry,* 23**,** 283-337.

MAUZY, M. J., ENOMOTO, S., LANCTO, C. A., ABRAHAMSEN, M. S. & RUTHERFORD, M. S. 2012. The *Cryptosporidium parvum* transcriptome during in vitro development. *PloS One,* 7.

MEAD, P. S., SLUTSKER, L., DIETZ, V., MCCAIG, L. F., BRESEE, J. S., SHAPIRO, C., GRIFFIN, P. M. & TAUXE, R. V. 1999. Food-related illness and death in the United States. *Emerging Infectious Diseases,* 5**,** 607.

MERRIMAN, H. 2013. CHAPTER Infectious Diseases. *Acute Care Handbook for Physical Therapists***,** 313.

MIYAMOTO, Y. & ECKMANN, L. 2015. Drug development against the major diarrhea-causing parasites of the small intestine, *Cryptosporidium* and Giardia. *Frontiers in Microbiology,* 6**,** 1208.

MORGAN, A. S., CIACCIO, P. J., TEW, K. D., KAUVAR, L. M. & CIACCIO, F. 1996. Isozyme-specific glutathione S-transferase inhibitors potentiate drug sensitivity in cultured human tumor cell lines. *Cancer Chemotherapy and Pharmacology,* 37**,** 363-370.

MOROU, E., ISMAIL, H. M., DOWD, A. J., HEMINGWAY, J., LABROU, N., PAINE, M. & VONTAS, J. 2008. A dehydrochlorinase-based pH change assay for determination of DDT in sprayed surfaces. *Analytical Biochemistry,* 378**,** 60-64.

MORSE, S. S., MAZET, J. A., WOOLHOUSE, M., PARRISH, C. R., CARROLL, D., KARESH, W. B., ZAMBRANA-TORRELIO, C., LIPKIN, W. I. & DASZAK, P. 2012. Prediction and prevention of the next pandemic zoonosis. *The Lancet,* 380**,** 1956-1965.

MUNYAMPUNDU, J.-P., XU, Y.-P. & CAI, X.-Z. 2016. Phi class of glutathione S-transferase gene superfamily widely exists in nonplant taxonomic groups. *Evolutionary Bioinformatics,* 12**,** S35909.

NEBERT, D. W. & VASILIOU, V. 2004. Analysis of the glutathione S-transferase (GST) gene family. *Human genomics,* 1**,** 460.

O'DONOGHUE, P. J. 1995. *Cryptosporidium* and cryptosporidiosis in man and animals. *International Journal for Parasitology,* 25**,** 139-195.

OAKLEY, A. 2011. Glutathione transferases: a structural perspective. *Drug Metabolism Reviews,* 43**,** 138-151.

OAKLEY, A. J. 2005. Glutathione transferases: new functions. *Current Opinion in Structural Biology,* 15**,** 716-723.

PERBANDT, M., HÖPPNER, J., BURMEISTER, C., LÜERSEN, K., BETZEL, C. & LIEBAU, E. 2008. Structure of the extracellular glutathione S-transferase OvGST1 from the human pathogenic parasite Onchocerca volvulus. *Journal of Molecular Biology,* 377**,** 501-511.

PICKETT, C. B. & LU, A. Y. 1989. Glutathione S-transferases: gene structure, regulation, and biological function. *Annual Review of Biochemistry,* 58**,** 743-764.

PRADE, L., HUBER, R., MANOHARAN, T. H., FAHL, W. E. & REUTER, W. 1997. Structures of class pi glutathione S-transferase from human placenta in complex with substrate, transition-state analogue and inhibitor. *Structure,* 5**,** 1287-1295.

PISARSKI, K. 2019. The global burden of disease of zoonotic parasitic diseases: top 5 contenders for priority consideration. *Tropical Medicine and Infectious Disease,* 4**,** 44.

ROBINSON, A., HUTTLEY, G. A., BOOTH, H. S. & PHILIP, G. 2004. Modelling and bioinformatics studies of the human Kappa-class glutathione transferase predict a novel third glutathione transferase family with similarity to prokaryotic 2-hydroxychromene-2-carboxylate isomerases. *Biochemical Journal,* 379**,** 541-552.

ROSSJOHN, J., FEIL, S. C., WILCE, M. C., SEXTON, J. L., SPITHILL, T. W. & PARKER, M. W. 1997. Crystallization, structural determination and analysis of a novel parasite vaccine candidate: *Fasciola hepatica* glutathione S-transferase. *Journal of Molecular Biology,* 273**,** 857-872.

ROSSLE, N. F. & LATIF, B. 2013. Cryptosporidiosis as threatening health problem: a review. *Asian Pacific Journal of Tropical Biomedicine,* 3**,** 916-924.

RYAN, U. & HIJJAWI, N. 2015. New developments in *Cryptosporidium* research. *International Journal for Parasitology,* 45**,** 367-373.

RYAN, U., HIJJAWI, N. & XIAO, L. 2018. Foodborne cryptosporidiosis. *International Journal for Parasitology,* 48**,** 1-12.

RYAN, U., ZAHEDI, A. & PAPARINI, A. 2016. *Cryptosporidium* in humans and animals—a one health approach to prophylaxis. *Parasite Immunology,* 38**,** 535-547.

SALINAS, A. E. & WONG, M. G. 1999. Glutathione S-transferases-a review. *Current Medicinal Chemistry,* 6**,** 279-310.

SAVIOLI, L., SMITH, H. & THOMPSON, A. 2006. *Giardia* and *Cryptosporidium* join the 'neglected diseases initiative'. *Trends in Parasitology,* 22**,** 203-208.

SCIAN, M., LE TRONG, I., MAZARI, A. M., MANNERVIK, B., ATKINS, W. M. & STENKAMP, R. E. 2015. Comparison of epsilon-and delta-class glutathione S-transferases: the crystal structures of the

glutathione S-transferases DmGSTE6 and DmGSTE7 from *Drosophila melanogaster*. *Acta Crystallographica Section D: Biological Crystallography,* 71**,** 2089-2098.

SHEEHAN, D., MEADE, G. & FOLEY, V. M. 2001. Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochemical Journal,* 360**,** 1-16.

SHEHU, D., ABDULLAHI, N. & ALIAS, Z. 2019. Cytosolic Glutathione S-transferase in Bacteria: A Review. *Polish Journal of Environmental Studies,* 28.

SNYDER, M. J. & MADDISON, D. R. 1997. Molecular phylogeny of glutathione-S-transferases. *DNA and Cell Biology,* 16**,** 1373-1384.

SPARKS, H., NAIR, G., CASTELLANOS-GONZALEZ, A. & WHITE JR, A. 2015. Treatment of *Cryptosporidium*: what we know, gaps, and the way forward. *Current Tropical Mededical Reports*, 2 (3): 181–7.

SQUIRE, S. A. & RYAN, U. 2017. *Cryptosporidium* and *Giardia* in Africa: current and future challenges. *Parasites and Vectors,* 10**,** 195.

STRANGE, R. C., JONES, P. W. & FRYER, A. A. 2000. Glutathione S-transferase: genetics and role in toxicology. *Toxicology Letters,* 112**,** 357-363.

STRIEPEN, B. 2013. Parasitic infections: time to tackle cryptosporidiosis. *Nature News,* 503**,** 189.

THOMPSON, R. A., KOH, W. H. & CLODE, P. L. 2016. *Cryptosporidium*—What is it? *Food and Waterborne Parasitology,* 4**,** 54-61.

THOMPSON, R. A., OLSON, M., ZHU, G., ENOMOTO, S., ABRAHAMSEN, M. S. & HIJJAWI, N. 2005. *Cryptosporidium* and cryptosporidiosis. *Advances in Parasitology,* 59**,** 77-158.

TORRES-RIVERA, A. & LANDA, A. 2008. Glutathione transferases from parasites: a biochemical view. *Acta Tropica,* 105**,** 99-112.

TOWNSEND, D. M., MANEVICH, Y., HE, L., HUTCHENS, S., PAZOLES, C. J. & TEW, K. D. 2009. Novel Role for Glutathione S-Transferase π: Regulator of protein s-glutathionylation following oxidative and nitrosative stress. *Journal of Biological Chemistry,* 284**,** 436-445.

TSAVARIS, N. & SKOPELITIS, E. 2007. Cancer Drug Resistance. *Cancer Drug Resistance Research Perspectives***,** 7.

TSUCHIDA, S. 2000. Glutathione transferase. *Experimental protocols for reactive oxygen and nitrogen species. Oxford University Press, Oxford***,** 83-85.

WALKER, C. L. F., SACK, D. & BLACK, R. E. 2010. Etiology of diarrhea in older children, adolescents and adults: a systematic review. *PLoS neglected tropical diseases,* 4.

WILCE, M. C. & PARKER, M. W. 1994. Structure and function of glutathione S-transferases. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology,* 1205**,** 1-18.

WU, B. & DONG, D. 2012. Human cytosolic glutathione transferases: structure, function, and drug discovery. *Trends in Pharmacological Sciences,* 33**,** 656-668.

XIAO, L. & FENG, Y. 2008. Zoonotic cryptosporidiosis. *FEMS Immunology & Medical Microbiology,* 52**,** 309-323.

XU, C., LI, C. Y.-T. & KONG, A.-N. T. 2005. Induction of phase I, II and III drug metabolism/transport by xenobiotics. *Archives of Pharmacal Research,* 28**,** 249.

YANG, G., XU, Z., PENG, S., SUN, Y., JIA, C. & ZHAI, M. 2016. In planta characterization of a tau class glutathione S-transferase gene from *Juglans regia* (JrGSTTau1) involved in chilling tolerance. *Plant Cell Reports,* 35**,** 681-692.

YANG, Y.-L., SERRANO, M. G., SHEORAN, A. S., MANQUE, P. A., BUCK, G. A. & WIDMER, G. 2009. Over-expression and localization of a host protein on the membrane of *Cryptosporidium parvum* infected epithelial cells. *Molecular and Biochemical Parasitology,* 168**,** 95-101.

ZHENG, G.-Q., KENNEY, P. M., ZHANG, J. & LAM, L. K. 1992. Inhibition of benzo [a] pyrene-induced tumorigenesis by myristicin, a volatile aroma constituent of parsley leaf oil. *Carcinogenesis,* 13**,** 1921-1923.

CHAPTER 2:

COMPARATIVE ANALYSES AND STRUCTURAL INSIGHTS OF NEW CLASS GLUTATHIONE TRANSFERASES IN *CRYPTOSPORIDIUM* SPECIES

Mbalenhle Sizamile Mfeka [1], José Martínez-Oyanedel [2], Wanping Chen [3], Ikechukwu Achilonu [4], Syed Khajamohiddin [5]*, Thandeka Khoza [1]*

[1] Department of Biochemistry, School of Life Sciences, University of KwaZulu-Natal (Pietermaritzburg campus), Scottsville, 3209, KwaZulu-Natal, South Africa.

[2] Laboratorio de Biofísica Molecular, Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias Biológicas, Universidad de Concepción, Barrio Universitario S/N, Casilla 160_C, Concepción, Chile.

[3] Department of Molecular Microbiology and Genetics, University of Göttingen, Göttingen 37077, Germany.

[4] Protein Structure-Function Research Unit, School of Molecular and Cell Biology, University of the Witwatersrand, Braamfontein, Johannesburg, South Africa.

[5] Department of Biochemistry and Microbiology, Faculty of Science and Agriculture, University of Zululand, KwaDlangezwa 3886, South Africa.

* Corresponding authors' email:

khajamohiddinsyed@gmail.com and khozat1@ukzn.ac.za

2.1. Abstract

Cryptosporidiosis, caused by protozoan parasites of the genus *Cryptosporidium*, is estimated to rank as a leading cause in the global burden of neglected zoonotic parasitic diseases. This diarrheal disease is the second leading cause of death in children under 5 years of age. Based on the *C. parvum* transcriptome data, glutathione transferase (GST) has been suggested as a drug target against this pathogen. GSTs are diverse multifunctional proteins involved in cellular defense and detoxification in organisms and help pathogens to alleviate chemical and environmental stress. In this study, we performed genome-wide data mining, identification, classification and *in silico* structural analysis of GSTs in fifteen *Cryptosporidium* species. The study revealed the presence three GSTs in each of the *Cryptosporidium* species analyzed in the study. Based on the percentage identity and comprehensive comparative phylogenetic analysis, we assigned *Cryptosporidium* species GSTs to three new GST classes, named Vega (ϑ), Gamma (γ) and Psi (ψ). The study also revealed an atypical thioredoxin-like fold in the *C. parvum* GST1 of the Vega class, whereas *C. parvum* GST2 of the Gamma class and *C. melagridis* GST3 of the Psi class has a typical thioredoxin-like fold in the N-terminal region. This study reports the first comparative analysis of GSTs in *Cryptosporidium* species.

2.2. Introduction

Cryptosporidiosis is a zoonotic parasitic disease that is caused by *Cryptosporidium* spp. (Savioli *et al*., 2006, Pisarski, 2019, Pumipuntu and Piratae, 2018). This disease is estimated to be among the highest ranking causes in the global burdens of zoonotic parasitic disease, with an estimate of 8.37 million disability-adjusted life years (Hotez *et al*., 2014, Pisarski, 2019). Recently, large population studies revealed that cryptosporidiosis has become a fast-growing burden to children under the age of five years (Kotloff *et al*., 2013, Sow *et al*., 2016). Moreover, the Global Enteric Multicenter Study showed that *Cryptosporidium* is significantly associated with diarrheal disease among children <24 months of age in sub-Saharan Africa and South Asia (Sow *et al*., 2016). Similar studies also found *Cryptosporidium* to be the second leading cause of moderate to severe diarrhea in infants after Rotavirus (Kotloff *et al*., 2013). It is interesting to note that vaccines/treatment are already available or fast being developed for three of four diarrheal pathogens (*Rotavirus, Shigella* and heat-stable, enterotoxigenic *Escherichia coli*), the exception being *Cryptosporidium,* highlighting the need to address this disease (Striepen, 2013). Despite the global burden of cryptosporidiosis, to date nitazoxanide (NTZ) is the only treatment available for this disease. NTZ only appears to be effective in patients with a good immune response, whilst having limited efficacy in malnourished children and ineffective in immunocompromised people (Amadi *et al*., 2009, Bhalchandra *et al*., 2018, Cabada and White Jr, 2010). The lack of effective treatment for cryptosporidiosis, coupled with the fact that it is now considered the most common cause of human parasitic diarrhea in the world, highlights the need for more research on *Cryptosporidium* to identify new drug targets and thus develop new drugs (Widmer *et al*., 2020).

Cryptosporidiosis is typically characterized by nausea, profuse watery diarrhea, abdominal cramps, vomiting and low-grade fever, which manifest after 14 days and last up to 2.5 months in immune-competent patients (Certad *et al*., 2017, Leitch and He, 2011). These symptoms are usually self-limiting in immune-competent patients; however, in immunocompromised hosts they can be devastating, with the disease manifesting as life-threatening and often becoming extraintestinal (Leitch and He, 2011). The gastrointestinal infection can spread to other sites, such as the gall bladder, biliary tract, pancreas and pulmonary system. Cryptosporidiosis can be contracted through the fecal-oral route, through contact with infected animals or humans or contaminated food or water (Leitch and He, 2011).

Of the *Cryptosporidium* species that exist, *C. hominis* and *C. parvum* are responsible for the highest level of clinically relevant infections worldwide (Pumipuntu and Piratae, 2018). The remaining species have mild zoonotic properties causing moderate-to-severe diarrhea in humans (Pumipuntu and Piratae, 2018). *Cryptosporidium* species are reported to have an efficient defense mechanism that allows it to cope with a wide range of environmental stresses such as changes in temperature, drugs, free radicals, as well as the host's immune responses at various life stages (Certad *et al*., 2017). Genome analysis of *C. parvum* revealed that it contains various defense proteins such as glutathione transferase (GST), glutathione peroxidase and superoxide dismutase, which are known for detoxification, signal modulation and aromatic amino acid catabolism (Abrahamsen *et al*., 2004). The existence of these enzymes may provide *C. parvum* with the abilities to maintain its parasitic lifecycle, enabling it to survive and persist in its host.

Among the above-mentioned enzymes, GST is found to be expressed in all stages of the *C. parvum* parasite's life cycle (Mauzy *et al*., 2012), thus making it a promising therapeutic target (Khoza *et al*., 2018). GSTs have been studied as drug targets against infectious agents and metabolic disorders (Harwaldt *et al*., 2002, Pljesa-Ercegovac *et al*., 2018, Rao *et al*., 2000). GSTs are a diverse group of multifunctional proteins that are distributed ubiquitously in eukaryotes and prokaryotes (Allocati *et al*., 2012, Sheehan *et al*., 2001). These enzymes play an important role in cellular defense and detoxification (Frova, 2006, Sheehan *et al*., 2001, Wilce and Parker, 1994). They catalyze the nucleophilic conjugation of the reduced tripeptide glutathione (GSH) thiol group to the electrophilic substrates to convert them to less harmful, more soluble compounds. Based on the location, the GST superfamily is divided into three sub-families namely, soluble or cytosolic GSTs, mitochondrial GSTs and membrane-associated proteins involved in eicosanoid and gluthatione metabolism (MAPEG) with the cytosolic GSTs being the most characterized (Table S1). The GSTs are generally divided into classes based on amino acid sequence similarity, with GSTs within each class sharing similar immunological cross-reactivity and specificity towards the electrophilic substrate and sensitivity to inhibitors (Glisic *et al*., 2015, Salinas and Wong, 1999, Sheehan *et al*., 2001). GSTs within each class typically share as little as 60% amino acid sequence identity however, some classes can share from as little as 40% (Allocati *et al.,* 2009, Frova, 2006, Oakley, 2011, Sheehan *et al*., 2001, Soranzo *et al*., 2004). It is generally accepted that the assignment of different GSTs to specific

classes must fall within these limits, with sequences sharing less than 25-30% designated to their own class (Allocati *et al*., 2009, Frova, 2006, Oakley, 2011, Sheehan *et al*., 2001, Soranzo *et al*., 2004). Information on different GST classes found in organisms, their cellular localization and functions are listed in Table S1.

Typical GSTs are dimeric in structure and each monomer is divided into two domains (Frova, 2006, Sheehan *et al*., 2001). The N-terminal domain of conical GSTs assumes a topology resembling the thioredoxin fold with a βαβ-ββα motif. This domain also houses an important conserved region of the active site where a catalytically active Tyr, Ser or Cys is found to interact with the GSH thiol group. The C-terminal domain of typical GSTs is all helical and connected by a short linker sequence called the *cis*-Pro loop with a highly-conserved proline residue in *cis* conformation (Frova, 2006). The active site is comprised of the glutathione binding site (G-site) and the hydrophobic substrate binding site (H-site), located in the N-terminal and C–terminal domain respectively. The G-site exclusively binds glutathione and is highly conserved, whilst the H-site accepts more variability so to accommodate an extensive range of toxic electrophilic substances (Frova, 2006, Sheehan *et al*., 2001).

Despite the importance of GSTs, especially as potential drug targets against *Cryptosporidium* (Khoza *et al*., 2018), to the best of our knowledge, no literature is available to date on *Cryptosporidium* GSTs with regards to their distribution, the GST classes and structural information. Thus, this study is aimed at addressing this research gap. In this study, genome data mining, identification, phylogenetic and structural analysis of GSTs in fifteen *Cryptosporidium* species was carried out.

2.3. Methods

2.3.1. Species and Database

Cryptosporidium species genomes that are available for public use at the *Cryptosporidium* database or CryptoDB (Heiges *et al*., 2006) (https://cryptodb.org/cryptodb/app; release 48 beta, 27 August 2020; accessed on 14 September 2020) and at National Center for Biotechnology information (NCBI) (Agarwala *et al*., 2018) (https://www.ncbi.nlm.nih.gov/datasets/genomes/?txid=5806; accessed on 14 September 2020) were used in the study. The *Cryptosporidium* pathogens examined in this study include ones from both humans and other mammals (Table 2.1).

Table 2.1: *Cryptosporidium* species used in the study and their host specificity.

Species and isolates	Host range	Reference(s)
Cryptosporidium andersoni isolate 30847	Cattle, sheep, bactrian camel, gerbil	Liu *et al*., 2016
Cryptosporidium hominis isolate TU502_2012	Humans	Ifeonu *et al*., 2016
Cryptosporidium hominis isolate 30976	Humans	Guo *et al*., 2015
Cryptosporidium hominis TU502	Humans	Xu *et al*., 2004
Cryptosporidium hominis UdeA01	Humans	Isaza *et al*., 2015
Cryptosporidium meleagridis strain UKMEL1	Turkey, chicken, bobwhite quail, dog	Ifeonu *et al*., 2016
Cryptosporidium parvum Iowa II	Mammals, including humans	Abrahamsen *et al*., 2004, Bankier *et a*l., 2003
Cryptosporidium tyzzeri isolate UGA55	Domestic mice	Sateriale *et al*., 2019
Cryptosporidium ubiquitum isolate 39726	Deer, sheep, goat, squirrel, mouse	Liu *et al*., 2016
Cryptosporidium muris RN66	Mice and cats	Mc Donald *et al*., 1992, Uni *et al*., 1987
Cryptosporidium baileyi strain TAMU-09Q1	Chickens and Black-headed full	Ng *et al*., 2006, Pavlásek, 1993
Cryptosporidium viatorum isolate UKVIA1	Humans and rats	Elwin *et al*., 2012, Koehler *et al*., 2018
Cryptosporidium sp. *chipmunk* LX-2015	Mice, squirrels, chipmunks	Guo *et al*., 2015a, Prediger *et al*., 2017, Stenger *et al*., 2015
Cryptosporidium ryanae isolate 45019	Cattle	Fayer *et al*., 2008
Cryptosporidium bovis isolate 42482	Sheep, goats and cattle	Fayer *et al*., 2005

2.3.2. Genome Data Mining, Identification and Classification of GSTs

Cryptosporidium species genomes available at CryptoDB (Heiges *et al*., 2006) were mined for GSTs. Two different methods followed for GST mining. First, the genomes of *Cryptosporidium* species were mined using the term "glutathione transferase" in the search tool. Second, the species genomes were blasted with GST proteins from *Homo sapiens* (protein ID:P08263) (Board and Webb, 1987) and *C. parvum* Iowa II (protein ID: EAK89476.1) (Bankier *et al*., 2003, Abrahamsen *et al*., 2004). The BLASTP mined proteins revealed a range of apicomplexan species which were filtered out to show only *Cryptosporidium* species. The hit proteins were then collected and subjected to protein family analysis using the Pfam (El-Gebali *et al*., 2019) and InterPro (Mitchell *et al*., 2019) programs. The results were analyzed and the hit proteins that were classified as GST by Pfam (PF14497, PF13417 and, PF17172) (El-Gebali *et al*., 2019) and InterPro (IPR036282, IPR004045 and IPR010987) (Mitchell *et al*., 2019) were selected.

For the collection of more hits, *Cryptosporidium* species genomes available at NCBI database (Agarwala *et al*., 2018) was blasted with two GST proteins from *C. andersoni* 30847 (cand_012830 & cand_023790) and from *C. meleagridis* UKMEL1 (CmeUKMEL1_05845) that were collected from CryptoDB (Heiges *et al*., 2006). The hit proteins were screened for GSTs following the method described above.

A final total count was presented by deleting the duplicated GSTs. The selected GSTs were then grouped into different classes or groups based on their percentage identity, following the conventional criterion of less than 25-30% identity being a new class (Allocati *et al*., 2009, Frova, 2006, Oakley, 2011, Sheehan *et al*., 2001, Soranzo *et al*., 2004).

2.3.3. Analysis of Homology

The percentage identity between GSTs was deduced using Clustal Omega (Sievers *et al*., 2011). The full-length GSTs were subjected to Clustal analysis which produced the percentage identity amongst each of the proteins as matrix identity results. These results were laid out in an Excel spreadsheet where the results were analyzed to identify the percentage identity between GSTs.

2.3.4. Collection of Different GST Classes' Protein Sequences

For comparative analysis, GST protein sequences belonging to different GST classes were collected using multiple methods to build a library for phylogenetic analysis. On the European Molecular Biology Laboratory (EMBL) site (Madeira *et al*., 2019), GSTs sequences that are placed under the GST superfamily (IPR040079) were retrieved. The GST classes namely CLIC (IPR002946), Alpha (IPR003080), Mu class (IPR003081), Pi (IPR003082), Omega (IPR005442), Zeta (IPR005955) and Sigma (IPR003083) were collected under EMBL. More sequences were obtained through text search using the UniProt protein knowledge base (Consortium, 2019). A specific GST class was searched on the site and the hits obtained were further verified using Pfam (El-Gebali *et al*., 2019) and InterPro (Mitchell *et al*., 2019) to ensure uniformity with the GSTs collected from the EMBL site (Madeira *et al*., 2019). The remaining GSTs that were not in the databases were retrieved from published articles.

The *Cryptosporidium* species GST sequences along with protein sequences of different GST classes used in the phylogenetic analysis are presented in Appendix B.

2.3.5. Phylogenetic Analysis

The GST sequences in supplementary dataset 1 were used to make a phylogenetic tree for inferring their evolutionary relationship. First, all the GST protein sequences were aligned by MAFFT v6.864 embedded on the Trex-online server (Boc *et al*., 2012). Then, the alignment was automatically submitted to the server for inferring the tree with different models and the optimized tree was selected. Finally, the tree was submitted to iTOL for viewing and annotation (Letunic and Bork, 2019). Thioredoxin from *Oryctolagus cuniculus* (protein ID: P08628) was used as an outgroup.

For the construction of the phylogenetic tree of the *Cryptosporidium* GST proteins, the protein sequences were aligned using MUSCLE software (Edgar, 2004) embedded in MEGA7 (Kumar *et al*., 2016). The evolutionary history was inferred by using the maximum likelihood method with 100 bootstrap replication based on the JTT matrix-based model (Jones *et al*., 1992). Evolutionary analyses were conducted in MEGA7.

2.3.6. Cellular Localization and Transmembrane Helices Prediction

Cellular localization of GSTs was predicted using the Bologna Unified Subcellular Component Annotator (BUSCA) (Savojardo *et al*., 2018). BUSCA is the latest, accurate program available for the prediction of proteins' subcellular localization; it integrates different computational tools such as identifying signal and transit peptides (DeepSig and TP-pred3), GPI-anchors (PredGPI) and transmembrane domains (ENSEMBLE3.0 and BetAware) with tools for discriminating subcellular localization of both globular and membrane proteins (BaCelLo, MemLociand SChloro) (Savojardo *et al*., 2018). The outcomes of these different programs were processed and integrated to predict subcellular localization of both eukaryotic and bacterial proteins (Savojardo *et al*., 2018). Prediction of transmembrane helices in GSTs was done using TMHMM Server v. 2.0 (Möller *et al*., 2001). This program is well known for its high degree of accuracy in the prediction of transmembrane helices and discrimination between soluble and membrane proteins.

2.3.7. Template Identification

To construct 3D models of proteins, reference protein structures previously solved by crystallization or nuclear magnetic resonance (NMR) are needed. These would serve to simulate not only the fold of a protein but also a full atom model to build. These proteins are referred to as templates. Either single or multiple templates can be used in constructing the 3D model of a protein (Fiser, 2010). In this study, three different web servers, namely NCBI BLAST (v2.10.1) (Altschul *et al*., 1990), i-TASSER (v5.1) (Yang and Zhang, 2015) and PHYRE (v2.0) (Kelley *et al*., 2015), were consulted to identify the most suitable templates for GST proteins. Based on the highest percentage identity and sequence coverage, the best templates were selected for modeling each GST protein. In cases where the templates had the same percentage identity and sequence coverage, we selected the template with the highest resolution for modelling.

2.3.8. Protein Sequence Alignment for Modelling

T-COFFEE webserver (Di Tommaso *et al*., 2011) was used for aligning the GST proteins and the template sequences. The aligned files were downloaded in FASTA format and modified to generate files to be used for protein modelling (Webb and Sali, 2016).

2.3.9. Protein Modelling, Optimization and Validation

The MODELLER v9.21 program (Webb and Sali, 2016) was used to build GST models. Multiple structures were produced by Modeller 9.21. The model with the best DOPE assessment was selected as the output structure to be used. The structures modeled were viewed using PyMOL (Schrodinger, 2010). The model for each GST was then subjected to evaluation for stereochemistry and energetic quality at the Structural Analysis and Verification Server (http://servicesn.mbi.ucla.edu/SAVES/) and ProsaII (https://prosa.services.came.sbg.ac.at/) (Wiederstein and Sippl, 2007). Based on the validation results, the protein models were then refined on the GalaxyWeb Refiner server (Ko *et al*., 2012). After refinement, the models were again subjected to evaluation and validation using programs such as ERRAT (Colovos and Yeates, 1993), Verify3D (Eisenberg *et al*., 1997), PROCHECK (Laskowski et al., 1993, Laskowski et al., 2006), and RAMPAGE (Wang *et al*., 2016) and ProsaII (Wiederstein and Sippl, 2007).

2.4. Results and Discussion

2.4.1. Two Different Sizes of GSTs Present in *Cryptosporidium* Species

Genome data mining of fifteen *Cryptosporidium* species revealed the presence of three GST genes in each of the species genomes (Table 2.2). The presence of more than one GST gene is common in eukaryotic species (Frova, 2006). Among 45 GSTs, 30 were found to have the characteristic GST motifs (Oakley, 2011, Sheehan et al., 2001), such as the N-terminal domain, which houses the G site, and C-terminal domain, which determines the substrate specificity (H-site) (Table 2.2 and Figure S1). The remaining 15 GSTs have one of the characteristics GST motifs indicating either these sequences are diverse or fragmented or not properly annotated (Table 2.2). These GSTs were considered incomplete and were not included for further analysis unless indicated. Future genome editing and better gene prediction programs will help in getting the complete sequences for these GSTs and possibly predicting characteristic N- and C-terminal motifs. In total, 30 GSTs were taken for further analysis. Analysis of GST protein sizes revealed the presence of two different lengths of GSTs in *Cryptosporidium* species (Table 2.2). One type of GST protein is shorter in size with amino acids ranging between 157 and 268, and another type of GST protein is longer in size, with amino acids ranging between 373 and 466 (Table 2.2). GSTs from *Cryptosporidium* species seem to be the longest in amino acid length, as most of the GSTs reported in other organisms to date are 200-250 amino acids in length (Frova, 2006). Furthermore, it can be noted that the addition in

length is found only on the outer N- and C-terminal regions, with the center of the protein containing the GST-superfamily domains (Table 2.2). In order to assess whether *Cryptosporidium* species GST proteins are indeed properly annotated gene products, we further analyzed the gene structure. Interestingly, all the longer GSTs had a single exon, thus no introns, but shorter GSTs were the products of one to four exons (Table 2.2). This could be indicative of shorter GSTs being prone to having multiple isoforms owing to gene shuffling. Due to the presence of these multiple introns, the production of more diverse short GSTs can be expected compared to longer GSTs (Xu *et al*., 2012).

Table 2.2: Glutathione transferase (GST) analysis in *Cryptosporidium* species. The GST number in column 2 is an indication of the number of GSTs that a specific species possesses. Whilst the number on column 3 indicates the group the protein belongs to (based on the percentage identity) (Allocati *et al*., 2009, Frova, 2006, Oakley, 2011, Sheehan *et al*., 2001, Soranzo *et al*., 2004).

Species	**Total number of GSTs**	**GST number**	**Protein ID**	**Protein size (No of Amino acids)**	**Characteristic GST motifs location**		**Gene structure (No. of exons)**
					N Terminal	**C terminal**	
Cryptosporidium andersoni isolate 30847	3	GST1	cand_012830[$]	197	12-97	95-195	3 exons
		GST2	cand_023790[$]	466	67-149	166-319	1 exon
		GST3	OII73498.1[*]	260	-	124-235	1 exon
Cryptosporidium hominis isolate TU502_2012	3	GST1	ChTU502y2012_407g2365[$]	186	1-62	64-186	2 exons
		GST2	ChTU502y2012_421g0615[$]	428	69-151	146-315	1 exon
		GST3	ChTU502y2012_303g0055/OLQ15919.1[$]	268	-	153-236	1 exon
Cryptosporidium hominis isolate 30976	3	GST1	GY17_00002363[$]	186	1-62	60-183	2 exons
		GST2	GY17_00000733 [$]	428	69-151	146-315	1 exon
		GST3	PPS94453.1[*]	268	-	152-236	1 exon
Cryptosporidium hominis TU502	3	GST1	XP_667744.1[*]	161	1-62	64-161	1 exon
		GST2	Chro.80347 [$]	428	69-151	146-315	1 exon

		GST3	XP_666781.1*	268	-	154-236	1 exon
Cryptosporidium hominis UdeA01	3	GST1	CUV07467.1*	161	1-62	64-161	1 exon
		GST2	CHUDEA8_2970$	428	69-151	146-315	1 exon
		GST3	CUV04748.1*	268	-	154-236	1 exon
Cryptosporidium meleagridis strain UKMEL1	3	GST1	CmeUKMEL1_03350$	193	9-94	96-193	3 exons
		GST2	CmeUKMEL1_14570$	428	69-151	146-315	1 exon
		GST3	CmeUKMEL1_05845$	268	31-118	101-243	1 exon
Cryptosporidium parvum Iowa II	3	GST1	cgd7_4780 $	186	1-62	60-183	2 exons
		GST2	cgd8_2970$	429	69-151	146-315	1 exon
		GST3	cgd2_3730$	268	-	156-236	1 exon
Cryptosporidium tyzzeri isolate UGA55	3	GST1	CTYZ_00001095$	186	1-62	60-186	2 exons
		GST2	CTYZ_00000322$	429	69-151	146-315	1 exon
		GST3	TRY52903.1*	268	-	153-236	1 exon
Cryptosporidium ubiquitum isolate 39726	3	GST1	cubi_03151$	213	1-89	91-213	4 exons
		GST2	cubi_03523$	428	69-151	146-315	1 exon
		GST3	XP_028873506.1*	266	-	159-235	1 exon
Cryptosporidium muris RN66	3	GST1	XP_002141168.1*	160	1-60	58-158	2 exons
		GST2	XP_002140043.1*	466	-	211-312	1 exon

		GST3	XP_002142877.1*	260	-	164-233	1 exon
Cryptosporidium baileyi strain TAMU-09Q1	3	GST1	JIBL01000090.1*	156	1-57	59-156	1 exon
		GST2	JIBL01000106.1*	390	36-118	113-275	1 exon
		GST3	JIBL01000138.1*	236	1-87	69-223	1 exon
Cryptosporidium viatorum isolate UKVIA1	3	GST1	QZWW01000010.1*	161	1-62	64-161	1 exon
		GST2	QZWW01000018.1*	428	69-151	146-315	1 exon
		GST3	QZWW01000026.1*	249	-	134-217	1 exon
Cryptosporidium sp. *chipmunk* LX-2015	3	GST1	JXRN01000042.1*	205	1-106	108-205	1 exon
		GST2	JXRN01000009.1*	425	69-151	-	1 exon
		GST3	JXRN01000023.1*	250	-	135-217	1 exon
Cryptosporidium ryanae isolate 45019	3	GST1	VHLK01000064.1*	166	-	37-154	1 exon
		GST2	VHLK01000046.1*	373	36-118	113-274	1 exon
		GST3	VHLK01000056.1*	230	1-85	89-221	1 exon
Cryptosporidium bovis isolate 42482	3	GST1	VHIT01000033.1*	147	-	30-142	1 exon
		GST2	VHIT01000012.1*	376	21-103	98-264	1 exon

		GST3	VHIT01000028.1*	227	1-85	98-221	1 exon

Symbols: $, protein ID from CryptoDatabase; *, protein ID from NCBI database; -, characteristic GST domain not identified.

2.4.2. *Cryptosporidium* Species GSTs Are Cytosolic in Nature

Most of the GSTs identified in organisms are cytosolic in nature, with the exception of GSTs belonging to the classes MAPEG and Kappa (mitochondrial) (Table S1). In order to identify the cellular localization, we subjected *Cryptosporidium* species GST protein sequences to the TMHMM Server v. 2.0 for the prediction of transmembrane helices in their structure (Möller *et al*., 2001) and the BUSCA server (Savojardo *et al*., 2018) for identifying possible localization in a cell. TMHMM prediction revealed that none of the *Cryptosporidium* species GSTs had transmembrane helices, indicating they were soluble and thus possibly cytosolic (Table S2). To authenticate our results, we also subjected 395 GSTs belonging to 17 different classes to TMHMM prediction (Table S3). The TMHMM predicted the presence of no transmembrane helices in previously designated cytosolic GSTs, whereas transmembrane helices were predicted for previously designated microsomal GSTs (Table S3). This indicated that the TMHMM results on the prediction of no transmembrane helices in *Cryptosporidium* species GSTs were in agreement with previous annotations. Furthermore, BUSCA indicated that all 30 *Cryptosporidium* species GSTs were cytosolic (Table S4). Based on these *in silico* results, we concluded that the 30 *Cryptosporidium* species GSTs were cytosolic in nature.

2.4.3. *Cryptosporidium* Species GSTs Belongs to New Classes

Phylogenetic analysis of *Cryptosporidium* species GSTs revealed that the 30 GSTs could be grouped into three different groups (Figure 2.1). The shorter GSTs were grouped together (Group 1) and so were the longer GSTs (Group 2). Interestingly, despite the short amino acid length, four GSTs diverged from these two groups (Group 3) (Figure 2.1). Analysis of the amino acid percentage identity among *Cryptosporidium* species GSTs further confirmed that they indeed belonged to three different groups. Group 1 GSTs shared an amino acid percentage identity of 54-100%, whereas groups 2 and 3 shared identities of 48-100% and 42-71%, respectively. Group 3 GSTs had 13-21% identity with Group 2 GSTs and 14-22% identity to Group 1 GSTs. The percentage identity between Groups 1 and 2 was 17-25%. This indicates that all three groups of *Cryptosporidium* species GSTs indeed belonged to three different classes as the percentage identity between these groups was below 25-30%, qualifying them to be their own class (Allocati *et al*., 2009, Frova, 2006, Oakley, 2011, Sheehan *et al*., 2001, Soranzo *et al*., 2004).

Although the above results clearly indicated that *Cryptosporidium* species GSTs belong to three different groups, it was still not clear whether they fell under one of the GST classes described in the literature (Table S1). Thus, the comprehensive phylogenetic analysis of proteins belonging to 17 known GST classes and *Cryptosporidium* species GSTs was carried out (Figure 2.2). Phylogenetic analysis revealed that *Cryptosporidium* species GSTs did not align with any of the 17 pre-existing GST classes and formed three new groups (Figure 2.2). This clearly indicates that *Cryptosporidium* species GSTs belong to three different new GST classes. Thus, we named groups 1, 2 and 3 of *Cryptosporidium* GSTs Vega (ϑ), Gamma (γ) and Psi (ψ), respectively. A point to be noted is that all the GST proteins aligned together as per their GST class on the phylogenetic tree, indicating our phylogenetic analysis is correct and thus we concluded that *Cryptosporidium* species GSTs indeed belong to new GST classes.

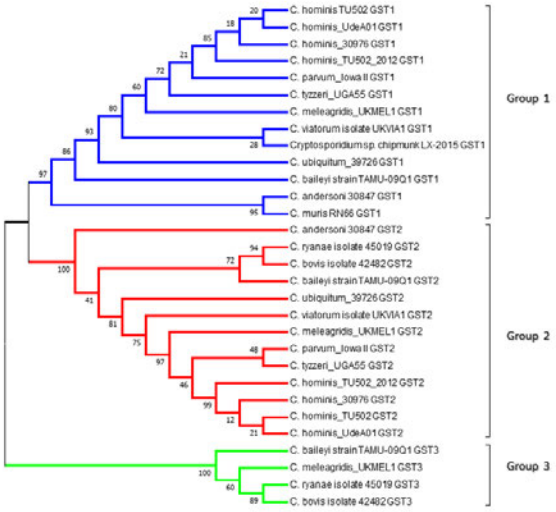


Figure 2.1: Phylogenetic analysis of glutathione transferase (GST) proteins from *Cryptosporidium* species. The evolutionary history was inferred by using the maximum likelihood method based on the JTT matrix-based model (Jones *et al*., 1992). Evolutionary analyses were conducted in MEGA7 (Kumar *et al*., 2016). The percentage of trees (bootstrap value) in which the associated taxa clustered together is shown next to the branches.

2.4.4. *Cryptosporidium parvum* GST1 of Vega Class Has Atypical Thioredoxin-Like Fold

Identification of three new GST classes in *Cryptosporidium* species in this study necessitated examination of the structural aspects of these new classes to see if any deviations or novel folds might be present, compared to the canonical structure of GSTs (Oakley, 2011, Sheehan *et al*., 2001). Analysis of the primary structure revealed that all *Cryptosporidium* species GSTs

have N- and C-terminal regions characteristic of GSTs that usually contain a G-site and H-site Oakley, 2011, Sheehan *et al*., 2001), respectively (Table 2.2 and Figure S1). All GSTs have the highly conserved proline amino acid residue (Figure S1) that is part of the *cis*-Pro loop responsible for connecting the N- and C-terminal regions in order to maintain the GST structural integrity (Allocati *et al*., 1999). It was observed from Figure S1 that Psi class GSTs have a Tyr residue in the N-terminal domain in close proximity to the expected active site Tyr. The same was observed with the Vega class GSTs with the expectation *of C. muris* and *C. baileyi*. Vega and Psi GSTs have a few tyrosine residues in the N-terminal region, but they are not at a position that is considered part of an active site (Oakley, 2011, Sheehan *et al*., 2001) (Figure S1). Similarly, the majority of the Gamma class GSTs consist of an active site Tyr residue with the exception *C. andersoni, C. baileyi, C. ryanae* and *C. bovis* species. In these species, Phe replaces the active site Tyr residue. Mutagenesis studies have shown that the presence of Phe at the supposed position of the active site Tyr significantly reduces the catalytic activity. This highlights the critical role played by the active site Tyr in the catalytic activity of GST (Liu *et al*., 1998, Stenberg *et al*., 1991). The effect of these mutations in the context of *Cryptosporidium* GSTs is yet to be studied.

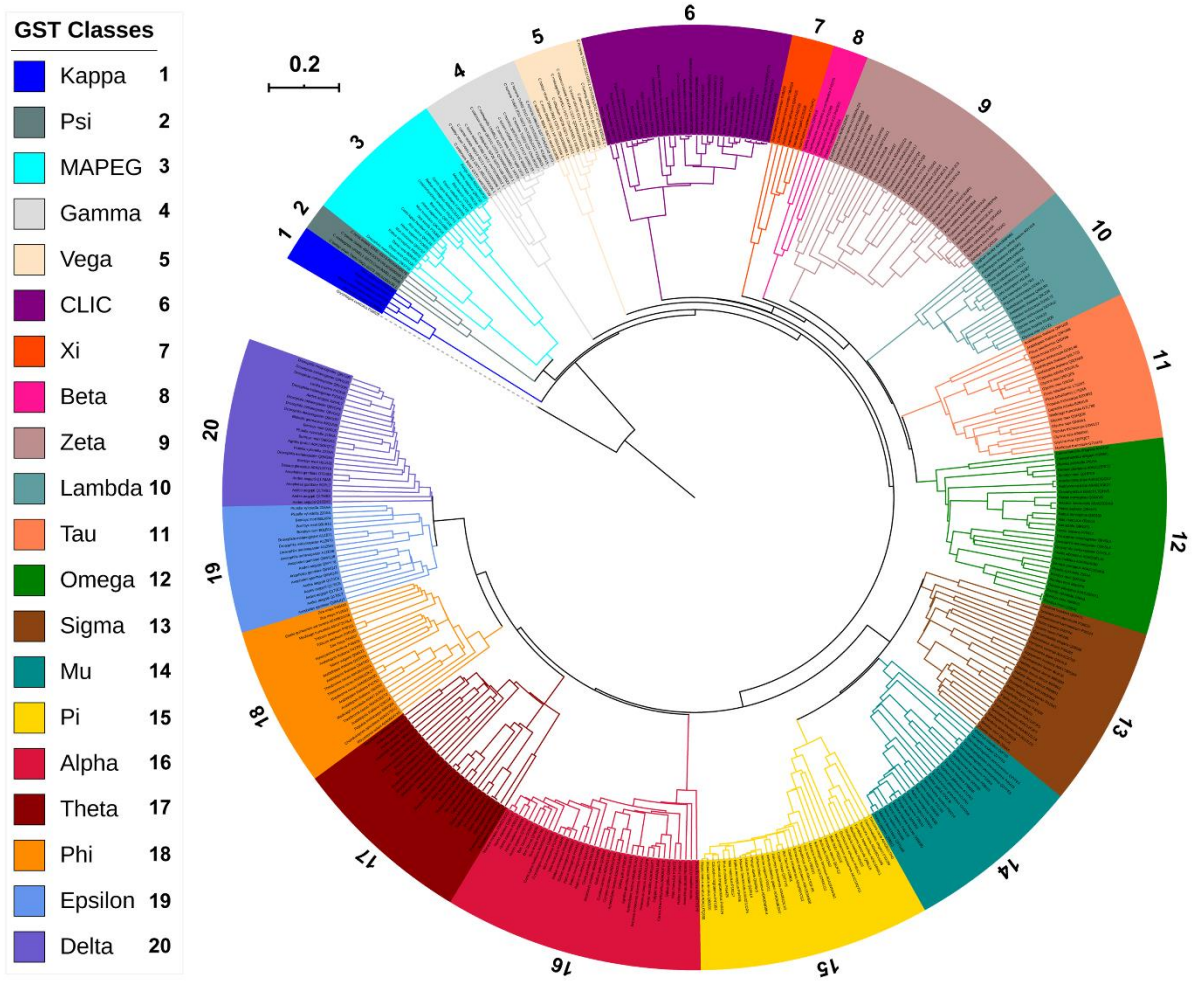


Figure 2.2: Phylogenetic tree of the glutathione transferases (GSTs) protein sequences of *Cryptosporidium* species with GSTs from 17 different GST classes. Thioredoxin from *Oryctolagus cuniculus* (protein ID: P08628) is used as an outgroup. Three new GST classes reported in this study from *Cryptosporidium* species named Vega, Gamma and Psi are also shown in the tree.

Multiple sequence alignments of Vega and Gamma GSTs revealed that amino acids in the N- and C-terminal regions of these GSTs are highly conserved (Figure S1). For this reason, we selected *C. parvum* GSTs 1 and 2 (*Cp*GST1 and *Cp*GST2) as representative of the Vega and Gamma GST classes for structural analysis along with *C. meleagridis* UKMEL1 GST3 (C*m*GST3) for the Psi class. Structural analysis of the three GSTs was carried out using *in silico* homology modeling. The structural analysis was aimed at assessing only the secondary structural elements that are characteristic of GST proteins (Oakley, 2011, Sheehan *et al*., 2001). These GST models are not aimed to assess the binding affinities or the residues involved in binding to different ligands. In order to build 3D models we performed a template search at three different webpages, namely NCBI (Altschul *et al*., 1990), PHYRE (Kelley *et al*., 2015) and I-

TASSER (Yang and Zhang, 2015). The templates found were of low sequence identity but had relatively good coverage (Tables S5). This was expected, since these GSTs are new. We then proceeded to build 3D models using a multiple template method, as this approach is known to improve the quality of homology models (Larsson *et al*., 2008). We built 3D models for all three GSTs, attempting single and multiple templates, while also using different combinations of the available templates listed in Table S5. The best 3D models with good quality closest to the templates were chosen for the structural analysis.

Here, we present the combination of templates that gave *Cp*GST1, *Cp*GST2 and *Cm*GST3 models. The templates used to model *Cp*GST1 were a *Bombyx mori* Sigma class GST (3VPQ-A) (Yamamoto *et al*., 2013) that had 94% coverage and 26% identity and a *Penaeus vannamei* Mu class (5AN1-A) (Juárez-Martínez *et al*., 2017) with 98% coverage and 23% identity (Figure 2.3 and Table S5). For *Cp*GST2 the templates were both from *Homo sapiens* Alpha class (1K3Y-B) (Le Trong *et al*., 2002) and Pi class (19GS-A) (Oakley *et al*., 1999), with sequence identity at 21%, coverage at 94% and 22% identity and 84% coverage (Figure 2.4 Table S5), respectively. The C*m*GST3 templates used were from *Caenorhabditis elegans* Pi class GST (1ZL9-A) (https://www.rcsb.org/structure/1ZL9) with 94% coverage and 21% identity and a *Homo sapiens* Alpha class (1K3Y-B) (Le Trong *et al*., 2002) with 98 % coverage and 22% identity (Figure 2.5 and Table S5).

For each GST, 20 models were built using the MODELLER v9.21 program (Webb and Sali, 2016). The best model evaluated by DOPE score was selected and subjected to structural quality analysis. The selected model for each GST was then refined on the GalaxyWeb Refiner server (Ko *et al*., 2012) and further subjected to structural quality evaluation using different programs such as ERRAT (Colovos and Yeates, 1993), Verify3D (Eisenberg *et al*., 1997), PROCHECK (Laskowski *et al*., 1993, Laskowski *et al*., 2006), RAMPAGE (Wang *et al*., 2016) and ProsaII (Wiederstein and Sippl, 2007). The overall quality of the models was assessed by the combination of these programs' values and by comparing these with the templates' structural evaluation scores (Tables S6 and S7). The models generated for C*p*GST1 and C*p*GST2 were found to be of good quality, as different structural validation programs indicated that the quality of the model structures was close to the quality of the template structures (Tables S6 and S7). The model generated for C*m*GST3 had all parameters in acceptable range including Z-score of -3.68 indicating the model is of good quality with the exception of Verify3D where

26% residues had an average 3D-1D score > 0.2 (Tables S6 and S7). The three GST models generated in the study, along with their corresponding sequence alignments with their templates, are presented in Figures 3-5.

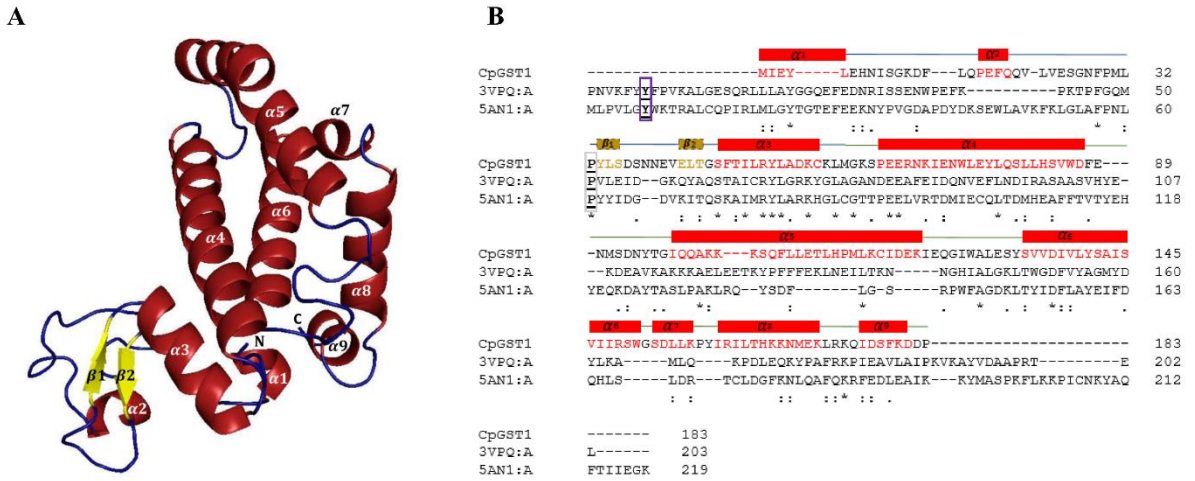


Figure 2.3: *In silico* structural analysis of Vega class representative *Cryptosporidium parvum* glutathione transferase 1 (C*p*GST1). 3D model of C*p*GST1 (A) and its amino acid sequence alignment with templates (B). Secondary structural annotations were done as per modeled structure where α-helices and corresponding amino acids are colored in red while the β-sheets and their corresponding amino acids are colored in yellow. The active-site tyrosine and the *cis*-proline residues are boxed in purple and grey respectively. The template Protein Data Bank codes, 3VPQ-A and 5AN1-A, represents GST protein crystal structures from *Bombyx mori* (Sigma class GST) and *Penaeus vannamei* (Mu class GST).

Structural analysis revealed the presence of 2β-sheets and 3α-helices in the N-terminal region and 6α-helices in the C-terminal region of C*p*GST1 (Figure 2.3). The overall structure of C*p*GST1 at the N-terminal domain seems completely different compared to the canonical GST N-terminal domain (Oakley, 2011, Sheehan *et al*., 2001). The N-terminal region of C*p*GST1 did not have the typical thioredoxin-like fold, nor did it follow the βαβ-α-ββα motif; it was rather composed of two antiparallel β-sheets and 3α-helices (Figure 2.3). It is rare to find GSTs that do not possess the conventional thioredoxin βαβ-α-ββα motif. Kappa class GSTs, which are mitochondrial GSTs, are the closest GSTs that do not follow the traditional thioredoxin fold but have still been found to carry out a similar molecular function as conical GSTs (Atkinson and Babbitt, 2009, Ladner *et al*., 2004, Lallement *et al*., 2014). This is also common for MAPEG GST and the mPGES-1 (microsomal ProstaGlandin E-Synthase type 1) subfamily of proteins, as they too are a group of structurally unrelated proteins with GSH transferase activities

(Frova, 2006, Lallement *et al*., 2014). Because the GST superfamily shares such vast variations in terms of their structural conformation, this ααββα conformation of C*p*GST1 can be considered a unique Vega class feature.

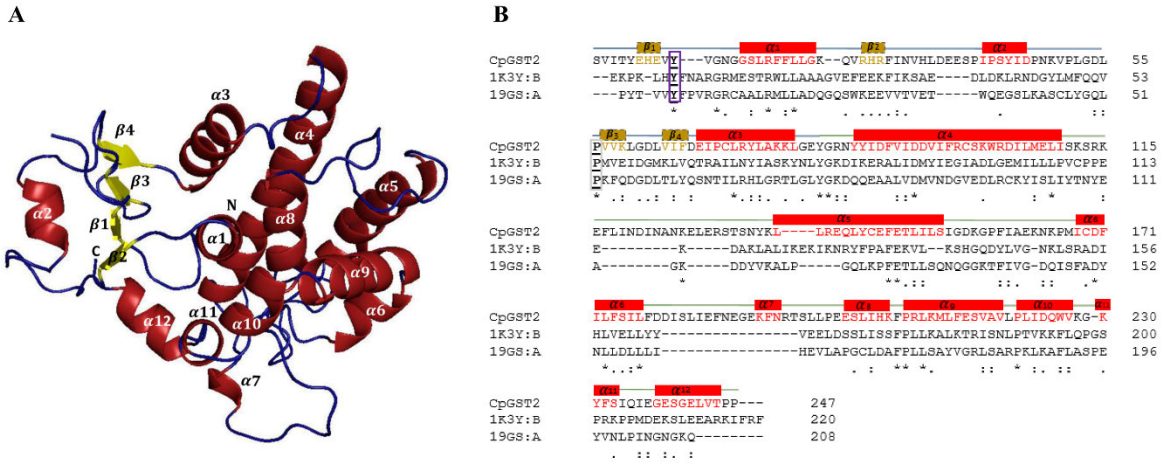


Figure 2.4: *In silico* structural analysis of Gamma class representative *Cryptosporidium parvum* glutathione transferase 2 (C*p*GST2). 3D model of C*p*GST2 (A) and its amino acid sequence alignment with templates (B). Secondary structural annotations were done as per modeled structure where α-helices and corresponding amino acids are colored in red while the β-sheets and their corresponding amino acids are colored in yellow. The active-site tyrosine and the *cis*-proline residues are boxed in purple and grey respectively. The template Protein Data Bank codes, 1K3Y-B and 19GS-A, represents GST protein crystal structures of Alpha class (1K3Y-B) and Pi class (19GS-A) GSTs from humans**.**

In contrast to the C*p*GST1 model, the C*p*GST2 and C*m*GST3 models N-terminal domain follows the thioredoxin-like fold, which is characteristic of cytosolic enzymes in the GST superfamily (Oakley, 2011, Sheehan *et al*., 2001, Wilce and Parker, 1994). The N-terminal domain was complete with 4β-sheets and 3α-helices following a βαβ and ββα arrangement with the two motifs linked by an α2 (Figures 4 and 5). The C-terminal domain contains helices with each model C*p*GST2 and C*m*GST3 having a varying number of helices (Figures 2.4 and 2.5). It has been suggested that an increase in the number of helices in the C-terminal domain, may allow for a broader substrate range and/or offer a deeper catalytic pocket that facilitates the conjugation of larger substrates (Fritz-Wolf *et al*., 2003, Ji *et al*., 1992).

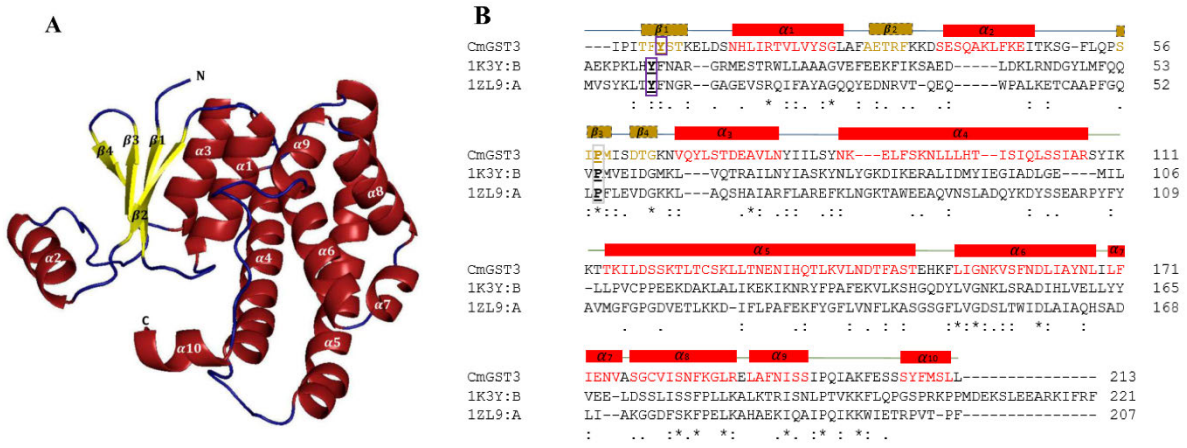


Figure 2.5: *In silico* structural analysis of Psi class representative *Cryptosporidium meleagridis* strain UKMEL1 GST3 glutathione transferase 3 (C*p*GST3). 3D model of C*m*GST3 (A) and its amino acid sequence alignment with templates (B). Secondary structural annotations were done as per modeled structure where α-helices and corresponding amino acids are colored in red while the β-sheets and their corresponding amino acids are colored in yellow. The active-site tyrosine and the *cis*-proline residues are boxed in purple and grey respectively. The template Protein Data Bank codes, 1K3Y-B and 1ZL9-A, represents GST protein crystal structures from Human (Alpha class GST) and *Caenorhabditis elegans* (Pi class GST).

2.5. Conclusion

In this genomic era, *in silico* based comparative studies at genome level or at protein family level have become an important tool to uncover novel aspects in organisms. This study is such an example, where genomes of *Cryptosporidium* species were mined for glutathione transferases (GSTs), enzymes playing a key role in cellular defence and detoxification that are also a potential drug target against pathogens and metabolic disorders. Analysis revealed an interesting feature, namely the presence of two different sizes of GSTs (short and long) in these species. The longer GST proteins were found to be longer than the GSTs found in other organisms, with the size attributed to C- and N-terminal extensions. One of the major findings of the study is the identification of GSTs belonging to three new GST classes in *Cryptosporidium* species. In addition, *Cryptosporidium parvum* GST1 had an atypical thioredoxin fold in the N-terminal region with an αα-ββ-α motif rather than the typical thioredoxin-like fold with a βαβ-α-ββα motif. Future studies includes functional and structural (X-ray or NMR) characterization of *Cryptosporidium* species GSTs. The study results serve as reference for future mining and annotation of GSTs *Cryptosporidium* species.

2.6. References

ABRAHAMSEN, M. S., TEMPLETON, T. J., ENOMOTO, S., ABRAHANTE, J. E., ZHU, G., LANCTO, C. A., DENG, M., LIU, C., WIDMER, G. & TZIPORI, S. 2004. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science,* 304**,** 441-445.

AGARWALA, R., BARRETT, T., BECK, J., BENSON, D. A., BOLLIN, C., BOLTON, E., BOUREXIS, D., BRISTER, J. R., BRYANT, S. H., CANESE, K., CAVANAUGH, M., CHAROWHAS, C., CLARK, K., DONDOSHANSKY, I., FEOLO, M., FITZPATRICK, L., FUNK, K., GEER, L. Y., GORELENKOV, V., GRAEFF, A., HLAVINA, W., HOLMES, B., JOHNSON, M., KATTMAN, B., KHOTOMLIANSKI, V., KIMCHI, A., KIMELMAN, M., KIMURA, M., KITTS, P., KLIMKE, W., KOTLIAROV, A., KRASNOV, S., KUZNETSOV, A., LANDRUM, M. J., LANDSMAN, D., LATHROP, S., LEE, J. M., LEUBSDORF, C., LU, Z., MADDEN, T. L., MARCHLER-BAUER, A., MALHEIRO, A., MERIC, P., KARSCH-MIZRACHI, I., MNEV, A., MURPHY, T., ORRIS, R., OSTELL, J., O'SULLIVAN, C., PALANIGOBU, V., PANCHENKO, A. R., PHAN, L., PIEROV, B., PRUITT, K. D., RODARMER, K., SAYERS, E. W., SCHNEIDER, V., SCHOCH, C. L., SCHULER, G. D., SHERRY, S. T., SIYAN, K., SOBOLEVA, A., SOUSSOV, V., STARCHENKO, G., TATUSOVA, T. A., THIBAUD-NISSEN, F., TODOROV, K., TRAWICK, B. W., VAKATOV, D., WARD, M., YASCHENKO, E., ZASYPKIN, A. & ZBICZ, K. 2018. Database resources of the national center for biotechnology information. *Nucleic Acids Research,* 46**,** D8-D13.

ALLOCATI, N., CASALONE, E., MASULLI, M., CECCARELLI, I., CARLETTI, E., PARKER, M. W. & DI ILIO, C. 1999. Functional analysis of the evolutionarily conserved proline 53 residue in Proteus mirabilis glutathione transferase B1-1. *FEBS Letters,* 445**,** 347-350.

ALLOCATI, N., FEDERICI, L., MASULLI, M. & DI ILIO, C. 2009. Glutathione transferases in bacteria. *The FEBS Journal,* 276**,** 58-75.

ALLOCATI, N., FEDERICI, L., MASULLI, M. & DI ILIO, C. 2012. Distribution of glutathione transferases in Gram-positive bacteria and Archaea. *Biochimie,* 94**,** 588-596.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. Journal of Molecular Biolology, 215**,** 403-10.

AMADI, B., MWIYA, M., SIANONGO, S., PAYNE, L., WATUKA, A., KATUBULUSHI, M. & KELLY, P. 2009. High dose prolonged treatment with nitazoxanide is not effective for cryptosporidiosis in HIV positive Zambian children: a randomised controlled trial. *BMC Infectious Diseases,* 9**,** 195.

ATKINSON, H. J. & BABBITT, P. C. 2009. Glutathione transferases are structural and functional outliers in the thioredoxin fold. *Biochemistry,* 48**,** 11108-11116.

BANKIER, A. T., SPRIGGS, H. F., FARTMANN, B., KONFORTOV, B. A., MADERA, M., VOGEL, C., TEICHMANN, S. A., IVENS, A. & DEAR, P. H. 2003. Integrated mapping, chromosomal sequencing and sequence analysis of *Cryptosporidium parvum*. *Genome Research,* 13**,** 1787-1799.

BHALCHANDRA, S., CARDENAS, D. & WARD, H. D. 2018. Recent breakthroughs and ongoing limitations in *Cryptosporidium* research. *F1000Research,* 7, F1000 Faculty Rev-1380 .

BOARD, P. G. & WEBB, G. C. 1987. Isolation of a cDNA clone and localization of human glutathione S-transferase 2 genes to chromosome band 6p12. *Proceedings of the National Academy of Sciences,* 84**,** 2377-2381.

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**,** W573-9.

CABADA, M. M. & WHITE JR, A. C. 2010. Treatment of cryptosporidiosis: do we know what we think we know? *Current Opinion in Infectious Diseases,* 23**,** 494-499.

CERTAD, G., VISCOGLIOSI, E., CHABÉ, M. & CACCIÒ, S. M. 2017. Pathogenic mechanisms of *Cryptosporidium* and *Giardia*. *Trends in Parasitology,* 33**,** 561-576.

COLOVOS, C. & YEATES, T. O. 1993. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Science,* 2**,** 1511-9.

CONSORTIUM, U. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research,* 47**,** D506-D515.

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**,** W13-7.

EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research,* 32**,** 1792-7.

EISENBERG, D., LUTHY, R. & BOWIE, J. U. 1997. VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymolog,* 277**,** 396-404.

EL-GEBALI, S., MISTRY, J., BATEMAN, A., EDDY, S. R., LUCIANI, A., POTTER, S. C., QURESHI, M., RICHARDSON, L. J., SALAZAR, G. A. & SMART, A. 2019. The Pfam protein families database in 2019. *Nucleic Acids Research,* 47**,** D427-D432.

ELWIN, K., HADFIELD, S. J., ROBINSON, G., CROUCH, N. D. & CHALMERS, R. M. 2012. *Cryptosporidium viatorum* n. sp.(Apicomplexa: Cryptosporidiidae) among travellers returning to Great Britain from the Indian subcontinent, 2007–2011. *International Journal for Parasitology,* 42**,** 675-682.

FAYER, R., SANTÍN, M. & TROUT, J. M. 2008. *Cryptosporidium ryanae* n. sp.(Apicomplexa: Cryptosporidiidae) in cattle (Bos taurus). *Veterinary Parasitology,* 156**,** 191-198.

FAYER, R., SANTÍN, M. & XIAO, L. 2005. *Cryptosporidium bovis* n. sp.(Apicomplexa: Cryptosporidiidae) in cattle (*Bos taurus)*. *Journal of Parasitology,* 91**,** 624-629.

FISER, A. 2010. Template-based protein structure modeling. *Computational Biology.* Springer.

FRITZ-WOLF, K., BECKER, A., RAHLFS, S., HARWALDT, P., SCHIRMER, R. H., KABSCH, W. & BECKER, K. 2003. X-ray structure of glutathione S-transferase from the malarial parasite *Plasmodium falciparum*. *Proceedings of the National Academy of Sciences,* 100**,** 13821-13826.

FROVA, C. 2006. Glutathione transferases in the genomics era: new insights and perspectives. *Biomolecular Engineering,* 23**,** 149-169.

GLISIC, B., MIHALJEVIC, I., POPOVIC, M., ZAJA, R., LONCAR, J., FENT, K., KOVACEVIC, R. & SMITAL, T. 2015. Characterization of glutathione-S-transferases in zebrafish (*Danio rerio)*. *Aquatic Toxicology,* 158**,** 50-62.

GUO, Y., CEBELINSKI, E., MATUSEVICH, C., ALDERISIO, K. A., LEBBAD, M., MCEVOY, J., ROELLIG, D. M., YANG, C., FENG, Y. & XIAO, L. 2015a. Subtyping novel zoonotic pathogen *Cryptosporidium chipmunk* genotype I. *Journal of Clinical Microbiology,* 53**,** 1648-1654.

GUO, Y., TANG, K., ROWE, L. A., LI, N., ROELLIG, D. M., KNIPE, K., FRACE, M., YANG, C., FENG, Y. & XIAO, L. 2015b. Comparative genomic analysis reveals occurrence of genetic recombination in virulent *Cryptosporidium hominis* subtypes and telomeric gene duplications in *Cryptosporidium parvum*. *Bmc Genomics,* 16**,** 320.

HARWALDT, P., RAHLFS, S. & BECKER, K. 2002. Glutathione S-transferase of the malarial parasite *Plasmodium falciparum*: characterization of a potential drug target. *Biological Chemistry,* 383**,** 821-830.

HEIGES, M., WANG, H., ROBINSON, E., AURRECOECHEA, C., GAO, X., KALUSKAR, N., RHODES, P., WANG, S., HE, C.-Z. & SU, Y. 2006. CryptoDB: a *Cryptosporidium* bioinformatics resource update. *Nucleic Acids Research,* 34**,** D419-D422.

HOTEZ, P. J., ALVARADO, M., BASÁÑEZ, M.-G., BOLLIGER, I., BOURNE, R., BOUSSINESQ, M., BROOKER, S. J., BROWN, A. S., BUCKLE, G. & BUDKE, C. M. 2014. The global burden of disease study 2010: interpretation and implications for the neglected tropical diseases. *PLoS Neglected Tropical Diseases,* 8.

IFEONU, O. O., CHIBUCOS, M. C., ORVIS, J., SU, Q., ELWIN, K., GUO, F., ZHANG, H., XIAO, L., SUN, M. & CHALMERS, R. M. 2016. Annotated draft genome sequences of three species of *Cryptosporidium: Cryptosporidium meleagridis* isolate UKMEL1, *C. baileyi* isolate TAMU-09Q1 and *C. hominis* isolates TU502_2012 and UKH1. *FEMS Pathogens and Disease,* 74**,** ftw080.

ISAZA, J. P., GALVAN, A. L., POLANCO, V., HUANG, B., MATVEYEV, A. V., SERRANO, M. G., MANQUE, P., BUCK, G. A. & ALZATE, J. F. 2015. Revisiting the reference genomes of human pathogenic

Cryptosporidium species: reannotation of *C. parvum* Iowa and a new *C. hominis* reference. *Scientific Reports,* 5**,** 16324.

JI, X., ZHANG, P., ARMSTRONG, R. N. & GILLILAND, G. L. 1992. The three-dimensional structure of a glutathione S-transferase from the Mu gene class. Structural analysis of the binary complex of isoenzyme 3-3 and glutathione at 2.2-. ANG. resolution. *Biochemistry,* 31**,** 10169-10184.

JONES, D. T., TAYLOR, W. R. & THORNTON, J. M. 1992. The rapid generation of mutation data matrices from protein sequences. *Computer Applications in Biosciences,* 8**,** 275-82.

JUÁREZ-MARTÍNEZ, A. B., SOTELO-MUNDO, R. R. & RUDIÑO-PIÑERA, E. 2017. Crystal structure of a class-mu glutathione S-transferase from whiteleg shrimp Litopenaeus vannamei: structural changes in the xenobiotic binding H-site may alter the spectra of molecules bound. *Journal of Biochemical and Molecular Toxicology,* 31**,** e21838.

KELLEY, L. A., MEZULIS, S., YATES, C. M., WASS, M. N. & STERNBERG, M. J. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols,* 10**,** 845.

KHOZA, T., MFEKA, S. M. & ACHILONU, I. 2018. Targeting *Cryptosporidium* GST for rational drug discovery against parasitic infectious diseases. *Biochemistry and Molecular Biology Journal,* 4.

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

KOEHLER, A. V., WANG, T., HAYDON, S. R. & GASSER, R. B. 2018. *Cryptosporidium viatorum* from the native Australian swamp rat Rattus lutreolus-An emerging zoonotic pathogen? *International Journal for Parasitology: Parasites and Wildlife,* 7**,** 18-26.

KOTLOFF, K. L., NATARO, J. P., BLACKWELDER, W. C., NASRIN, D., FARAG, T. H., PANCHALINGAM, S., WU, Y., SOW, S. O., SUR, D. & BREIMAN, R. F. 2013. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *The Lancet,* 382**,** 209-222.

KUMAR, S., STECHER, G. & TAMURA, K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology Evolution,* 33**,** 1870-1874.

LADNER, J. E., PARSONS, J. F., RIFE, C. L., GILLILAND, G. L. & ARMSTRONG, R. N. 2004. Parallel evolutionary pathways for glutathione transferases: structure and mechanism of the mitochondrial class kappa enzyme rGSTK1-1. *Biochemistry,* 43**,** 352-361.

LALLEMENT, P.-A., BROUWER, B., KEECH, O., HECKER, A. & ROUHIER, N. 2014. The still mysterious roles of cysteine-containing glutathione transferases in plants. *Frontiers in Pharmacology,* 5**,** 192.

LARSSON, P., WALLNER, B., LINDAHL, E. & ELOFSSON, A. 2008. Using multiple templates to improve quality of homology models in automated homology modeling. *Protein Science,* 17**,** 990-1002.

LASKOWSKI, R., MACARTHUR, M. & THORNTON, J. 2006. PROCHECK: validation of protein-structure coordinates.

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.

LE TRONG, I., STENKAMP, R. E., IBARRA, C., ATKINS, W. M. & ADMAN, E. T. 2002. 1.3-Å resolution structure of human glutathione S-transferase with S-hexyl glutathione bound reveals possible extended ligandin binding site. *Proteins: Structure, Function, and Bioinformatics,* 48**,** 618-627.

LEITCH, G. J. & HE, Q. 2011. Cryptosporidiosis-an overview. *Journal of Biomedical Research,* 25**,** 1-16.

LETUNIC, I. & BORK, P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research,* 47**,** W256-W259.

LIU, S., ROELLIG, D. M., GUO, Y., LI, N., FRACE, M. A., TANG, K., ZHANG, L., FENG, Y. & XIAO, L. 2016. Evolution of mitosome metabolism and invasion-related proteins in *Cryptosporidium*. *BMC Genomics,* 17**,** 1006.

LIU, S., STOESZ, S. P. & PICKETT, C. B. 1998. Identification of a Novel Human GlutathioneS-Transferase Using Bioinformatics. *Archives of Biochemistry and Biophysics,* 352**,** 306-313.

MADEIRA, F., PARK, Y. M., LEE, J., BUSO, N., GUR, T., MADHUSOODANAN, N., BASUTKAR, P., TIVEY, A. R., POTTER, S. C. & FINN, R. D. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Research,* 47**,** W636-W641.

MAUZY, M. J., ENOMOTO, S., LANCTO, C. A., ABRAHAMSEN, M. S. & RUTHERFORD, M. S. 2012. The *Cryptosporidium parvum* transcriptome during in vitro development. *PLoS One,* 7**,** e31715.

MCDONALD, V., DEER, R., UNI, S., ISEKI, M. & BANCROFT, G. 1992. Immune responses to *Cryptosporidium muris* and *Cryptosporidium parvum* in adult immunocompetent or immunocompromised (nude and SCID) mice. *Infection and Immunity,* 60**,** 3325-3331.

MITCHELL, A. L., ATTWOOD, T. K., BABBITT, P. C., BLUM, M., BORK, P., BRIDGE, A., BROWN, S. D., CHANG, H.-Y., EL-GEBALI, S. & FRASER, M. I. 2019. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research,* 47**,** D351-D360.

MÖLLER, S., CRONING, M. D. & APWEILER, R. 2001. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics,* 17**,** 646-653.

NG, J., PAVLASEK, I. & RYAN, U. 2006. Identification of novel *Cryptosporidium* genotypes from avian hosts. *Applied and Environmental Microbiology,* 72**,** 7548-7553.

OAKLEY, A. 2011. Glutathione transferases: a structural perspective. *Drug Metabolism Reviews,* 43**,** 138-151.

OAKLEY, A. J., BELLO, M. L., NUCCETELLI, M., MAZZETTI, A. P. & PARKER, M. W. 1999. The ligandin (non-substrate) binding site of human Pi class glutathione transferase is located in the electrophile binding site (H-site). *Journal of Molecular Biology,* 291**,** 913-926.

PAVLÁSEK, I. 1993. The black-headed gull (Larus ridibundus L.), a new host for *Cryptosporidium baileyi* (Apicomplexa: Cryptosporidiidae). *Veterinární Medicína,* 38**,** 629-638.

PISARSKI, K. 2019. The global burden of disease of zoonotic parasitic diseases: top 5 contenders for priority consideration. *Tropical Medicine and Infectious Disease,* 4**,** 44.

PLJESA-ERCEGOVAC, M., SAVIC-RADOJEVIC, A., MATIC, M., CORIC, V., DJUKIC, T., RADIC, T. & SIMIC, T. 2018. Glutathione transferases: potential targets to overcome chemoresistance in solid tumors. *International Journal of Molecular Sciences,* 19**,** 3785.

PREDIGER, J., HORČIČKOVÁ, M., HOFMANNOVA, L., SAK, B., FERRARI, N., MAZZAMUTO, M. V., ROMEO, C., WAUTERS, L. A., MCEVOY, J. & KVÁČ, M. 2017. Native and introduced squirrels in Italy host different *Cryptosporidium* spp. *European Journal of Protistology,* 61**,** 64-75.

PUMIPUNTU, N. & PIRATAE, S. 2018. Cryptosporidiosis: A zoonotic disease concern. *Veterinary World,* 11**,** 681.

RAO, U., SALINAS, G., MEHTA, K. & KLEI, T. R. 2000. Identification and localization of glutathione S-transferase as a potential target enzyme in Brugia species. *Parasitology research,* 86**,** 908-915.

SALINAS, A. E. & WONG, M. G. 1999. Glutathione S-transferases-a review. *Current Medicinal Chemistry,* 6**,** 279-310.

SATERIALE, A., ŠLAPETA, J., BAPTISTA, R., ENGILES, J. B., GULLICKSRUD, J. A., HERBERT, G. T., BROOKS, C. F., KUGLER, E. M., KISSINGER, J. C. & HUNTER, C. A. 2019. Protective Immunity in a Genetically Tractable Natural Mouse Model of Cryptosporidiosis. *Cell Host and Microbe, 57*.

SAVIOLI, L., SMITH, H. & THOMPSON, A. 2006. *Giardia* and *Cryptosporidium* join the 'neglected diseases initiative'. *Trends in Parasitology,* 22**,** 203-208.

SAVOJARDO, C., MARTELLI, P. L., FARISELLI, P., PROFITI, G. & CASADIO, R. 2018. BUSCA: an integrative web server to predict subcellular localization of proteins. *Nucleic Acids Research,* 46**,** W459-W466.

SCHRODINGER, L. 2010. The PyMOL molecular graphics system. *Version,* 1**,** 0.

SHEEHAN, D., MEADE, G. & FOLEY, V. M. 2001. Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochemical Journal,* 360**,** 1-16.

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.

SORANZO, N., GORLA, M. S., MIZZI, L., DE TOMA, G. & FROVA, C. 2004. Organisation and structural evolution of the rice glutathione S-transferase gene family. *Molecular Genetics and Genomics,* 271**,** 511-521.

SOW, S. O., MUHSEN, K., NASRIN, D., BLACKWELDER, W. C., WU, Y., FARAG, T. H., PANCHALINGAM, S., SUR, D., ZAIDI, A. K. & FARUQUE, A. S. 2016. The burden of *Cryptosporidium* diarrheal disease among children< 24 months of age in moderate/high mortality regions of sub-Saharan Africa and South Asia, utilizing data from the Global Enteric Multicenter Study (GEMS). *PLoS Neglected Tropical Diseases,* 10.

STENBERG, G., BOARD, P. G. & MANNERVIK, B. 1991. Mutation of an evolutionarily conserved tyrosine residue in the active site of a human class Alpha glutathione transferase. *FEBS Letters,* 293**,** 153-155.

STENGER, B. L., CLARK, M. E., KVÁČ, M., KHAN, E., GIDDINGS, C. W., PREDIGER, J. & MCEVOY, J. M. 2015. North American tree squirrels and ground squirrels with overlapping ranges host different *Cryptosporidium* species and genotypes. *Infection, Genetics and Evolution,* 36**,** 287-293.

STRIEPEN, B. 2013. Parasitic infections: time to tackle cryptosporidiosis. *Nature News,* 503**,** 189.

UNI, S., ISEKI, M., MAEKAWA, T., MORIYA, K. & TAKADA, S. 1987. Ultrastructure of *Cryptosporidium muris* (strain RN 66) parasitizing the murine stomach. *Parasitology Research,* 74**,** 123-132.

WANG, W., XIA, M., CHEN, J., DENG, F., YUAN, R., ZHANG, X. & SHEN, F. 2016. Data set for phylogenetic tree and RAMPAGE Ramachandran plot analysis of SODs in *Gossypium raimondii* and *G. arboreum*. *Data Brief,* 9**,** 345-8.

WEBB, B. & SALI, A. 2016. Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics,* 54**,** 5.6.1-5.6.37.

WIDMER, G., CARMENA, D., KVÁČ, M., CHALMERS, R. M., KISSINGER, J. C., XIAO, L., SATERIALE, A., STRIEPEN, B., LAURENT, F. & LACROIX-LAMANDÉ, S. 2020. Update on *Cryptosporidium* spp.: highlights from the Seventh International *Giardia* and *Cryptosporidium* Conference. *Parasite,* 27.

WIEDERSTEIN, M. & SIPPL, M. J. 2007. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research,* 35**,** W407-W410.

WILCE, M. C. & PARKER, M. W. 1994. Structure and function of glutathione S-transferases. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology,* 1205**,** 1-18.

XU, G., GUO, C., SHAN, H. & KONG, H. 2012. Divergence of duplicate genes in exon–intron structure. *Proceedings of the National Academy of Sciences,* 109**,** 1187-1192.

XU, P., WIDMER, G., WANG, Y., OZAKI, L. S., ALVES, J. M., SERRANO, M. G., PUIU, D., MANQUE, P., AKIYOSHI, D. & MACKEY, A. J. 2004. The genome of *Cryptosporidium hominis*. *Nature,* 431**,** 1107-1112.

YAMAMOTO, K., HIGASHIURA, A., SUZUKI, M., ARITAKE, K., URADE, Y., UODOME, N. & NAKAGAWA, A. 2013. Crystal structure of a *Bombyx mori* sigma-class glutathione transferase exhibiting prostaglandin E synthase activity. *Biochimica et Biophysica Acta (BBA)-General Subjects,* 1830**,** 3711-3718.

YANG, J. & ZHANG, Y. 2015. I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Research,* 43**,** W174-W181.

CHAPTER 3:

EXPRESSION AND PURIFICATION OF RECOMBINANT *CRYPTOSPORIDIUM PARVUM* GLUTATHIONE TRANSFERASE

Mbalenhle Sizamile Mfeka[1] and Thandeka Khoza [1]*

Department of Biochemistry, School of Life Sciences, University of KwaZulu-Natal (Pietermaritzburg campus), Scottsville, 3209, KwaZulu-Natal, South Africa; smmfeka850@outlook.com (M.S.M); khozat1@ukzn.ac.za (T.K.)

*Correspondence khozat1@ukzn.ac.za (T.K.)

3.1. Abstract

Glutathione transferases (GSTs) from *Cryptosporidium* species are a newly classified group of proteins thought to be key to the survival of *Cryptosporidium* spp. However not much is known about structure and function of this parasitic GST. GSTs are a family of multifunctional proteins which belong to phase II detoxifaction group of enzyme. They function to catalyze the nucleophilic conjugation of previously reduced glutathione thiol group to the electrophilic substrate converting them to less harmful more soluble compounds. Here we report the recombinant overexpression and purification of a gamma class *Cryptosporidium parvum* GST2 protein using *Escherichia coli*. Several expression vectors were tested for the expression of high yields of soluble GST protein. GST was successfully overexpressed in *E. coli* BL21 (DE3) cells using a pCOLD-derived vector with a molecular chaperon trigger factor and N-terminal His-tag. The expressed protein was purified to homogeneity using affinity and gel filtration chromatography. The gel filtration further revealed the *Cryptosporidium* GST under native conditions are dimeric, much like the pre-existing GSTs. The successful purification of GST paves the way for structural and functional characterization studies, which will be essential to understand the structure and function of these newly identified enzymes.

Keywords: Cryptosporidiosis; *Cryptosporidium parvum*; Glutathione transferase; Gamma Class GST; Expression; pCOLD

3.2. Introduction

Cryptosporidium spp. are an intracellular group of apicomplexan protists, responsible for the contraction of a mild to severe gastro-intestinal disease called cryptosporidiosis (Certad *et al*., 2017). This disease is both zoonotic and anthroponotic in its transmission and poses a large threat on socio economic and developmental growth in struggling third world countries (Xiao and Feng, 2008).

A system present in all the *Cryptosporidium* spp. with great potential for cryptosporidiosis intervention is the detoxification system (Abrahamsen *et al*., 2004). This system is comprised of phase I and phase II drug metabolizing enzymes in conjunction with phase III transporters, which safely detoxify and eliminate a wide range of xenobiotic compounds. Glutathione transferases (GST) are amongst the phase II detoxification enzymes that bacterial and eukaryotic species use in this system (Frova, 2006, Xu *et al*., 2005). They primarily function to catalyse the nucleophilic addition of a reduced glutathione (GSH) thiol group to the electrophilic centres of organic xenobiotic compounds. This allows the compounds to be more soluble and less toxic so to be compartmentalized and eliminated by phase III transporter enzymes or excreted, protecting the cells from cytotoxic compounds in the process (Hayes *et al*., 2005). Depending on the class of GST enzymes, some might have additional functions, serving as isomerases, thiol transferases, peroxidases, being involved in prostaglandin synthesis and other non-catalytic functions (Sheehan *et al*., 2001). In the context of *Cryptosporidium* GST, both the catalytic and non catalytic functions are yet to be established due to this enzyme being recently discovered. Also, from the bioinformatics studies it can be anticipated that the *Cryptosporidium* GSTs would have novel functions due to the extreme divergence in amino acid similarity and structural difference when compared to pre-existing GSTs (Mfeka *et al*., 2020).

Wide genome mining of *Cryptosporidium* spp. had showed that of three GST genes present in this species, *Cp*GST2 was the only GST that had the active site tyrosine in the expected active site position and encompassed the traditional GST structural features (Mfeka *et al*., 2020). This GST was selected for further expression and purification studies as it had a close resemblance of the existing GST classes. Moreover, this specific GST is reported to be expressed in all stages of the parasites lifecycle and associated with a rapid turnover of MDR cases (Liu *et al*., 2001, Mauzy *et al*., 2012). Additionally, the presence of this enzyme in all

stages of the parasites life cycle makes it a suitable drug target to treat the disease at any time once detected (Mauzy *et al*., 2012).

Escherichia coli is a largely popular expression system providing high level production of recombinant proteins for structural genomics and proteomics however, the system does garner a few complications (Wingfield, 2015). The most frequently experienced being the production of recombinant proteins in the form of inclusion bodies (Butt *et al*., 2005). Similar complications were experienced in attempts to fulfil the aim of this study which was to bridge the research gap about these scarcely covered *Cryptosporidium* GSTs through expression and purification. In this study, the GST was cloned into a pET11 vector and pCOLD vectors which contain cold shock proteins and trigger factor chaperones to help protein production and correct folding in a soluble form. This protein was then purified to examine the oligomerization patterns in comparison to pre-existing GSTs.

3.3. Materials and Methods

3.3.1. Materials

Genomic/ plasmid work: Acrylamide, agarose and $MgCl_2$ (Merck, RSA). Agar bacteriological, 1 Kb DNA ladder, gel loading dye, purple, OneTaq® 2x master mix with standard buffer, restriction enzymes BamH1 High Fidelity®, Nde1 and restriction enzyme buffer Cut smart® buffer (New England Biolabs, USA). GeneJET Pmasmid Miniprep Kit, FastRuler middle range DNA ladder and GeneRuler 1Kb DNA ladder (ThermoFisher Scientific, USA). EZ-Vision™ Blue light DNA Dye (VWR Life Sciences, USA).

Protein expression: NaCl (Merck, RSA). Pancreatic digest of casein (tryptone), yeast extract and blue prestained protein standard, broad range (New England Biolabs, USA). Lysozyme (Sigma-Aldrich, USA). The gel filtration standards (Bio-Rad, USA). IPTG and ampicillin (Glentham Life Sciences, UK).

Protein purification: glycine, Imidazole, KCl, K_2PO_4, KH_2PO_4, NaCl, Tris (hydroxymethyl aminomethane) (Merck, RSA). NaOH, $NaPOH_2$ (Sigma-Aldrich, USA).

Protein visualization: bromophenol blue, commassie brilliant blue R250, sodium dodecyl sulfate (SDS), ethylenediaminenetetraacetric acid (EDTA) and glycine (Merck, RSA). Acetic acid, methanol, ethyl acetate, TEMED, 2-mercaptoethanol and BLUeye prestained ladder (Sigma-Aldrich, USA). PageRuler® prestained protein ladder (ThermoFisher Scientific, USA), Ammonium persulfate (Bio-Rad, USA).

All other reagents and chemicals used were of analytical grade.

***Escherichia coli* cell lines:**

E. coli JM109 from Sigma-Aldrich (USA) were used for plasmid storage. *E. coli* BL21 (DE3) from ThermoFisher Scientific (USA) were used for colony PCR and recombinant protein expression.

Constructs:

The genomic GST coding sequence of *Cryptosporidium parvum* CryptoDB (transcript ID: cgd8_2970) was synthesized by GenScript (USA) and cloned into pCOLD1 (TaKaRa) and

pCOLDTF (TaKaRa) within the BamH1 and Nde1 restriction sites. pET11-*Cp*GST was gifted by Dr Ikechukwu which was also synthesized by GenScript.

Equipment: The incubators used were the orbital shaker incubator MRC (RSA) and still incubator from Merck (USA). The vortex used was the Vortex genie 2, purchased from Scientific Industries (USA) and the digital heat block from Beckman was used. The magnetic stirrer was purchased from Labs-Mart (CA). The pH meter Starter 2100 was from OHAUS (UK), precision balance series scale from LABOTEC (RSA) and the endure electrophoresis system was from Labnet International (USA). The power pac used was from Bio-Rad (USA). The centrifuges used were the UV mini 1240 UV-Vis spectrophotometer from SHIMADZU (JPN) and the V-630 Absorbance spectrophotometer from JASCO (USA). The centrifuges used were the spectrafuge 16M from Labnet International (USA), Avanti® J-26 XPI centrifuge from Beckman Coulter (USA). The sonicator and the NanoDrop 200 spectrophotometer was from ThermoFisher Scientific (USA). For PCR the T100™ Thermal cycle from Bio-Rad (USA) was used. For capturing gel images the G:BOX was used from Syngene (INDIA). The columns used for protein purification were the HisTrap FF 5 mL column packed with Ni^{2+} Sepharose 6 Fast Flow purchased from Merck (USA) and the HiPrep™ 16/60 Sephacryl® S-200 HR prepacked gel filtration column from GE-Healthcare (USA). The columns were connected to the ÄKTA start from GE Healthcare (USA). The purification samples were concentrated in the Amicon® Ultra-2 Centrifugal Filters purchased from Sigma-Aldrich (USA).

3.3.2. Methods

3.3.2.1. Preparation of Competent *E. coli* Cells:

A vial of *E. coli* JM109 cells glycerol stocks was spread plated across 2 x YT agar plate (1.6 % (w/v) tryptone, 1 % (w/v) yeast extract, 0.5 % (w/v) NaCl, 1.5% (w/v) agar bacteriological. Without flaming the hockey stick shaped glass rod, the rod was aseptically spread across four other plates to dilute the bacteria to obtain single colonies. An additional sixth 2 x YT agar plate was left un-inoculated as a media sterility control. The plates were then inverted and incubated for 16 h at 37 °C. A single colony was then selected from plate five and inoculated into 50 mL 2 x YT media to be incubated for another 16 h at 37 °C shaking at 200 rotations per minute (rpm). The overnight culture was then diluted into a 1:20 in 2 x YT media and grown to OD_{600} ~ 0.4-0,6. The sample was then centrifuged at 5000 x g for 10 min at 4 °C. The pellet was resuspended in 10 mL $MgCl_2$ (100 mM) and incubated in ice for 4 h. The sample was then

centrifuged again under the same conditions followed by the resuspension of the pellet once again with 1 mL $CaCl_2$ (100 mM) and 1 mL glycerol (80 % v/v). The cells were then sub-aliquoted into sterile sample tubes and stored at -80 °C for further use. The same procedure was executed for *E. coli* BL21 (DE3) cells.

3.3.2.2. Transformation of Gene Constructs Into *E. coli* Cells:

The gene constructs of *Cp*GST were resuspended in miliQ water to get a final concentration of 173 ng/µL of *Cp*GST-pET11a, 264 ng/ µL of *Cp*GST1-*Cp*GST and 251 ng/ µL of pCOLDTF-*Cp*GST.

The constructs were transformed into competent *E. coli* JM109 cells for storage and propagation and into *E. coli* BL21(DE3) for expression. For the transformation process, 20 µL of the competent cells were incubated with 1 µL of the gene constructs in ice for 30 min. The cells were then heat shocked at 42 °C for 90 s with a heating block and cooled rapidly on ice for 2 min. Proceeding this was incubation with nutrient rich media at 37 °C for 1 h, which is used for optimum growth competent cells for efficient transformation. This media was prewarmed SOC media (2% (w/v) tryptone, 0.5% (w/v) yeast extract, 250 mM KCl, 1 M glucose, 2 M MgCl2). Thereafter the cells were spread plated aseptically into 2 x YT selection agar plates containing 50 µg/mL ampicillin and incubated overnight at 37 °C.

3.3.2.3. Plasmid DNA Isolation

After transformation of the gene constructs into *E. coli* JM109 for plasmid storage and propagation, a single colony was inoculated into 5 mL of 2 x YT media supplemented with 50 mg/µL ampicillin. The inoculate was grown overnight at 37°C, 200 rpm. Thereafter the cells were collected through centrifugation at 12 000 x g for 10 min. The plasmid DNA was then harvested using the Thermo Scientific GeneJET Plasmid Miniprep Kit as per the manufacturer instructions. To determine the concentration of the plasmid DNA isolated the NanoDrop spectrophotometer was used.

3.3.2.4. Restriction Digest of Mini Prepped Plasmid Isolates

To confirm the presence of the *Cp*GST insert in the pET and pCOLD vectors that were previously isolated a restriction digest was performed. The plasmid DNA (1 µg) was incubated with BamH1-HF for a single digest and BamH1-HF with Nde1 for a double digest along with the cut smart buffer and milliQ water to make up a volume of 25 µL in sterile sample tubes.

The tubes were incubated for an hour at 37°C. The restriction enzyme digest was terminated by addition of loading dye into digestion reaction. The restriction enzyme products were electrophorized on an agarose gel (0.75 % (w/v) agarose, 40 mM Tris, 20mM acetic acid, 1 mM EDTA, pH 8.3). This was done alongside a 1 Kb DNA ladder at 80V using the Bio-Rad power pac. The gel was visualized using the G box imaging system. The standard curves used for gel analysis are on Appendix C Figure S2.

3.3.2.5. Primer selection

The pCOLD vector primers were used for DNA amplification. The coding sequence of both the pCOLD1 and pCOLDTF were obtained from Genscript. The coding sequences selected for the primers were a few base pairs upstream and downstream the restriction enzyme sites. Primers were obtained from Inqaba Biotechnical Industries. The vectors shared the same primer sequence as taken from the TaKaRa (JPN). Thermocalc which is a confirmation software was used to determine whether self-complementary regions appeared and the annealing temperatures of the primers (Andersson *et al*., 2002). The primers synthesized were as follows:

Table 3.1: pCOLD primers synthesized at Inqaba Biotechnical Industries

Primer name	Sequence 5'-3'	Melting temperature
pCOLDTF Forward	CCACTTTCAACGAGCTGATG	51.78 °C
pCOLD1 Forward	ACGCCATATCGCCGAAAGG	53.25 °C
pCOLD Reverse	GGCAGGGATCTTAGATTCTG	51.78 °C

3.3.2.5.1. Colony PCR

To confirm the presence of the pCOLD-*Cp*GST insert in the transformed *E. coli* BL21(DE3) a colony PCR was used. Five colonies were picked from each of the selection plates containing the newly transformed *E. coli* BL21(DE3) cells with pCOLD1-*Cp*GST and pCOLDTF-*Cp*GST. The colonies were diluted with 10 µL of nuclease free distilled water. Thereafter the Taq 2X Master Mix was used with the components in Table 3.2. The PCR product was run on a 1% agarose gel and was visualized using the G box imaging system. The standard curves used for gel analysis are on Appendix C Figure S3.

Table 3.2: Colony PCR components

Components	Volume (µL)	Final concentration
Forward primer (10 µM)	0.5	0.2 µM
Reverse primer (10 µM)	0.5	0.2 µM
Template DNA	2	< 1 µg
Taq 2X Master Mix	12.5	1 X
Nuclease free water	9.5	—

The template DNA was amplified using the following conditions; 95 °C for 30 sec, followed by 30 cycles of (95 °C for 15 sec, 46 °C for 15 sec, 68 °C for 90 sec) with the final extension being at 68 °C for 5 min. The PCR products were visualised on an agarose gel (1 % (w/v) agarose, 40 mM Tris, 20mM acetic acid, 1 mM EDTA, pH8.3) run alongside a FastRuler middle range DNA Ladder (ThermoFisher Scientific, USA).

3.3.2.6. Recombinant Expression of *Cp*GST

For the recombinant expression of C*p*GST, *E. coli* BL21 (DE3) cells were transformed with pET11a-C*p*GST, pCOLD1-C*p*GST and pCOLDTF-*C*pGST as explained in section 3.2.2. Single colonies were selected to be inoculated overnight in 10 mL of 2 x YT broth supplemented with 50 µg/mL ampicillin (37°C at 200 rpm, 16 h). The overnight cultures were then diluted in 1:100 with fresh 2 x YT broth supplemented with 100 µg/mL ampicillin and incubated once more (37°C at 200 rpm). At mid-log phase A_{600} - (0.6-0.8), the cells with the pET vectors were induced with isopropyl-β-D-1-thiogalactopyranoside (IPTG) at varying concentrations (1 mM, 3 mM and 5 mM) with the induction time varying as well (4h, 8h and 16h). Thereafter the cultures were incubated (37°C at 200 rpm). The cells with the pCOLD vectors at mid-log phase A_{600} - (0.6-0.8), were cold shocked for 30 min in a 15°C water bath. Thereafter the cultures were induced with IPTG to a final concentration of 1 mM. The cells were then incubated for 24 h (15°C at 200 rpm). Before harvesting the cells, 50 mL samples were put aside for sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis and then harvested separately by centrifugation at 6000 x g for 15 min at 4°C. The supernatant was discarded with the pellets stored at -80 °C for the promotion of cell lysis or until they were used for C*p*GST purification.

The samples separated for gel analysis as done by Laemmli (1970) were thawed at room temperature and resuspended in phosphate buffered saline (1X, pH 7.4, 137 mM NaCl, 2.7 mM KCl, 8 mM Na_2HPO_4, and 2 mM KH_2PO_4). Thereafter lysozyme (1 mg/mL) was supplemented to the suspension and incubated for 30 min at 37 °C. The samples were further lysed on ice by sonification (30 sec x 8). The soluble and insoluble fractions were then harvested by centrifugation at 10 000 x g for 30 minutes. The pellet contained the insoluble fraction while the supernatant contained the soluble fraction. The expression samples were then analysed using 12.5 % SDS PAGE as described by Laemmli (1970).

3.3.2.7. Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Prior to loading, all samples were denatured by boiling with reducing treatment (125 mM Tris-HCl, 4 % (w/v) SDS, 20 % (v/v) glycerol, 10 % (v/v) 2-mercaptoethanol) for 3 min and cooled on ice until loaded. The gels were connected to a BIO-RAD PowerPac ™ and protein separation occurred at 40 V, 18 mA per gel in tank buffer (250 mM Tris-HCl, 192 mM glycine, 0.1 (w/v) SDS, pH 8.3). The gels were then stained overnight with staining solution (0.125 % (w/v) coomassie blue R-250, 50 % (v/v) methanol, 10 % (v/v) acetic acid. They were then destained overnight with destaining solution (50 % (v/v) methanol, 10 % (v/v) acetic acid) until background was clear and captured with the G-Box imaging system.

3.3.2.8. Purification

The pellet from expression was thawed on ice. Five grams of the harvested cells were resuspended in 10 mL of buffer 1 (50 mM Tris-HCl, 500 mM NaCl, 15mM Imidazole, 0.02% NaN_3, pH 7.5,) to optimize protein binding to the column. The cell suspension was lysed by sonication on ice (8 x 30 sec) and centrifuged at 10 000 x g for 30 minutes. The soluble fraction (supernatant) was then diluted to 50 mL and subjected to immobilized metal affinity chromatography (IMAC) using the HisTrap FF column. The column was washed with milliQ water and pre-equilibrated with 5 column volumes of buffer 1. The supernatant was then passed through the column followed by 10 column volumes of buffer 1 to wash away any unbound proteins. The bound protein was eluted with increasing concentrations of imidazole starting with buffer 2 (50 mM Tris-HCl, 500 mM NaCl, 50 mM Imidazole, 0.02% NaN_3, pH 7.5). Thereafter buffer 3 (50 mM Tris-HCl, 500 mM NaCl, 150 mM Imidazole, 0.02% NaN_3, pH 7.5) and buffer 4 (50 mM Tris-HCl, 500 mM NaCl, 250 mM Imidazole, 0.02% NaN_3, pH 7.5) were used to elute strongly bound proteins. 1.0 mL fractions were collected. The harvested cell

pellet, supernatant, flow through, wash and eluents were analysed using reducing SDS-PAGE as previously described in section 3.3.2.7. The standard curves used for gel analysis are on Appendix C Figure S4.

3.3.2.9. Size Exclusion Chromatography (SEC)

The IMAC purified protein samples eluted using buffer 4 was pooled together and subjected to buffer exchange in preparation for size exclusion chromatography. The sample was buffer exchanged overnight at 4 °C against buffer 5 (50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 0.02% NaN3). The buffer exchanged sample was then concentrated using centrifugal filters which was spun down at 5000 x g for 30 min. The sephacryl S-200 column (120mL with a flow rate of 0.5mL/min) was washed with 5 column volumes of MilliQ water and equilibrated with a further 5 column volumes of buffer 5. Thereafter 500 µL of the concentrated sample was pumped into the column and further washed with 1.5 column volumes of buffer 5. The eluted proteins were monitored at 280 nm with the absorbance detected through the ÄKTA start. 0.5 mL samples were collected with the peaks analysed using reducing SDS-PAGE as previously described in section 3.3.2.7. The standard curve used for gel analysis are on Appendix C Figure S5.

3.3.2.10. Protein Concentration Determination

The protein concentration was determined spectrophotometrically using Beer-Lamberts Law

$$\boldsymbol{A = \varepsilon_{\lambda} cl} \qquad \textbf{Equation 1}$$

Where ε_{λ} is the molar absorption coefficient at λ, *c* is the concentration and *l* is the pathlength of the light through the solution. The molar absorption coefficient of trigger factor-C*p*GST was determined using equation 2:

$$\boldsymbol{\varepsilon(M^{-1}cm^{-1}) = 5500\,(\sum Trp) + 1340\,(\sum Tyr) + 150\,(\sum Cys)} \qquad \textbf{Equation 2}$$

where $\sum$ Trp) is the sum of the tryptophan residues, $\sum$ Tyr is the sum of the tyrosine residues and $\sum$ Cys is the sum of the cysteine residues all within the protein. The constants are representative of molar absorption coefficients for the mentioned residues. The molar absorption coefficient at 280 nm for C*p*GST with the trigger factor and the cleavage site was determined to be 49530 M^{-1}cm-1.

3.4. Results

Integrating bioinformatics and wet lab experimentation was done for a more focused, well rounded investigation into Gamma C*p*GST2 hereon referred to as C*p*GST. Seeing that the expression, purification and characterization of this protein has never been reported on, here the C*p*GST gene was cloned into various vectors for optimum recombinant expression. Subsequently, pCOLD vectors were then used to express the C*p*GST. Before the expression studies were conducted, the C*p*GST constructs received from GeneScript were confirmed using restriction digest and colony PCR. This step was necessary to ensure that the correct construct was used.

3.4.1. C*p*GST Confirmation

To confirm the presence of the C*p*GST inserts within the respective pCOLD and pET vectors a restriction enzyme digest was performed. All the gene constructs were single digested with BamH1 to confirm the construct size and double digested with BamH1 and Nde1 to release the C*p*GST gene from the vectors. Figure 3.1 shows the undigested plasmid from the three constructs (pET11a-C*p*GST, pCOLD1-C*p*GST and pCOLDTF-C*p*GST) consists of three to four bands of varying intensities as expected for circular DNA taking up multiple conformations on an agarose gel (Snustad and Simmons, 2015). A single digest of the C*p*GST contructs resulted in the band size that corresponds to the size of the plasmid with the C*p*GST insert. For an example, a single digest of pET11a-C*p*GST shows a band size of approximately 7144 bp which equals to size of pET 11a vector (5675 bp) plus a CpGST gene (1299 bp). The double digested C*p*GST constructs with BamH1 and NdeI showed two bands with different sizes. The higher and lower molecular weight bands corresponds to the size of the vector and the C*p*GST gene, respectively (Figure 3.1). Therefore, the results obtained from restriction digests confirmed the presence of the C*p*GST gene in both pET11a and pCOLD derived vectors.

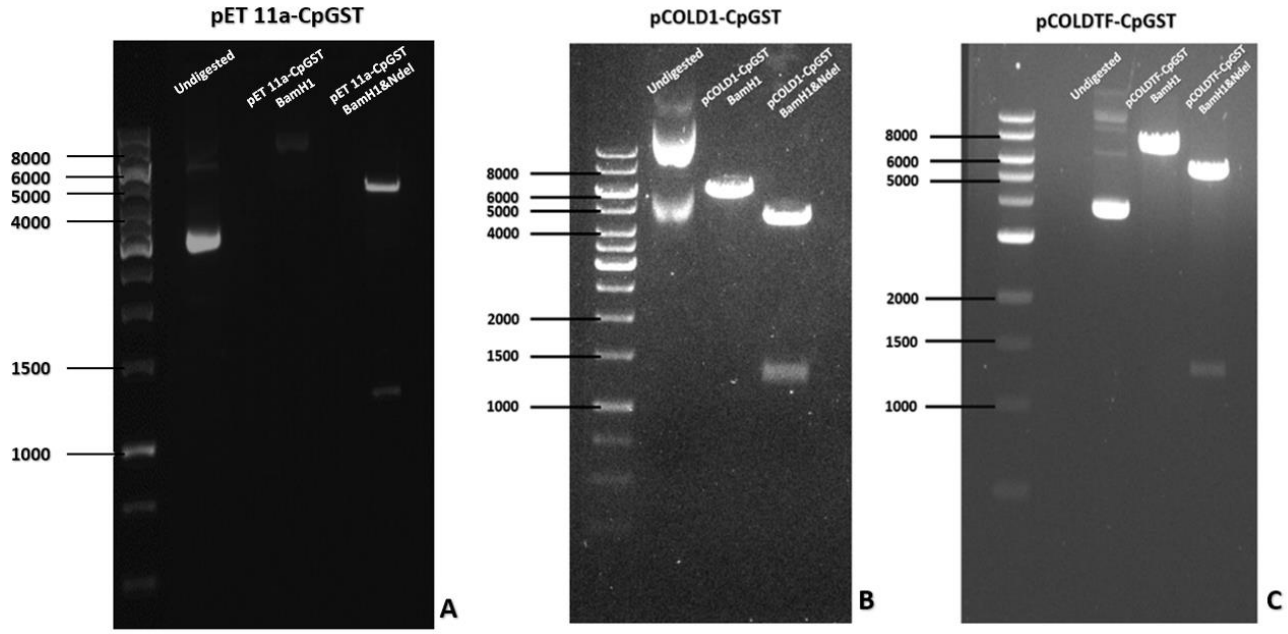


Figure 3.1: Restriction enzyme digests of gene constructs pET11a-C*p*GST, pCOLD1-C*p*GST and pCOLDTF-C*p*GST cut with BamHI and NdeI evaluated on 0.75 % agarose gel. The restriction digested samples of the plasmid isolates of pET 11a-C*p*GST, pCOLD1-C*p*GST and pCOLDTF-C*p*GST shown in A, B and C respectively. The C*p*GST construct were single digested with BamH1 and double digested with BamH1 and Nde1 for 1 h. The samples loaded in the agarose gels were the undigested, BamH1 single digested and the double digested gene contructs with BamH1 and NdeI respectively.

Prior to expression studies, the C*p*GST constructs were transformed to *E.coli* BL21 (DE3) expression cell line and the resulting transformants were screened for the presence of C*p*GST plasmid using colony PCR. Figure 3.2 confirms that the selected transformants contained the plasmid of interest. Figure 3.2a, showed that the insert size in pCOLD1 was 1584 bp while Figure 3.2b, revealed that the size of the insert in pCOLDTF was 1704 bp. The obtained insert size in both pCOLD1 and pCOLDTF were not same. Furthermore, the sizes obtained were larger than the size of the C*p*GST gene which is 1299 bp. The difference in size of insert is attributed to the fact that vector primers were used instead of specific gene primers. The forward primer anneals upstream the trigger factor coding sequence as pointed out in Appendix C Figure S6 and the reverse primer anneals downstream the multiple cloning site thus resulting in the amplification of additional bps flanking the anealation sites. Screening of pET11a-C*p*GST transformant was not done due to the lack of pET forward and reverse primers as well as C*p*GST gene specific primers.

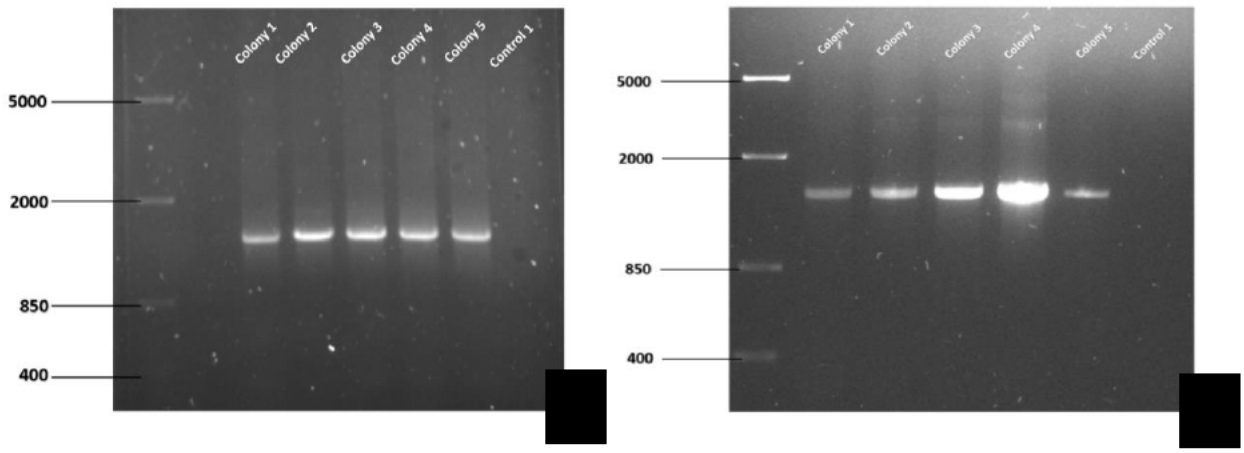


Figure 3.2: Colony PCR of pCOLD1-C*p*GST and pCOLDTF-C*p*GST evaluated on 1 % agarose gel. Colony PCR samples of C*p*GST from their respective vectors from *Escherichia coli*. The samples loaded were the DNA Ladder, different colony PCR products and a no template control. A: C*p*GST in pCold1 with a 1584bp PCR product. B: C*p*GST in pColdTF with the PCR product being 1704 bp. The estimated size of this PCR product is 1704 bp.

3.4.2. Over Expression of Recombinant C*p*GST

Expression studies then proceeded with the protein being recombinantly produced in *E. coli* BL21 (DE3) cells which were all grown in 2 x YT media. Initially, pET11a-C*p*GST was investigated. Various expression conditions were utilized for optimal expression of C*p*GST. Figure 3.3 shows that C*p*GST did not express when cloned into the pET11a vector regardless of the various conditions experimented with being IPTG concentration, temperature and induction time respectively.

In order to overcome the lack of expression a vector that allows for the controlled expression of difficult and somewhat toxic proteins was used. This vector was the pCOLD vector which uses cold shock protein A (cspA) promoter for the controlled expression of proteins at reduced temperatures (Qing *et al.*, 2004). In results not shown, the expression conditions were optimized, with the optimum expression conditions being determined to be 15 °C, using 1 mM IPTG induction for 24 h. The expression of C*p*GST using the pCOLD1 vector was successful, however the recombinant protein was insoluble (Figure 3.4). This was indicated by a highly expressed protein which corresponds to the molecular weight of C*p*GST in the insoluble fraction. This same sized protein was absent in the soluble fraction.

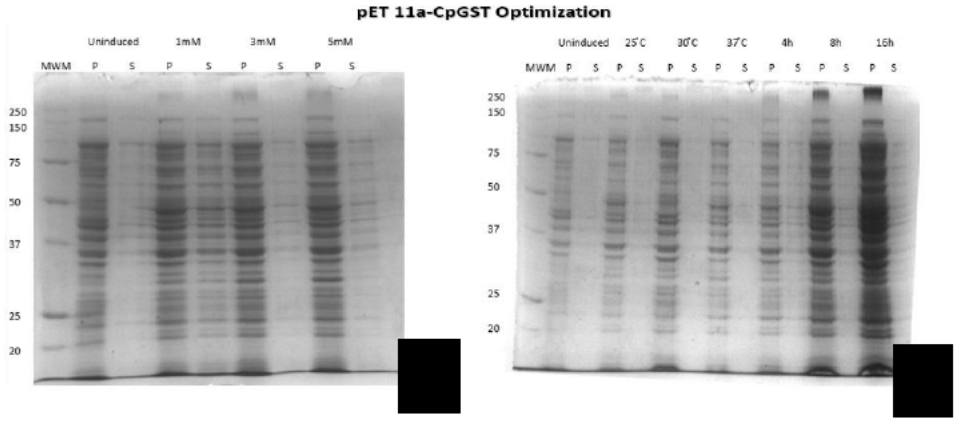


Figure 3.3: 12.5% Reducing SDS-PAGE gel showing expression of C*p*GST transformed into *E. coli* BL21 (DE3) as a function of IPTG concentration, induction temperature and induction time. Expression attempts of C*p*GST in pET11a were performed using a variety of conditions, changing the IPTG concentration, induction temperature and induction time. A: The IPTG concentration was optimized testing 0.1mM, 0.3mM and 0.5mM. B: The induction temperature was optimized growing induced cultures at 25, 30 and 37° C and different induction times of 4h, 8h and 16h. The samples were loaded with equal volumes of reducing treatment boiled for 2 min. The gel was stained with Commassie blue R-250. Over-expression was observed at 8 and 16 hours of induction. The lanes were labelled as (MWM) molecular weight marker; (p) cell pellet; (s) cell supernatant).

Protein purification from inclusion bodies has disadvantages such as obtaining inactive proteins, denatured protein with a considerably lower yield than anticipated with soluble protein products (Wingfield, 2015). As a result, it was reasonable to explore pCOLDTF vector for expression of soluble C*p*GST. Proteins expressed using pCOLDTF are expressed tagged trigger factor. Trigger factor is a 49 kDa molecular chaperone which reported to mediate co-translational folding of the newly synthesized polypeptide in *E.coli* cells resulting in producing correctly folded protein (Saini, 2014). As anticipated, the expression of pCOLDTF-C*p*GST construct resulted in overexpression of a soluble protein at about 100 kDa (Figure 3.4b). This molecular weight size corresponds C*p*GST (~ 49 kDa) tagged to trigger factor (47 kDa). This is further supported by the expression pCOLDTF vector in the absence of C*p*GST which results in overexpression of an ~ 50 kDa protein.

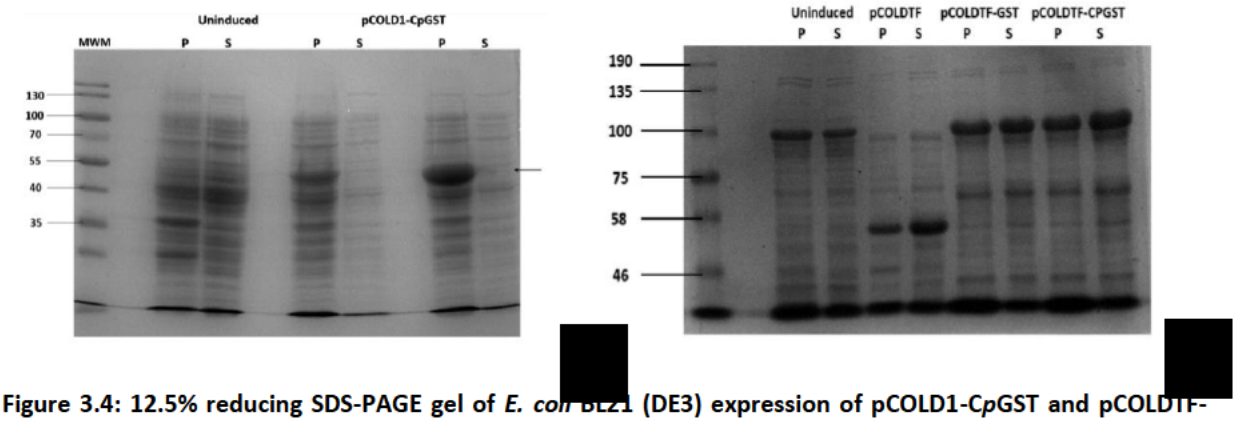


Figure 3.4: 12.5% reducing SDS-PAGE gel of *E. coli* BL21 (DE3) expression of pCOLD1-C*p*GST and pCOLDTF-C*p*GST. Expression studies saw the expression of CpGST in pCOLD vectors. The *E. coli* BL21(DE3) cells were harvested and separated into their insoluble pellet fraction (p) and soluble supernatant fractions (s). A: pCOLD1-C*p*GST expression proteins were shown comparing the uninduced expression sample to the 1 mM IPTG induced expression samples. B: pCOLDTF-C*p*GST expression proteins were shown comparing the uninduced control samples, and the no-insert pCOLDTF expression control to the pCOLDTF-C*p*GST expression samples. The overexpression of a protein was observed ~50 kDa using the pCOLD1-C*p*GST construct. Similarly the overexpression of a protein at ~100 kDa was also observed as expected as the theoretical expressed protein is 49 kDa with a trigger factor of 47 kDa using the pCOLDTF-C*p*GST construct.

3.4.3. C*p*GST Purification

After expression of C*p*GST in pCOLDTF the soluble cell fraction was then purified using affinity chromatography, followed by gel filtration. Initially, the sample was subjected to IMAC purification with the bound proteins being eluted with increasing concentrations of imidazole in the elution buffer. The eluents were then analysed on SDS-PAGE to determine whether C*p*GST constructs could be purified to homogeneity using affinity chromatography.

As Figure 3.5 shows, a great amount of the protein content found in the lysate passed through the column showing very little affinity for the Ni^{2+} resin used. A small amount of the protein of interest was lost in the unbound fraction and with 15 mM immadozole column wash. At this concentration (15 mM immadozle) a majority of proteins less than 55 kDa in size were also washed away from the column. The bound protein was eluted with increasing imidazole concentrations (50-250 mM) in the elution buffer. Figure 3.5 shows that the majority of C*p*GST eluted at 150 mM as indicated by band at approximately 100 kDa. However, C*p*GST co-eluted with multiple proteins of various sizes. At 250 mM imidazole a C*p*GST sample with fewer contaminants was obtain, free of the contaminants previously seen in the previous samples. The proteins eluted however, were low in quantity and concentration as evidenced by the size of the peak and the slim band seen in the gel.

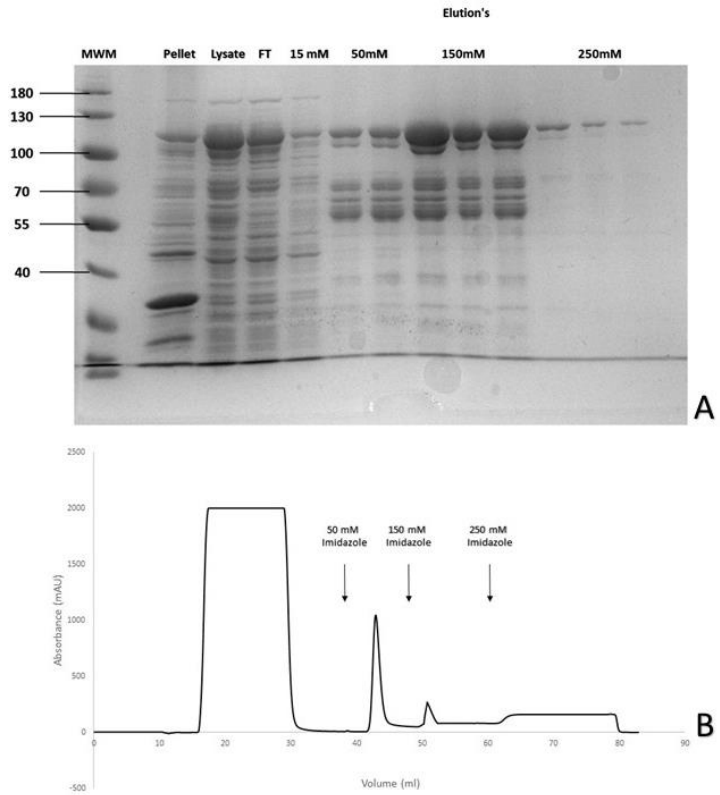


Figure 3.5: Reducing SDS-PAGE analysis of pCOLDTF-C*p*GST affinity chromatography purification. The soluble expression sample was purified using a HisTrap FF column and the bound protein was eluted using an imidazole containing buffer. The proteins eluted had two prominent proteins which were approximately 100 kDa in size. A: SDS-PAGE gel of IMAC purification. The lanes were labelled as (MWM) molecular weight marker; (FT) flow through and the elution buffers (50 mM Tris-HCl, 500 mM NaCl, 0.02% NaN3, pH 7.5) with increasing concentrations of imidazole from 15-250 mM. B: Elution profile of expressed C*p*GST captured using the ÄKTA start.

To further purify the C*p*GST which eluted from the affinity column, the eluent collected at 250 mM imidazole was pooled and concentrated to 1.5 mL and loaded into the gel filtration column. The gel filtration elution profile in Figure 3.6b shows two peaks. The molecular weight of the protein/s in the first peak is ~ 200 kDa. This size correlates with the theoretical size of a dimerized C*p*GST tagged with the trigger factor (47 kDa). In fact when this sample was ran reducing SDS-page gel, interestingly two bands of about similar molecular weight (100 kDa) were observed. The larger of the two proteins band suggests that C*p*GST under native condition forms a homodimer resulting in the 200 kDa sized protein. It is unclear at this stage whether the smaller sized protein also ~100 kDa is a contaminating protein or a partially degraded C*p*GST protein. On the other hand, the second peak detected was of ~ 66.4 kDa

proteins similar to that of the co-eluted contaminant seen in Figure 3.5a. The SDS-PAGE gel showed that the remaining proteins which were co-eluted with C*p*GST in the affinity column were purified further from the recombinant protein. This lead to the C*p*GST purified to homogeneity.

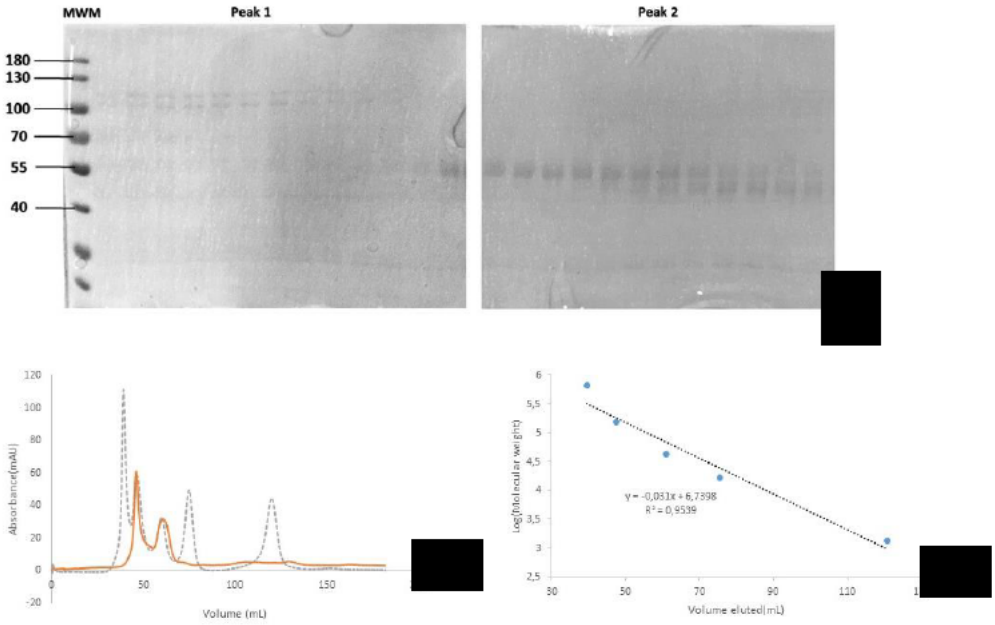


Figure 3.6: SEC elution profile of trigger factor tagged C*p*GST. The column was equilibrated with 50 mM Tris HCl at pH 7.0, 500 mM NaCl and 0.02% NaN3. The flow-rate was 0.500 ml/min and the elution was monitored by absorbance spectroscopy at A_{280}. A 12.5 % reducing SDS-PAGE of the IMAC proteins eluted in SEC showing proteins in peak 1 and 2 : B: The size exclusion profile of C*p*GST-TF (orange) alongside the standard proteins of known sizes as the gel filtration protein markers (grey). C: The size exclusion standard curve with bovine thryglobulin (670 kDa), bovine g-globulin (158 kDa), chicken ovalbumin (44 kDa), horse myoglobulin (17 kDa) and vitamin B12 (1.35 kDa).

To further determine the purity of peak 1 from dust and non protein specimen, a UV absorption scan was done between 240 nm to 340 nm, taken in accumulations of 3. The blank used was the imidazole containing elution buffer. The UV scan seen in Figure 3.7 revealed a peak at 277 nm as expected and no peaks thereafter. This confirmed the absence of UV absorbing impurities and specs of unwanted particles, or copurifying cofactors within the purified protein.

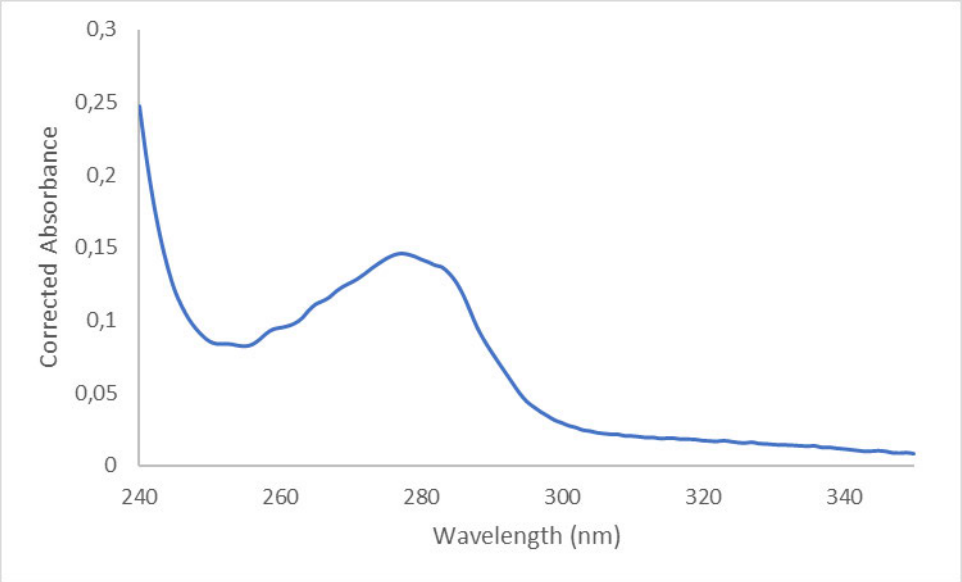


Figure 3.7: UV Scan of IMAC purified pCOLDTF-C*p*GST within a range of 240-340 nm. The purified protein was initially diluted (1:50) followed by the scan being taken across a 240-340 nm range to determine any impurities. The highest peak observed was at 277 nm and no other peaks were seen thereafter**.**

3.4.4. Protein Concentration Determination

Once the purity was confirmed, the protein concentration was determined. Typically a Bradford assay could have been used. However to get a true quantification, clear of any bias or possible false positives, the predicted molar extinction coefficient at A_{280} was used. A single microcentrifuge tube containing the eluted sample was used for the duration of the project. A serial dilution of the protein was performed, and the absorbance of the samples were taken in accumulations of three at 280 nm. The absorbance values were corrected against the blank and the plot seen in Figure 3.8 was used to determine the concentration of trigger factor tagged C*p*GST. The calculation took into consideration the trigger factor tagged C*p*GST and determined the concentration of C*p*GST to be 137 µM.

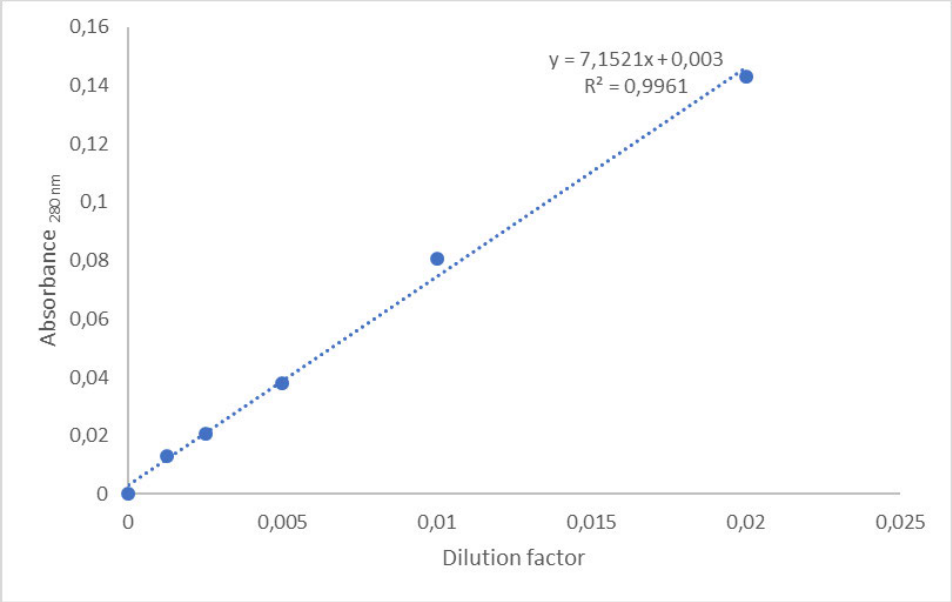


Figure 3.8: Concentration determination of purified C*p*GST with the trigger factor

3.5. Discussion

Glutathione transferases are a promising set of proteins to target in attempts to treat cryptosporidiosis. Their detoxification abilities enables the parasitic cells to reduce a wide range of xenobiotic compounds to less toxic, less soluble substances whilst continuously propagating within the host (Frova, 2006). Although there is an abundance of information of GSTs, much is yet to be discovered about *Cryptosporidium* GSTs. Thus cloning, expression and purification techniques were used in this study to reveal more about the *Cryptosporidium parvum* Glutathione transferase 2 protein found to bear the most resemblance to pre-existing GSTs.

The C*p*GST gene was then cloned into three vectors, namely pET11a, pCOLD1 and pCOLDTF to obtain the best system suitable for C*p*GST expression. Both PCR and restriction digest experiments successfully confirmed the presence of the inserts in the respective vectors (Figure 3.1 and 3.2). In both experiments, the gene size obtained was consistent with the theoretical C*p*GST gene reported in the *Cryptosporidium* database (http://CryptoDB.org) and expasy (Gasteiger *et al*., 2005).

The C*p*GST protein was then recombinantly expressed in the before mentioned pET, pCOLD1 and pCOLFTF vectors. pET vector systems were initially utilized as they are reported to be one of the most powerful, efficient systems for recombinant expression in *E. coli* cells (Bernaudat *et al*., 2011). The utilization of the T7 RNA polymerase promoter within these vectors allows for the tight regulation of high yield transcripts (Studier *et al*., 2009). The expression trials carried out using the pET11a-C*p*GST construct did not yield a recombinant C*p*GST product. Generally, when no protein can be detected after expression trials, protein toxicity is most often thought to be the cause (Young *et al*., 2012). This occurs when the recombinant protein performs functions which can be harmful to the host cell either before or after induction. If these functions are disruptive to the homeostasis or proliferation of the host cell, either slower growth rate, low cell density or death will be observed. Though rare, this typically occurs in *E. coli* BL21 (DE3) cells and could be the reason as to why no expression occurred (Mierendorf *et al*., 2000, Saccardo *et al*., 2016). Alternatively, the lack of expression is often seen when the metabolic activity of the host is not favourable for target protein, stressing the expression system so much so that insoluble proteins are produced in the form of aggregates

(Mierendorf *et al*., 2000). This often leads to the denaturation of the target protein if not packaged in inclusion bodies.

In instances where protein production is not successful even after various optimization attempts, alternative vectors, promoters, cell expression lines and even co-expression with chaperones through fusion tags are often used. In the case of this study, pCOLD vectors were used to aid in successful protein expression due to other avenues not being available or affordable. The existence of cold shock proteins (Csp) such as CspA within these vectors allows for high protein yield after their induction at lower temperatures (Qing *et al*., 2004). Their function as RNA chaperones allowed for regulated transcription and translation at lower rates resulting in successful production of recombinant C*p*GST (Jiang *et al*., 1997, Saini *et al*., 2014). Although the protein produced was insoluble, the size of the protein expressed in the pCOLD1 vector was ~ 50 kDa as anticipated from the theoretical size. Although pCOLD1 allowed for the recombinant expression of C*p*GST, recovering a protein from inclusion bodies can be often be cumbersome as it could result in a denatured, inactive protein product after resolubilization (Georgiou and Valax, 1996). Retrieving an insoluble protein from the aggregated state also reduces the recovered yield one has to work with for characterization purposes (Wingfield, 2015). Additionally the purification steps taken to carefully extract and purify the protein of interest are rather long and do not guarantee an enzymatically active protein product after resolubilization (Trimpin, 2009). Another disadvantage when working with proteins in inclusion bodies is the steps, material and cost of denaturing and unfolding the protein for the purposes of refolding in the correct conformation or partially folding the protein can often times be ineffective and time wasting (Wingfield, 2015). The further analysis and purification of soluble, correctly folded protein is often preferred.

To avoid the previously mentioned complications experienced with insoluble proteins, a pCOLDTF vector was used which utilizes the trigger factor chaperone to assist in the production of a soluble protein. The trigger factor itself assists and protects the nascent chain with long hydrophobic stretches during protein production and the beginning folding stages whilst also fast tracking peptidylpropyl cis-trans isomerization (Saini *et al*., 2014). As the recombinant protein is made, the chaperones are constantly recruited to shield the polypeptides on the ribosome through productive *de novo* folding all to prevent degradation

and aggregation (Hoffmann *et al*., 2006). This lead to the soluble production of C*p*GST with the chaperone attached in Figure 3.4b.

The over-expressed protein was then easily purified from the other bacterial proteins though affinity chromatography. This exploited the affinity that the his-tagged proteins have on the Ni^{2+} ions in the column. Most of the bacterial proteins either flowed through the column or were washed away with low concentrations of imidazole as non specific binding was observed at imidazole concentration lower than 250 mM (Bornhorst and Falke, 2000). This occurrence is anticipated as at least 2% of all protein residues are histidine. Furthermore few proteins can have multiple adjacent histidine residues which could co-elute with the C*p*GST and be considered as contamination (Schmitt *et al*., 1993). To displace the proteins with stronger affinity to Ni^{2+}, higher concentrations were used resulting in a protein with fewer conaminants being eluted at 250 mM imidazole. The protein of interest was also eluted at lower concentrations of imidazole indicating various affinities the protein has to the Ni^{2+}.

The protein eluted at 250 mM imidazole in the IMAC column was then concentrated and subjected to SEC to separate C*p*GST from contaminants while confirming the size obtained from the SDS-PAGE gels. The two peaks obtained from the SEC pointed to the separation of two proteins of different sizes. Due to the fact that the gel filtration column was run under native conditions, the protein sizes expected for GSTs which typically exist as dimers would be doubled. As anticipated the first peak containing the protein of interest was double the GST monomer size. This could be due to dimerization that GSTs typically undergo in their native state (Sheehan *et al*., 2001). Another explanation for the oligomerization is presence of free GSH and other xenobiotic compounds within the expression environment that activates the conjugation of GSH to the foreign compound, leaving the enzyme in an active dimer form (Tripathi, 2007). This would not be reflected in the SDS page gel because of the denaturing effects of the buffers used and the heating of the protein samples in gel sample preparation (Laemmli, 1970).

3.6. Conclusion

This is the first report of on C*p*GST2 expression and purification. In this study, C*p*GST2 protein was not successfully expressed using the pET 11a vector and was expressed as an insoluble protein in pCOLD1 vector. Using pCOLDTF expression vector, C*p*GST2 was successfully expressed in a soluble form as 102 kDa protein. This protein was tagged to the trigger factor

chaperone which assisted the C*p*GST2 protein to fold into its native conformation. Future studies would thereafter encompass the removal of the trigger factor so to explore crystallization studies and simultaneously characterize the GST protein to determine Gamma class GST features.

3.7. References

ABRAHAMSEN, M. S., TEMPLETON, T. J., ENOMOTO, S., ABRAHANTE, J. E., ZHU, G., LANCTO, C. A., DENG, M., LIU, C., WIDMER, G. & TZIPORI, S. 2004. Complete genome sequence of the apicomplexan, *Cryptosporidium parvum*. *Science,* 304**,** 441-445.

ANDERSSON, J.-O., HELANDER, T., HÖGLUND, L., SHI, P. & SUNDMAN, B. 2002. Thermo-Calc & DICTRA, computational tools for materials science. *Calphad,* 26**,** 273-312.

BERNAUDAT, F., FRELET-BARRAND, A., POCHON, N., DEMENTIN, S., HIVIN, P., BOUTIGNY, S., RIOUX, J.-B., SALVI, D., SEIGNEURIN-BERNY, D. & RICHAUD, P. 2011. Heterologous expression of membrane proteins: choosing the appropriate host. *PloS One,* 6**,** e29191.

BORNHORST, J. A. & FALKE, J. J. 2000. [16] Purification of proteins using polyhistidine affinity tags. *Methods in Enzymology*, 326, 245-254.

BUTT, T. R., EDAVETTAL, S. C., HALL, J. P. & MATTERN, M. R. 2005. SUMO fusion technology for difficult-to-express proteins. *Protein Expression and Purification,* 43**,** 1-9.

CERTAD, G., VISCOGLIOSI, E., CHABÉ, M. & CACCIÒ, S. M. 2017. Pathogenic mechanisms of *Cryptosporidium* and *Giardia*. *Trends in Parasitology,* 33**,** 561-576.

FROVA, C. 2006. Glutathione transferases in the genomics era: new insights and perspectives. *Biomolecular Engineering,* 23**,** 149-169.

GASTEIGER, E., HOOGLAND, C., GATTIKER, A., WILKINS, M. R., APPEL, R. D. & BAIROCH, A. 2005. Protein identification and analysis tools on the ExPASy server. *The proteomics protocols handbook. Springer*, 571-607.

GEORGIOU, G. & VALAX, P. 1996. Expression of correctly folded proteins in *Escherichia coli*. *Current Opinion in Biotechnology,* 7**,** 190-197.

HAYES, J. D., FLANAGAN, J. U. & JOWSEY, I. R. 2005. Glutathione transferases. *Annual Review Pharmacolog and Toxicology,* 45**,** 51-88.

HOFFMANN, A., MERZ, F., RUTKOWSKA, A., ZACHMANN-BRAND, B., DEUERLING, E. & BUKAU, B. 2006. Trigger factor forms a protective shield for nascent polypeptides at the ribosome. *Journal of Biological Chemistry,* 281**,** 6539-6545.

JIANG, W., HOU, Y. & INOUYE, M. 1997. CspA, the major cold-shock protein of *Escherichia coli*, is an RNA chaperone. *Journal of Biological Chemistry,* 272**,** 196-202.

LAEMMLI, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* 227**,** 680-685.

LIU, J., CHEN, H., MILLER, D. S., SAAVEDRA, J. E., KEEFER, L. K., JOHNSON, D. R., KLAASSEN, C. D. & WAALKES, M. P. 2001. Overexpression of glutathione S-transferase II and multidrug resistance transport proteins is associated with acquired tolerance to inorganic arsenic. *Molecular Pharmacology,* 60**,** 302-309.

MAUZY, M. J., ENOMOTO, S., LANCTO, C. A., ABRAHAMSEN, M. S. & RUTHERFORD, M. S. 2012a. The *Cryptosporidium parvum* transcriptome during in vitro development. *PLoS One,* 7**,** e31715.

MFEKA, M. S., MARTÍNEZ-OYANEDEL, J., CHEN, W., ACHILONU, I., SYED, K. & KHOZA, T. 2020. Comparative analyses and structural insights of new class glutathione transferases in *Cryptosporidium* species. *Scientific Reports,* 10**,** 1-12.

MIERENDORF, R. C., MORRIS, B. B., HAMMER, B. & NOVY, R. E. 2000. Expression and purification of recombinant proteins using the pET system. *The Nucleic Acid Protocols Handbook.* Springer, 947-977.

QING, G., MA, L.-C., KHORCHID, A., SWAPNA, G., MAL, T. K., TAKAYAMA, M. M., XIA, B., PHADTARE, S., KE, H. & ACTON, T. 2004. Cold-shock induced high-yield protein production in *Escherichia coli*. *Nature Biotechnology,* 22**,** 877-882.

SACCARDO, P., CORCHERO, J. L. & FERRER-MIRALLES, N. 2016. Tools to cope with difficult-to-express proteins. *Applied Microbiology and Biotechnology,* 100**,** 4347-4355.

SAINI, P., WANI, S. I., KUMAR, R., CHHABRA, R., CHIMNI, S. S. & SAREEN, D. 2014. Trigger factor assisted folding of the recombinant epoxide hydrolases identified from *C. pelagibacter* and *S. nassauensis*. *Protein Expression and Purification,* 104**,** 71-84.

SCHMITT, J., HESS, H. & STUNNENBERG, H. G. 1993. Affinity purification of histidine-tagged proteins. *Molecular Biology Reports,* 18**,** 223-230.

SHEEHAN, D., MEADE, G. & FOLEY, V. M. 2001. Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochemical Journal,* 360**,** 1-16.

SNUSTAD, D. P. & SIMMONS, M. J. 2015. *Principles of genetics*, John Wiley & Sons.

STUDIER, F. W., DAEGELEN, P., LENSKI, R. E., MASLOV, S. & KIM, J. F. 2009. Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and BL21 (DE3) and comparison of the *E. coli* B and K-12 genomes. *Journal of Molecular Biology,* 394**,** 653-680.

TRIMPIN, S. & BRIZZARD, B. 2009. Analysis of insoluble proteins. *Biotechniques,* 46**,** 409-419.

TRIPATHI, T., RAHLFS, S., BECKER, K. & BHAKUNI, V. 2007. Glutathione mediated regulation of oligomeric structure and functional activity of *Plasmodium falciparum* glutathione S-transferase. *BMC Structural Biology,* 7**,** 1-10.

WINGFIELD, P. T. 2015. Overview of the purification of recombinant proteins. *Current Protocols in Protein Science,* 80**,** 6.1.1-6.1.35.

XIAO, L. & FENG, Y. 2008. Zoonotic cryptosporidiosis. *FEMS Immunology & Medical Microbiology,* 52**,** 309-323.

XU, C., LI, C. Y.-T. & KONG, A.-N. T. 2005. Induction of phase I, II and III drug metabolism/transport by xenobiotics. *Archives of Pharmacal Research,* 28**,** 249.

YOUNG, C. L., BRITTON, Z. T. & ROBINSON, A. S. 2012. Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. *Biotechnology Journal,* 7**,** 620-634.

CHAPTER 4:

CONCLUDING REMARKS

4.1. Closing Remarks

Cryptosporidiosis is classified as a neglected gastrointestinal disease with worldwide impact on immune compromised children and adults. However, as mentioned there is only one U.S. Food and Drug Administration (FDA) approved drug available for its treatment (Certad *et al*., 2017, Shirley *et al*., 2012,). This drug is called nitazoxanide and is ineffective in immune compromised patients. In addition, there are also no vaccines available or efficient preventative strategies put in place to resolve cryptosporidiosis. The few preventative measures taken to reduce the impact the disease has on the largely disadvantaged communities are adequate access to water supplies, improved sanitation and appropriate health education (Feasey *et al*., 2009). These attempts however are unfortunately short term and do not adequately reduce the global burden cryptosporidiosis causes. The absence of effective non-discriminatory treatments or vaccines for cryptosporidiosis highlights the urgency for the development of therapeutic intervention from this disease.

Glutathione transferases which are phase II detoxification enzymes have been identified as a therapeutic target in *Cryptosporidium* species (Mauzy *et al*., 2012) . In other apicomplexan species, detoxification proteins have been identified and exploited successfully as a therapeutic target (Fritz-Wolf *et al*., 2003). These proteins have been scarcely researched in *Cryptosporidium* species with their molecular and biophysical characteristics being unknown. The current study took to bridge that knowledge gap taking advantage of bioinformatics tools and molecular techniques such as protein expression and purification.

Each of the 15 *Cryptosporidium* species mined possessed three GST genes, of which two thirds were full length proteins. The cellular localization of these full length proteins were determined to be cytosolic in nature as opposed to being mitochondrial or membrane-associated proteins involved in eicosanoid and gluthatione metabolism. Cytosolic GSTs are said to be the most ancient within this group (Jemth and Mannervik, 1997). When determining the similarities found within the *Cryptosporidium* GSTs phylogenetically, the phylogenetic tree showed the GST separated into three different clades. The GSTs within each

clade had the following: i) GSTs with similar amino acid lengths. ii) One *Cryptosporidium* species within each clade, with the exception of group 3. Not a single clade had two proteins from one species. iii) The GSTs within each clade shared 42-100 % amino acid identity whilst the percentage identity between each clade dropped to as low as 10%. These divisions pointed to the notion of the *Cryptosporidium* GSTs not belonging to the same class. Based on the fact that GSTs within each class share secondary structural traits, immunological cross reactivity and sensitivity to inhibitors, determining the class that the *Cryptosporidium* GST belonged to might share further insight into these traits. However, the phylogenetic tree built with pre-existing GSTs showed the separation of the GST clades, highlighting the novelty of the *Cryptosporidium* GSTs. These results concluded that the separate clades formed three new classes of GSTs which were denoted Vega, Gamma and Psi class GSTs. Similarly, the *Plasmodium falciparum GST* which could not be assigned to any previously known classes, and thus designated its own class (Fritz-Wolf *et al*., 2003).

This was an exciting discovery, however to further investigate the structural elements the classes possessed, homology models were built. This would additionally reveal any class specific variations the novel GSTs had from the conical GST's. One GST from each class was chosen to be a class representative for structural features. The modelled GSTs all had the highly conserved proline residue that forms parts of the *cis*-Pro loop (Allocati *et al*., 1999). This is responsible for connecting the N- and C-terminal regions to maintain structural integrity. The active side tyrosine which is also usually conserved in GSTs was not found in the expected position in the Vega and Psi class GSTs. The Vega class GST was found to have an atypical N-terminal domain with only two beta sheets instead of the expected four found in most GST proteins. The Gamma and Psi class GSTs had the typical thioredoxin like fold with the N-terminal domain containing 4 β sheets and 3 α helices. The C-terminal domain for all three *Cryptosporidium* GST classes was all helical with much variation in sequence identity and secondary structure conformation to allow for the conjugation of various xenobiotic compounds (Frova, 2006).

To obtain further physiological information on *Cryptosporidium* GSTs, the least novel GST was expressed and purified. It was anticipated that the Gamma class GST had an open reading frame of which comprised of 1299 bp encoding a protein of 429 amino acids with a calculated

molecular size of 49 kDa. This size difference seen between the gamma class GST and pre-existing GST was attributable to the N-terminal and C-terminal extensions seen in Figure 4.1.

Figure 4.1: Schematic diagram of the full length C*p*GST with the InterPro and Pfam confirmed GST domains and the N- and C-terminal extensions.

The presence of the Gamma class *Cryptosporidium parvum* GST2 (C*p*GST) gene was successfully confirmed in the expression vectors selected for this study using PCR and restriction digest. Thereafter the three gene constructs were screened to determine the most efficient vector for the recombinant expression of C*p*GST. The pET 11 was not successful in the expression of the recombinant C*p*GST possibly due to protein toxicity (Young *et al*., 2012). The pCOLD1 vector was also unsatisfactory as a protein of the expected size (~ 50 kDa) was expressed in insoluble form. It is worth noting that this size corresponds to the theoretical 49 kDa value previously mentioned. This is double the size of traditional cytosolic GSTs which range from 24-29 kDa (Sheehan *et al*., 2001). This was yet another indication of the novelty of this new GST class. The pCOLDTF vector was then used which exploited the cold shock protein and trigger factor molecular chaperone which slowed protein translation in low temperatures and facilitated protein formation in the correct conformation (Qing *et al*., 2004, Saini *et al*., 2014). This resulted in a ~ 100 kDa recombinant protein due to expression of C*p*GST tagged with trigger factor which is 47 kDa.

The recombinant C*p*GST was purified successfully to homogeneity using two chromatographic techniques namely affinity chromatography and size exclusion chromatography. The affinity chromatography eluent was semi pure and thus further purified using size exclusion chromatography. The native conditions in which this purification step was performed allowed for the protein to be eluted as a dimer. The protein homogenous protein was eluted at double ~ 200 kDa which equated the size of two C*p*GST monomers each tagged with a trigger factor.

To our knowledge, these studies mentioned were the first to report the novel *Cryptosporidium* GST classes, and a recombinantly expressed and purified C*p*GST of this size (49 kDa).

Further studies for this project would be the thrombin cleavage of the molecular chaperone trigger factor from C*p*GST. Due to the similar sizes, the C*p*GST would then be purified from the trigger factor using ion exchange chromatography or affinity chromatography once again. The N-terminal trigger factor would remain bound to the Ni^{2+} column while the C*p*GST was eluted in the unfound fraction. The protein would then be free to be characterized and examined for enzyme activity.

To fully understand the role that the N-terminal and C-terminal extensions have on the C*p*GST the enzyme functionality, a range of experiments can be done. The designing of deletion mutants could be carried out to determine the effect the extensions have on the GST activity. Three mutants would be designed as depicted in Figure 4.2. Once these mutants have been successfully expressed and purified, they can be crystalized alongside the full-length C*p*GST to elucidate the structure of the protein. Following elucidation of theses structures, C*p*GST inhibitor screening and design experiments can thus begin through molecular modelling and docking of various ligands.

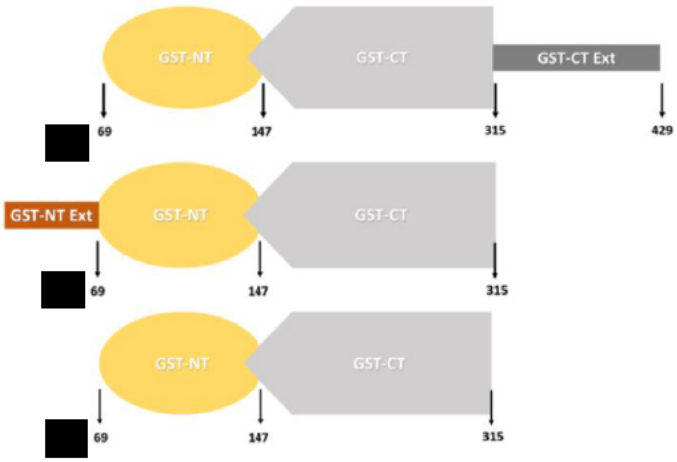


Figure 4.2: Schematic diagrams of mutated *Cryptosporidium parvum* GST with the InterPro and Pfam confirmed GST domains and the N- and C-terminal extensions. A: A full length GST without the N-terminal extension. B: A full length GST without the C-terminal extension. C: The common GST without any of the extensions.

4.2. References

ALLOCATI, N., CASALONE, E., MASULLI, M., CECCARELLI, I., CARLETTI, E., PARKER, M. W. & DI ILIO, C. 1999. Functional analysis of the evolutionarily conserved proline 53 residue in *Proteus mirabilis* glutathione transferase B1-1. *FEBS Letters,* 445**,** 347-350.

CERTAD, G., VISCOGLIOSI, E., CHABÉ, M. & CACCIÒ, S. M. 2017. Pathogenic mechanisms of *Cryptosporidium and Giardia*. *Trends in Parasitology,* 33**,** 561-576.

FEASEY, N., WANSBROUGH-JONES, M., MABEY, D. C. & SOLOMON, A. W. 2009. Neglected tropical diseases. *British Medical Bulletin,* 93**,** 179-200.

FRITZ-WOLF, K., BECKER, A., RAHLFS, S., HARWALDT, P., SCHIRMER, R. H., KABSCH, W. & BECKER, K. 2003. X-ray structure of glutathione S-transferase from the malarial parasite P*lasmodium falciparum*. *Proceedings of the National Academy of Sciences,* 100**,** 13821-13826.

FROVA, C. 2006. Glutathione transferases in the genomics era: new insights and perspectives. *Biomolecular Engineering,* 23**,** 149-169.

JEMTH, P. & MANNERVIK, B. 1997. Kinetic characterization of recombinant human glutathione transferase T1-1, a polymorphic detoxication enzyme. *Archives of Biochemistry and Biophysics,* 348**,** 247-254.

MAUZY, M. J., ENOMOTO, S., LANCTO, C. A., ABRAHAMSEN, M. S. & RUTHERFORD, M. S. 2012. The *Cryptosporidium parvum* transcriptome during in vitro development. *PloS One,* 7, e31715.

QING, G., MA, L.-C., KHORCHID, A., SWAPNA, G., MAL, T. K., TAKAYAMA, M. M., XIA, B., PHADTARE, S., KE, H. & ACTON, T. 2004. Cold-shock induced high-yield protein production in *Escherichia coli*. *Nature Biotechnology,* 22**,** 877-882.

SAINI, P., WANI, S. I., KUMAR, R., CHHABRA, R., CHIMNI, S. S. & SAREEN, D. 2014. Trigger factor assisted folding of the recombinant epoxide hydrolases identified from *C. pelagibacter* and *S. nassauensis*. *Protein Expression and Purification,* 104**,** 71-84.

SHEEHAN, D., MEADE, G. & FOLEY, V. M. 2001. Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochemical Journal,* 360**,** 1-16.

SHIRLEY, D.-A. T., MOONAH, S. N. & KOTLOFF, K. L. 2012. Burden of disease from cryptosporidiosis. *Current Opinion in Infectious Diseases,* 25**,** 555.

YOUNG, C. L., BRITTON, Z. T. & ROBINSON, A. S. 2012. Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. *Biotechnology Journal,* 7**,** 620-634.

APPENDIX A

CHAPTER 2 SUPPLEMENTARY DATASET

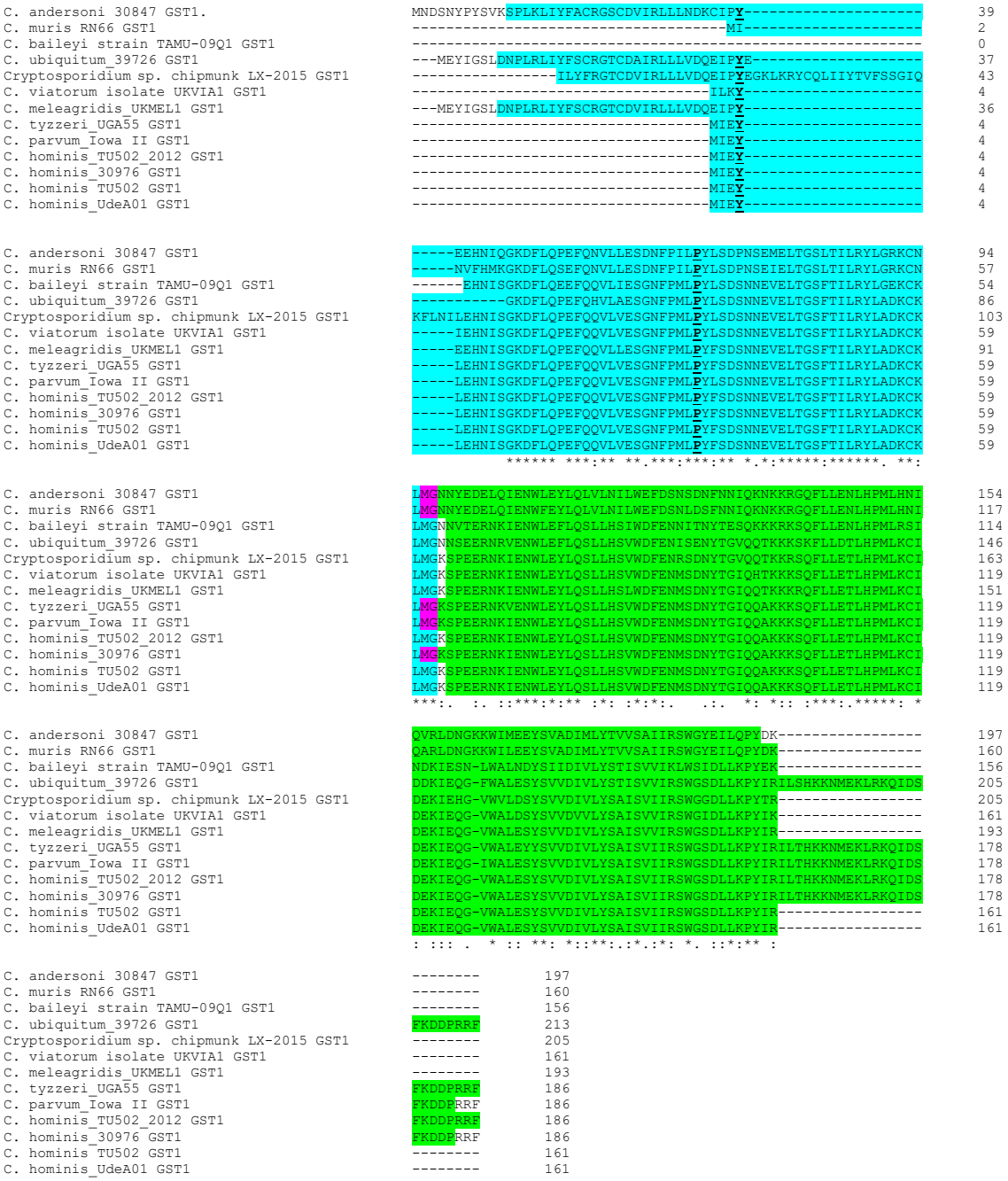
Table S1: Information on different glutathione transferase classes found in organisms.

GST Class	Cellular localization	General Information	Reference
Alpha	Cytosol	Found in a broad range of species. Involved in the biosynthesis of sex steroids and keto-steroid isomerase activity.	Deponte and Becker, 2005
Beta	Cytosol	Typically found in bacterial species. Known for conjugating antibiotics, assisting in antibiotic resistance to other organisms.	Shehu *et al*., 2019
CLIC	Cytosol	Found in a broad range of species. Enter intracellular membranes and form membrane channels.	Board and Menon, 2013
Delta and Epsilon	Cytosol	Typically found in insects. Thought to contribute to detoxication or antioxidative stress during development. Delta GSTs are also involved in oogenesis.	Scian *et al*., 2015, Udomsinprasert *et al*., 2005
Kappa	Mitochondrial	To date, found in primates and mice. Oligomerization of adiponectin.	Robinson *et al*., 2004
Lambda	Cytosol	Typically found in plants. Function is not yet known as they have no detectable GSH-conjugating activity.	Chronopoulou *et al*., 2017
MAPEG	Microsomal	Found in a broad range of species. Involved in production of leukotrienes and prostaglandin E and are mediators of inflammation.	Akil *et al*., 2012
Mu	Cytosol	Found in a broad range of species. Forms inhibitory complexes with ASK1, another member of the MAP kinase pathway.	Torres-Rivera and Landa, 2008
Omega	Cytosol	Found in a broad range of species. Catalyzes reduction and thioltransferase reactions.	Wu and Dong, 2012
Phi	Cytosol	Typically found in plants. Inhibits oxidative damage through the removal of endogenous cytotoxic hydroperoxides.	Munyampundu *et al*., 2016
Pi	Cytosol	Found in a broad range of species. Regulates JNK and TRAF signaling and catalyzes the S-glutathionylation reactions.	Prade *et al*., 1997

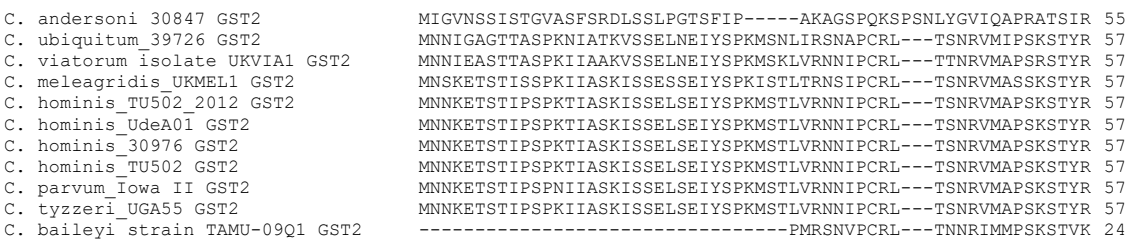
Sigma	Cytosol	Found in a broad range of species. Involved in prostaglandin synthesis by isomerization of PGH_2 – PGD_2.	Board and Menon, 2013
Tau	Cytosol	Typically found in plants. Involved in reactive oxygen species scavenging and improves plant chilling tolerance	Yang *et al*., 2016
Theta	Cytosol	Found in a broad range of species. Has dichloromethane dehalogenase activity for the degradation of dichloromethane to obtain energy	Shehu *et al*., 2019
Xi	Cytosol	Typically found in bacteria, fungi, and archaea. Aids in extreme haloalkaphilic conditions.	Di Matteo *et al*., 2019
Zeta	Cytosol	Found in a broad range of species. Involved in isomerization of maleyacetoacetate to fumaracetoacetate in tyrosine degradation pathway and biotransformation of dichloroacetic acid to glyoxylate	Board *et al*., 1997

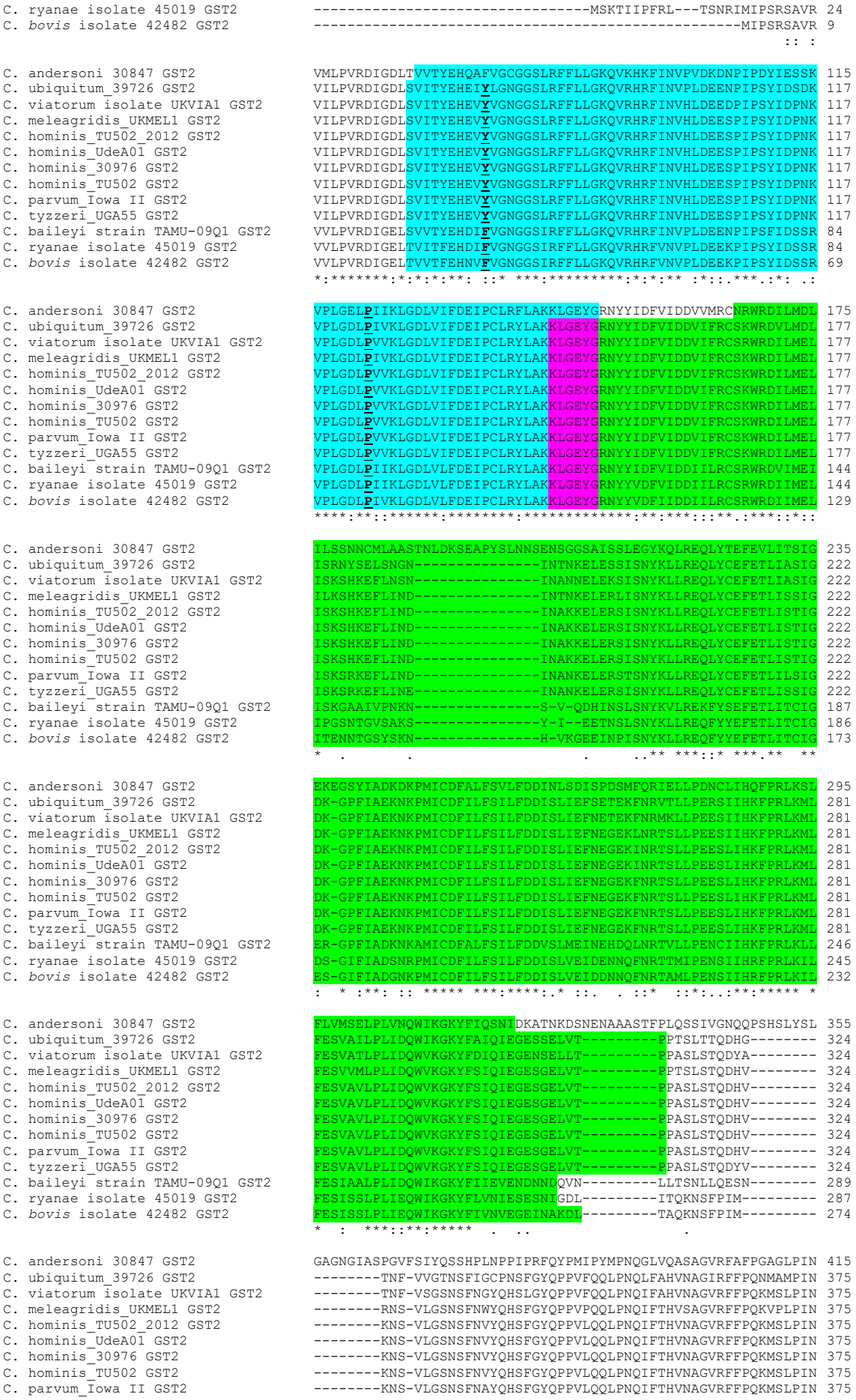
Symbol: *, Based on in silico prediction. Abbreviations: GSH, Glutathione; ASK1, Apoptosis signal-regulated kinase 1; MAP, Mitogen activated protein; JNK, c-Jun N-Terminal Kinase; TRAF, Tumor necrosis factor receptor (TNF)-associated factor; PGH2, Prostaglandin H2; PGD2, Prostaglandin D2

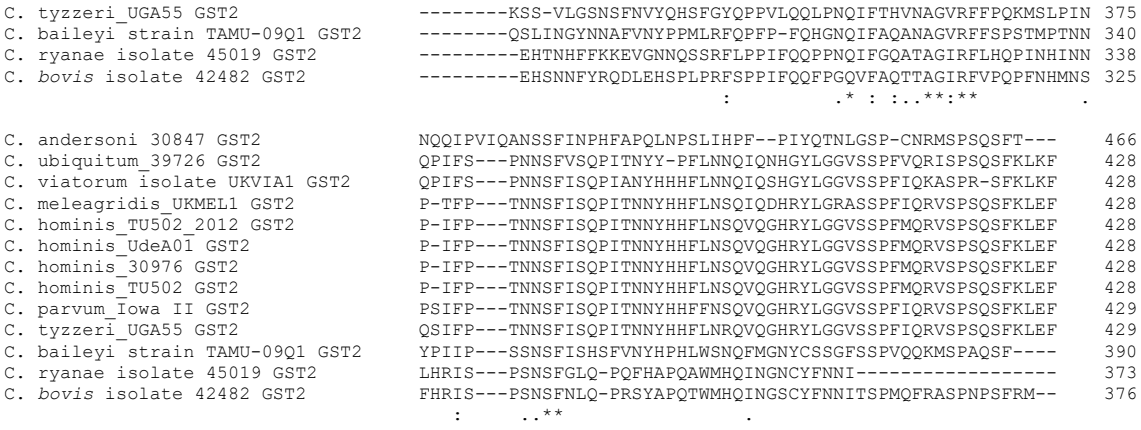
Group 1 (Vega (ϑ))



Group 2 (Gamma (γ))




C. ryanae isolate 45019 GST2 ----------MSKTIIPFRL---TSNRIMIPSRSAVR 24
C. bovis isolate 42482 GST2 ----------MIPSRSAVR 9
C. andersoni 30847 GST2 VMLPVRDIGDLTVVTYEHQAFVGCGGSLRFFLLGKQVKHKFINVPVDKDNPIPDYIESSK 115
C. ubiquitum_39726 GST2 VILPVRDIGDLSVITYEHEIYLGNGGSLRFFLLGKQVRHRFINVPLDEENPIPSYIDSDK 117
C. viatorum isolate UKVIA1 GST2 VILPVRDIGDLSVITYEHEVYVGNGGSLRFFLLGKQVRHRFINVHLDEEDPIPSYIDPNK 117
C. meleagridis_UKMEL1 GST2 VILPVRDIGDLSVITYEHEVYVGNGGSLRFFLLGKQVRHRFINVHLDEESPIPSYIDPNK 117
C. hominis_TU502_2012 GST2 VILPVRDIGDLSVITYEHEVYVGNGGSLRFFLLGKQVRHRFINVHLDEESPIPSYIDPNK 117
C. hominis_UdeA01 GST2 VILPVRDIGDLSVITYEHEVYVGNGGSLRFFLLGKQVRHRFINVHLDEESPIPSYIDPNK 117
C. hominis_30976 GST2 VILPVRDIGDLSVITYEHEVYVGNGGSLRFFLLGKQVRHRFINVHLDEESPIPSYIDPNK 117
C. hominis_TU502 GST2 VILPVRDIGDLSVITYEHEVYVGNGGSLRFFLLGKQVRHRFINVHLDEESPIPSYIDPNK 117
C. parvum_Iowa II GST2 VILPVRDIGDLSVITYEHEVYVGNGGSLRFFLLGKQVRHRFINVHLDEESPIPSYIDPNK 117
C. tyzzeri_UGA55 GST2 VILPVRDIGDLSVITYEHEVYVGNGGSLRFFLLGKQVRHRFINVHLDEESPIPSYIDPNK 117
C. baileyi strain TAMU-09Q1 GST2 VVLPVRDIGELSVVTYEHDIFVGNGGSIRFFLLGKQVRHRFINVPLDEENPIPSFIDSSR 84
C. ryanae isolate 45019 GST2 VVLPVRDIGELTVITFEHDIFVGNGGSIRFFLLGKQVRHRFVNVPLDEEKPIPSYIDSSR 84
C. bovis isolate 42482 GST2 VVLPVRDIGELTVVTFEHNVFVGNGGSIRFFLLGKQVRHRFVNVPLDEEKPIPSYIDSSR 69
C. andersoni 30847 GST2 VPLGELPIIKLGDLVIFDEIPCLRFLAKKLGEYGRNYYIDFVIDDVVMRCNRWRDILMDL 175
C. ubiquitum_39726 GST2 VPLGDLPIVKLGDLVIFDEIPCLRYLAKKLGEYGRNYYIDFVIDDVIFRCSKWRDVLMDL 177
C. viatorum isolate UKVIA1 GST2 VPLGDLPVVKLGDLVIFDEIPCLRYLAKKLGEYGRNYYIDFVIDDVIFRCSKWRDILMEL 177
C. meleagridis_UKMEL1 GST2 VPLGDLPIVKLGDLVIFDEIPCLRYLAKKLGEYGRNYYIDFVIDDVIFRCSKWRDILMEL 177
C. hominis_TU502_2012 GST2 VPLGDLPVVKLGDLVIFDEIPCLRYLAKKLGEYGRNYYIDFVIDDVIFRCSKWRDILMEL 177
C. hominis_UdeA01 GST2 VPLGDLPVVKLGDLVIFDEIPCLRYLAKKLGEYGRNYYIDFVIDDVIFRCSKWRDILMEL 177
C. hominis_30976 GST2 VPLGDLPVVKLGDLVIFDEIPCLRYLAKKLGEYGRNYYIDFVIDDVIFRCSKWRDILMEL 177
C. hominis_TU502 GST2 VPLGDLPVVKLGDLVIFDEIPCLRYLAKKLGEYGRNYYIDFVIDDVIFRCSKWRDILMEL 177
C. parvum_Iowa II GST2 VPLGDLPVVKLGDLVIFDEIPCLRYLAKKLGEYGRNYYIDFVIDDVIFRCSKWRDILMEL 177
C. tyzzeri_UGA55 GST2 VPLGDLPVVKLGDLVIFDEIPCLRYLAKKLGEYGRNYYIDFVIDDVIFRCSKWRDILMEL 177
C. baileyi strain TAMU-09Q1 GST2 VPLGDLPIIKLGDLVLFDEIPCLRYLAKKLGEYGRNYYIDFVIDDIILRCSRWRDVIMEI 144
C. ryanae isolate 45019 GST2 VPLGDLPIIKLGDLVLFDEIPCLRYLAKKLGEYGRNYYVDFVIDDIILRCSRWRDIIMEL 144
C. bovis isolate 42482 GST2 VPLGDLPIVKLGDLVLFDEIPCLRYLAKKLGEYGRNYYVDFIIDDIILRCSRWRDIIMEL 129
C. andersoni 30847 GST2 ILSSNNCMLAASTNLDKSEAPYSLNNSENSGGSAISSLEGYKQLREQLYTEFEVLITSIG 235
C. ubiquitum_39726 GST2 ISRNYSELSNGN----------INTNKELESSISNYKLLREQLYCEFETLIASIG 222
C. viatorum isolate UKVIA1 GST2 ISKSHKEFLNSN----------INANNELEKSISNYKLLREQLYCEFETLIASIG 222
C. meleagridis_UKMEL1 GST2 ILKSHKEFLIND----------INTNKELERLISNYKLLREQLYCEFETLISSIG 222
C. hominis_TU502_2012 GST2 ISKSHKEFLIND----------INAKKELERSISNYKLLREQLYCEFETLISTIG 222
C. hominis_UdeA01 GST2 ISKSHKEFLIND----------INAKKELERSISNYKLLREQLYCEFETLISTIG 222
C. hominis_30976 GST2 ISKSHKEFLIND----------INAKKELERSISNYKLLREQLYCEFETLISTIG 222
C. hominis_TU502 GST2 ISKSHKEFLIND----------INAKKELERSISNYKLLREQLYCEFETLISTIG 222
C. parvum_Iowa II GST2 ISKSRKEFLIND----------INANKELERSTSNYKLLREQLYCEFETLILSIG 222
C. tyzzeri_UGA55 GST2 ISKSRKEFLINE----------INANKELERSISNYKLLREQLYCEFETLISSIG 222
C. baileyi strain TAMU-09Q1 GST2 ISKGAAIVPNKN----------S-V-QDHINSLSNYKVLREKFYSEFETLITCIG 187
C. ryanae isolate 45019 GST2 IPGSNTGVSAKS----------Y-I--EETNSLSNYKLLREQFYYEFETLITCIG 186
C. bovis isolate 42482 GST2 ITENNTGSYSKN----------H-VKGEEINPISNYKLLREQFYYEFETLITCIG 173
C. andersoni 30847 GST2 EKEGSYIADKDKPMICDFALFSVLFDDINLSDISPDSMFQRIELLPDNCLIHQFPRLKSL 295
C. ubiquitum_39726 GST2 DK-GPFIAEKNKPMICDFILFSILFDDISLIEFSETEKFNRVTLLPERSIIHKFPRLKML 281
C. viatorum isolate UKVIA1 GST2 DK-GPFIAEKNKPMICDFILFSILFDDISLIEFNETEKFNRMKLLPEESIIHKFPRLKML 281
C. meleagridis_UKMEL1 GST2 DK-GPFIAEKNKPMICDFILFSILFDDISLIEFNEGEKLNRTSLLPEESIIHKFPRLKML 281
C. hominis_TU502_2012 GST2 DK-GPFIAEKNKPMICDFILFSILFDDISLIEFNEGEKINRTSLLPEESLIHKFPRLKML 281
C. hominis_UdeA01 GST2 DK-GPFIAEKNKPMICDFILFSILFDDISLIEFNEGEKINRTSLLPEESLIHKFPRLKML 281
C. hominis_30976 GST2 DK-GPFIAEKNKPMICDFILFSILFDDISLIEFNEGEKFNRTSLLPEESLIHKFPRLKML 281
C. hominis_TU502 GST2 DK-GPFIAEKNKPMICDFILFSILFDDISLIEFNEGEKFNRTSLLPEESLIHKFPRLKML 281
C. parvum_Iowa II GST2 DK-GPFIAEKNKPMICDFILFSILFDDISLIEFNEGEKFNRTSLLPEESLIHKFPRLKML 281
C. tyzzeri_UGA55 GST2 DK-GPFIAEKNKPMICDFILFSILFDDISLIEFNEGEKFNRTSLLPEESLIHKFPRLKML 281
C. baileyi strain TAMU-09Q1 GST2 ER-GPFIADKNKAMICDFALFSILFDDVSLMEINEHDQLNRTVLLPENCIIHKFPRLKLL 246
C. ryanae isolate 45019 GST2 DS-GIFIADSNRPMICDFILFSILFDDISLVEIDENNQFNRTTMIPENSIIHRFPRLKIL 245
C. bovis isolate 42482 GST2 ES-GIFIADGNKPMICDFILFSILFDDISLVEIDDNNQFNRTAMLPENSIIHRFPRLKIL 232
C. andersoni 30847 GST2 FLVMSELPLVNQWIKGKYFIQSNIDKATNKDSNENAAASTFPLQSSIVGNQQPSHSLYSL 355
C. ubiquitum_39726 GST2 FESVAILPLIDQWIKGKYFAIQIEGESSELVT---------PPTSLTTQDHG-------- 324
C. viatorum isolate UKVIA1 GST2 FESVATLPLIDQWVKGKYFDIQIEGENSELLT---------PPASLSTQDYA-------- 324
C. meleagridis_UKMEL1 GST2 FESVVMLPLIDQWVKGKYFSIQIEGESGELVT---------PPTSLSTQDHV-------- 324
C. hominis_TU502_2012 GST2 FESVAVLPLIDQWVKGKYFSIQIEGESGELVT---------PPASLSTQDHV-------- 324
C. hominis_UdeA01 GST2 FESVAVLPLIDQWVKGKYFSIQIEGESGELVT---------PPASLSTQDHV-------- 324
C. hominis_30976 GST2 FESVAVLPLIDQWVKGKYFSIQIEGESGELVT---------PPASLSTQDHV-------- 324
C. hominis_TU502 GST2 FESVAVLPLIDQWVKGKYFSIQIEGESGELVT---------PPASLSTQDHV-------- 324
C. parvum_Iowa II GST2 FESVAVLPLIDQWVKGKYFSIQIEGESGELVT---------PPASLSTQDHV-------- 324
C. tyzzeri_UGA55 GST2 FESVAVLPLIDQWVKGKYFSIQIEGESGELVT---------PPASLSTQDYV-------- 324
C. baileyi strain TAMU-09Q1 GST2 FESIAALPLIDQWIKGKYFIIEVENDNNDQVN---------LLTSNLLQESN-------- 289
C. ryanae isolate 45019 GST2 FESISSLPLIEQWIKGKYFLVNIESESNIGDL---------ITQKNSFPIM--------- 287
C. bovis isolate 42482 GST2 FESISSLPLIEQWIKGKYFIVNVEGEINAKDL---------TAQKNSFPIM--------- 274
C. andersoni 30847 GST2 GAGNGIASPGVFSIYQSSHPLNPPIPRFQYPMIPYMPNQGLVQASAGVRFAFPGAGLPIN 415
C. ubiquitum_39726 GST2 --------TNF-VVGTNSFIGCPNSFGYQPPVFQQLPNQLFAHVNAGIRFFPQNMAMPIN 375
C. viatorum isolate UKVIA1 GST2 --------TNF-VSGSNSFNGYQHSLGYQPPVFQQLPNQIFAHVNAGVRFFPQKMSLPIN 375
C. meleagridis_UKMEL1 GST2 --------RNS-VLGSNSFNWYQHSFGYQPPVPQQLPNQIFTHVSAGVRFFPQKVPLPIN 375
C. hominis_TU502_2012 GST2 --------KNS-VLGSNSFNVYQHSFGYQPPVLQQLPNQIFTHVNAGVRFFPQKMSLPIN 375
C. hominis_UdeA01 GST2 --------KNS-VLGSNSFNVYQHSFGYQPPVLQQLPNQIFTHVNAGVRFFPQKMSLPIN 375
C. hominis_30976 GST2 --------KNS-VLGSNSFNVYQHSFGYQPPVLQQLPNQIFTHVNAGVRFFPQKMSLPIN 375
C. hominis_TU502 GST2 --------KNS-VLGSNSFNVYQHSFGYQPPVLQQLPNQIFTHVNAGVRFFPQKMSLPIN 375
C. parvum_Iowa II GST2 --------KNS-VLGSNSFNAYQHSFGYQPPVLQQLPNQIFTHVNAGVRFFPQKMSLPIN 375



Group 3 (Psi class)

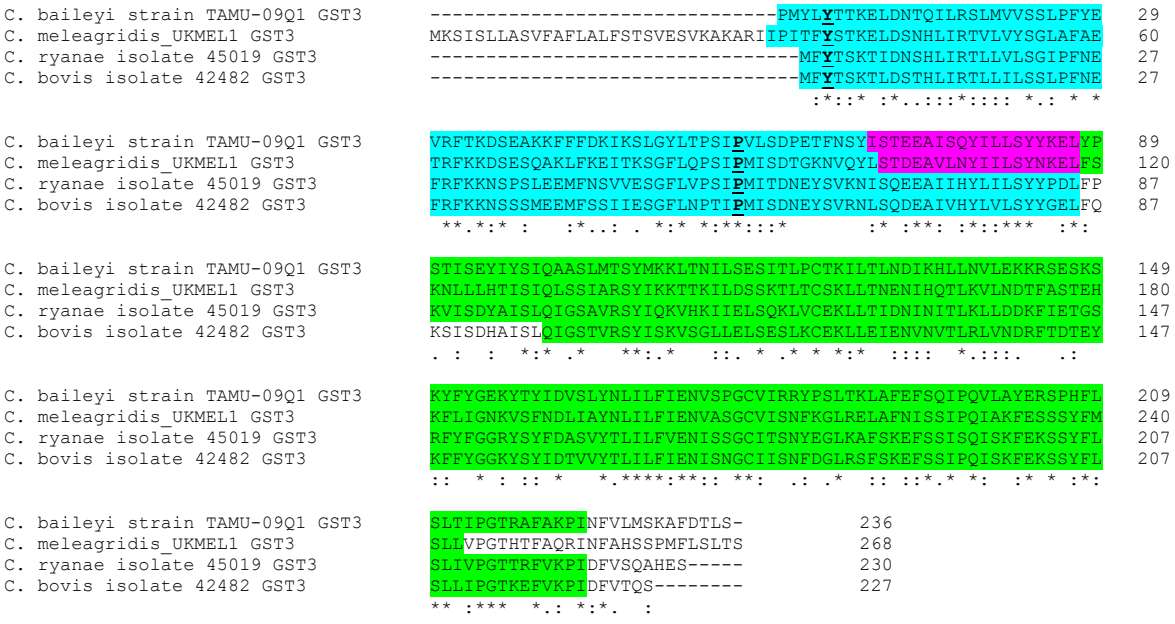


Figure S1: Multiple amino acid alignment of glutathione transferases (GSTs) from *Cryptosporidium* species. The conserved amino acids, active site tyrosine and the *cis*-proline are bold and underlined. The N-terminal region is highlighted in cyan, the C-terminal region is highlighted in green and the purple text indicates the overlapping region where the N- and C-terminal regions are shared.

Table S2: Prediction of transmembrane helices in glutathione transferase (GST) proteins of *Cryptosporidium* species and GSTs belonging to different classes. Prediction of transmembrane helices in GSTs were carried out using TMHMM - 2.0

C. andersoni 30847 GST1(cand_012830)	len=197	ExpAA=2.35	First60=0.01	PredHel=0	Topology=o
C. hominis TU502_2012 GST1(ChTU502y2012_407g2365/Q18145.1)	len=186	ExpAA=1.21	First60=0.00	PredHel=0	Topology=o
C. hominis 30976 GST1(GY17_00002363)	len=186	ExpAA=1.21	First60=0.00	PredHel=0	Topology=o
C. hominis TU502 GST1(XP_667744.1)	len=161	ExpAA=2.65	First60=0.00	PredHel=0	Topology=o
C. hominis UdeA01 GST1(CUV07467.1)	len=161	ExpAA=2.65	First60=0.00	PredHel=0	Topology=o
C. meleagridis UKMEL1 GST1(CmeUKMEL1_03350)	len=193	ExpAA=1.11	First60=0.01	PredHel=0	Topology=o
C. parvum Iowa II GST1(cgd7_4780)	len=186	ExpAA=1.48	First60=0.00	PredHel=0	Topology=o
C. tyzzeri UGA55 GST1(CTYZ_00001095)	len=186	ExpAA=3.20	First60=0.00	PredHel=0	Topology=o
C. ubiquitum 39726 GST1(cubi_03151)	len=213	ExpAA=0.57	First60=0.01	PredHel=0	Topology=o
C. muris RN66 GST1(XP_002141168.1)	len=160	ExpAA=1.78	First60=0.00	PredHel=0	Topology=o
Cryptosporidium sp. *chipmunk* LX-2015 *GST1 (JXRN01000042.1)*	len=205	ExpAA=3.35	First60=0.15	PredHel=0	Topology=o
C. viatorum isolate UKVIA1 GST1(QZWW01000010.1)	len=161	ExpAA=1.92	First60=0.00	PredHel=0	Topology=o
C. baileyi strain TAMU-09Q1 GST1(JIBL01000090.1)	len=156	ExpAA=2.50	First60=0.00	PredHel=0	Topology=o
C. baileyi strain TAMU-09Q1 GST2(JIBL01000106.1)	len=390	ExpAA=0.01	First60=0.01	PredHel=0	Topology=o
C. viatorum isolate UKVIA1 GST2(QZWW01000018.1)	len=428	ExpAA=0.02	First60=0.00	PredHel=0	Topology=o
C. ryanae isolate 45019 GST2(VHLK01000046.1)	len=373	ExpAA=0.03	First60=0.01	PredHel=0	Topology=o
C. andersoni 30847 GST2(cand_023790)	len=466	ExpAA=0.03	First60=0.00	PredHel=0	Topology=o
C. hominis TU502_2012 GST2(ChTU502y2012_421g0615)	len=428	ExpAA=0.01	First60=0.00	PredHel=0	Topology=o
C. hominis 30976 GST2(GY17_00000733)	len=428	ExpAA=0.01	First60=0.00	PredHel=0	Topology=o

C. hominis TU502 GST2(Chro.80347)	len=428	ExpAA=0.01	First60=0.00	PredHel=0	Topology=o
C. hominis UdeA01 GST2(CHUDEA8_2970)	len=428	ExpAA=0.01	First60=0.00	PredHel=0	Topology=o
C. meleagridis UKMEL1 GST2(CmeUKMEL1_14570)	len=428	ExpAA=0.01	First60=0.00	PredHel=0	Topology=o
C. parvum Iowa II GST2(cgd8_2970).	len=429	ExpAA=0.01	First60=0.00	PredHel=0	Topology=o
C. tyzzeri UGA55 GST2(CTYZ_0000322)	len=429	ExpAA=0.01	First60=0.00	PredHel=0	Topology=o
C. ubiquitum 39726 GST2(cubi_03523)	len=428	ExpAA=0.01	First60=0.00	PredHel=0	Topology=o
C. bovis isolate 42482 GST2(VHIT01000012.1)	len=376	ExpAA=0.33	First60=0.27	PredHel=0	Topology=o
C. meleagridis UKMEL1 GST3(CmeUKMEL1_05845)	len=268	ExpAA=11.30	First60=9.03	PredHel=0	Topology=o
C. baileyi strain TAMU-09Q1 GST3(JIBL01000138.1)	len=236	ExpAA=0.49	First60=0.19	PredHel=0	Topology=o
C. ryanae isolate 45019 GST3(VHLK01000056.1)	len=230	ExpAA=0.76	First60=0.19	PredHel=0	Topology=o
C. bovis isolate 42482 GST3(PRJNA545579)	len=227	ExpAA=2.32	First60=0.19	PredHel=0	Topology=o

Abbreviations: len, the length of the protein sequence; ExpAA, the expected number of amino acids in transmembrane helices; First60, The expected number of amino acids in transmembrane helices in the first 60 amino acids of the protein; PredHel, the number of predicted transmembrane helices by N-best; Topology, the topology predicted by N-best.

Table S3: Comparative analysis of transmembrane helices in eukaryotic glutathione transferase. Prediction of transmembrane helices in GSTs were carried out using TMHMM - 2.0. Abbreviation: PredHel, the number of predicted transmembrane helices.

GST Class	Location (experimental)	Number of GSTs analyzed	TMHMM 2.0 results
Alpha	Cytosol	37	PredHel=0
Beta	Cytosol	6	PredHel=0
CLIC	Cytosol	30	PredHel=0
Delta	Cytosol	29	PredHel=0
Epsilon	Cytosol	22	PredHel=0
Kappa	Mitochondrial	5	PredHel=0
Lambda	Cytosol	18	PredHel=0
MAPEG	Microsomal	20	PredHel=1-4
Mu	Cytosol	27	PredHel=0
Omega	Cytosol	32	PredHel=0
Phi	Cytosol	23	PredHel=0
Pi*	Cytosol	33	PredHel=0
Sigma	Cytosol	27	PredHel=0
Tau	Cytosol	21	PredHel=0
Theta	Cytosol	28	PredHel=0
Xi	Cytosol	4	PredHel=0
Zeta	Cytosol	33	PredHel=0

*, only one GST from *Chlorocebus sabaeus* (protein ID: A0A0D9R8K4) showed one transmembrane helix.

Table S4: Analysis of *Cryptosporidium* species glutathione transferases (GSTs) cellular localization using Bologna Unified Subcellular Component Annotator (BUSCA) web-server. All *Cryptosporidium* species GSTs predicated to be located in cytoplasm. The two GSTs transmembrane membrane helix scores are below the values (0.93 and 1 for transmembrane) that can be considered as membrane bound proteins GSTs.

GST	GO-id	GO-term	Score	Alternative localization	Features
C. muris RN66 GST1(XP_002141168.1).	GO:0005737	C:cytoplasm	0.7	GO:0005634 - C:nucleus (score=0.3)	
C. viatorum isolate UKVIA1 GST1(QZWW01000010.1)	GO:0005737	C:cytoplasm	1	-	
C. baileyi strain TAMU-09Q1 GST1(JIBL01000090.1)	GO:0005737	C:cytoplasm	0.7	GO:0005634 - C:nucleus (score=0.3)	
C. baileyi strain TAMU-09Q1 GST2(JIBL01000106.1)	GO:0005737	C:cytoplasm	0.7	GO:0005634 - C:nucleus (score=0.3)	
C. viatorum isolate UKVIA1 GST2(QZWW01000018.1)	GO:0005737	C:cytoplasm	0.7	GO:0005634 - C:nucleus (score=0.3)	
C. ryanae isolate 45019 GST2(VHLK01000046.1)	GO:0005737	C:cytoplasm	0.7	GO:0005634 - C:nucleus (score=0.3)	
C. andersoni 30847 GST2(cand_023790)	GO:0005739	C:mitochondrion	0.64	-	Mitochondrial Transit Peptide
*C. hominis*_TU502_2012 GST2(ChTU502y2012_421g0615)	GO:0005737	C:cytoplasm	0.7	GO:0005634 - C:nucleus (score=0.3)	
*C. hominis*_30976 GST2(GY17_00000733)	GO:0005737	C:cytoplasm	0.7	GO:0005634 - C:nucleus (score=0.3)	
*C. hominis*_TU502 GST2(Chro.80347)	GO:0005737	C:cytoplasm	0.7	GO:0005634 - C:nucleus (score=0.3)	
C. andersoni 30847 GST1(cand_012830)	GO:0012505	C:endomembrane system	0.73	GO:0031090 - C:organelle membrane (score=0.67)	Transmembrane Alpha Helix
*C. hominis*_UdeA01 GST2(CHUDEA8_2970)	GO:0005737	C:cytoplasm	0.7	GO:0005634 - C:nucleus (score=0.3)	
*C. meleagridis*_UKMEL1 GST2(CmeUKMEL1_14570)	GO:0005737	C:cytoplasm	0.7	GO:0005634 - C:nucleus (score=0.3)	
*C. parvum*_Iowa II GST2(cgd8_2970)	GO:0005737	C:cytoplasm	0.7	GO:0005634 - C:nucleus (score=0.3)	

*C. tyzzeri*_UGA55 GST2(CTYZ_0000322)	GO:0005737	C:cytoplasm	0.7	GO:0005634 - C:nucleus (score=0.3)	
*C. ubiquitum*_39726 GST2(cubi_03523)	GO:0005737	C:cytoplasm	0.7	GO:0005634 - C:nucleus (score=0.3)	
C. bovis isolate 42482 GST2(VHIT01000012.1)	GO:0005737	C:cytoplasm	0.7	GO:0005634 - C:nucleus (score=0.3)	
*C. meleagridis*_UKMEL1 GST3(CmeUKMEL1_05845)	GO:0012505	C:endomembrane system	0.84	GO:0005886 - C:plasma membrane (score=0.47)	Transmembrane Alpha Helix
C. baileyi strain TAMU-09Q1 GST3(JIBL01000138.1).	GO:0005737	C:cytoplasm	0.7	GO:0005634 - C:nucleus (score=0.3)	
C. ryanae isolate 45019 GST3(VHLK01000056.1)	GO:0005615	C:extracellular space	1	-	
C. bovis isolate 42482 GST3(PRJNA545579)	GO:0005615	C:extracellular space	1	-	
*C. hominis*_TU502_2012 GST1(ChTU502y2012_407g2365 /Q18145.1)	GO:0005737	C:cytoplasm	0.89	GO:0005634 - C:nucleus (score=0.11)	
*C. hominis*_30976 GST1(GY17_00002363)	GO:0005737	C:cytoplasm	0.89	GO:0005634 - C:nucleus (score=0.11)	
Cryptosporidium sp. chipmunk *LX-2015 GST1 (JXRN01000042.1)*	GO:0005615	C:extracellular space	1	-	-
C. hominis TU502 GST1(XP_667744.1)	GO:0005737	C:cytoplasm	1	-	
*C. hominis*_UdeA01 GST1(CUV07467.1)	GO:0005737	C:cytoplasm	1	-	
*C. meleagridis*_UKMEL1 GST1(CmeUKMEL1_03350)	GO:0005737	C:cytoplasm	1	-	
*C. parvum*_Iowa II GST1(cgd7_4780)	GO:0005737	C:cytoplasm	0.84	GO:0005634 - C:nucleus (score=0.16)	
*C. tyzzeri*_UGA55 GST1(CTYZ_00001095)	GO:0005737	C:cytoplasm	0.97	-	
*C. ubiquitum*_39726 GST1(cubi_03151)	GO:0005737	C:cytoplasm	1	-	

Table S5: Information on template hits obtained from different databases for homology modeling of *Cryptosporidium parvum* GSTs 1 and 2 and *Cryptosporidium meleagridis* UKMEL1 GST3 (CmGST3). Yellow-shaded templates were found to be the best combination to generate good quality 3D models.

***C. parvum* GST1 (C*p*GST1) hits**

	PDB-Chain	% Identity	Query Cover	Resolution	R-Value Free	R-Value Work
NCBI	4L5O-A	51	23 %	2.09	0.248	0.202
	3ISO-A	51	23 %	1.90	0.226	0.179
	2WRT-A	45	24 %	2.40	0.281	0.205
	1FHE-A	45	24 %	3.00	0.354	0.237
	5AN1-A	37	40 %	2.00	0.262	0.214
	1YQ1-A	28	59 %	3.00	0.297	0.224
PHYRE	1B8X-A	27	98 %	2.70	0.310	0.209
	5AN1-A	23	98 %	2.00	0.262	0.214
	1BG5-A	27	98 %	2.60	0.359	0.193
	1GTU-B	22	98 %	2.68	0.245	0.211
	1UA5-A	25	98 %	2.50	0.221	0.169
	1C72-A	21	98 %	2.80	0.190	0.190
	1VF1-A	20	98 %	1.77	0.217	0.192
	4Q5Q-A	20	98 %	1.93	0.227	0.180
	3C8E-B	19	96 %	2.71	0.181	0.182
	4MZW-A	19	97 %	1.92	0.194	0.152
	1GUM-A	18	98 %	3.00	0.270	0.253
	5H5L-A	18	98 %	2.00	0.238	0.173
i-TASSER	3VPQ-A	26	94 %	1.70	0.227	0.188
	4Q5F-A	22	94 %	2.45	0.239	0.188
	1M0U-A	25	93 %	1.75	0.232	0.213
	4ZB9	22	98 %	2.40	0.273	0.216
	4ECI	23	95 %	1.80	0.209	0.175
	5F05-A	19	96 %	1.70	0.181	0.144
	4Q5F	21	93 %	2.45	0.239	0.188
	5HFK-A	26	97 %	1.55	0.212	0.185
	1DUG-A	30	93 %	1.80	0.226	0.185
	1B8X-A	29	95 %	2.70	0.310	0.209

***C. parvum* GST2 (C*p*GST2) hits**

	PDB-Chain	% Identity	Query cover	Resolution	R-Value Free	R-Value Work
NCBI	5H5L-A	26	73 %	2.00	0.238	0.173
	2WS2-A	22	73 %	2.01	0.290	0.236
	2AAW-A	24	88 %	2.40	0.237	0.194
	2ON7-A	27	37 %	2.40	0.280	0.190
	1OKT-A	24	89 %	1.90	0.259	0.222
	3W8S-A	21	92 %	1.90	0.259	0.222
	1TW9-A	20	92 %	1.71	0.232	0.180
	4ZXG-A	24	86 %	1.70	0.229	0.183
	2ON5-A	21	74 %	1.90	0.227	0.178
	3FR3-A	22	86 %	1.90	0.252	0.211
	3FR6-A	24	86%	2.60	0.302	0.216
	4KDU-A	27	29%	1.60	0.200	0.168
PHYRE	5H5L-A	21	93 %	2.00	0.238	0.173
	1K3Y-B	21	94 %	1.30	0.207	0.137
	1M0U-B	19	93 %	1.75	0.232	0.213
	3VPT-A	19	93 %	1.90	0.253	0.204
	1GTU-B	19	94 %	2.88	0.245	0.211
	2WB9-A	18	90 %	1.59	0.222	0.175
	3ISO-B	18	96 %	1.90	0.226	0.179
	1B8X-A	17	95 %	2.70	0.310	0.209
	1BG5-A	17	95 %	2.60	0.359	0.193
	4W66-A	17	95 %	2.39	0.238	0.186
	5AN1-A	17	94 %	2.00	0.262	0.214
i-TASSER	3ISO-A	20	86 %	1.90	0.226	0.179
	2ON5-A	19	83 %	1.90	0.227	0.178
	1TU7-A	21	83 %	1.5	0.181	0.150
	19GS-A	22	84 %	1.9	0.249	0.212
	2ON7	19	83 %	2.4	0.284	0.190
	4Q5N-A	19	88 %	2.5	0.283	0.224
	19GS	22	80 %	1.9	0.249	0.212
	3GTU-B	19	86 %	2.80	0.270	0.225
	5H5L-A	23	79 %	2.00	0.238	0.173

	4W66-A	18	84 %	2.39	0.238	0.186

***C. meleagridis* UKMEL1 GST3 (C*m*GST3) hits**

	PDB-Chain	% Identity	Query cover	Resolution	R-Value Free	R-Value Work
NCBI	1HNA-A	19	92 %	1.85	0.226	0.226
	1XW5-A	19	92 %	1.80	0.232	0.206
	2C4J-A	19	92 %	1.35	0.213	0.194
	2DC5-A	19	92 %	1.60	0.202	0.188
	1GSU-A	22	73 %	1.94	0.292	0.210
	4GTU-A	19	92 %	3.30	0.315	0.245
	6KLD-A	52%	8 %	3.58	-	-
	6GSX-A	17%	98%	1.91	-	0.52
PHYRE	1K3Y-B	22	98 %	1.30	0.207	0.137
	1GUM-A	22	97 %	3.00	0.270	0.253
	1UA5-A	21	97 %	2.50	0.221	0.169
	1B8X-A	20	97 %	2.70	0.310	0.209
	2FHE-A	20	98 %	2.30	0.234	0.183
	1BG5-A	20	99 %	2.60	0.359	0.193
	4W66-A	19	97 %	2.60	0.359	0.193
	1C72-A	19	89 %	2.80	0.280	0.190
	1GTU-B	18	98 %	2.68	0.245	0.211
	4Q5Q-A	18	97 %	1.93	0.227	0.180
i-TASSER	3ISO-A	20	95 %	1.90	0.226	0.179
	4Q5F-A	18	94 %	2.45	0.239	0.188
	2AAW-C	22	94 %	2.40	0.237	0.194
	2ON7	19	93 %	2.40	0.280	0.190
	1ZL9	21	93 %	2.01	0.209	0.151
	1ZL9-A	21	94 %	2.01	0.209	0.151
	4RI6	20	93 %	1.52	0.183	0.151
	1XW6-A	20	94 %	1.90	0.247	0.227
	1GUL-A	28	92 %	2.70	0.260	0.248
	1B8X-A	24	92 %	2.70	0.310	0.209

Table S6: Validation of glutathione transferases (GSTs) from *Cryptosporidium parvum* (C*p*GST1 and C*p*GST2) and *Cryptosporidium meleagridis* GST3 (*Cm*GST3) protein models. For comparative analysis templates were also included in the study.

	3VPQ	5AN1	C*p*GST1 post refinement	1K3Y	19GS	C*p*GST2 post refinement	1ZL9	*Cm*GST3 post refinement
Verify	91.63%	87.67%	87.98%	88.69%	100%	83.40%	95.17 %	26.29 %
Errat	95.8974	92.8571	85.6322	97.1831	99	85.9649	97.4874	82.9146
Procheck errors	1	2	3	2	0	3	0	2
Procheck Warnings	4	4	2	4	7	2	6	3
Procheck pass	4	3	3	3	2	3	3	3
Z-score	-8.08	-8.44	-6.71	-7.89	-7.66	-5.81	-7.79	-3.68

Table S7: Glutathione transferases (GSTs) from *Cryptosporidium parvum* (C*p*GST1 and C*p*GST2) and *Cryptosporidium meleagridis* GST3 (C*m*GST3) protein models assessment by Ramachandran Plot. For comparative analysis templates were also included in the study.

	3VPQ	5AN1	C*p*GST1 post refinement	1K3Y	19GS	C*p*GST2 post refinement	1ZL9	C*m*GST3 post refinement
Favoured Region	165 (92.7%)	176 (90.7%)	155 (92.3%)	182 (92.4%)	167 (93.3%)	199 (90.9%)	167 (93%)	185 (92%)
Additionally allowed region	11 (6.2%)	16 (8.2%)	11 (6.5%)	14 (7.1%)	10 (5.6%)	17 (7.8%)	11 (6.1%)	14 (7%)
Generously allowed region	4 (1.1%)	2 (1%)	2 (1.2)	1 (0.5%)	2 (1.1%)	3 (1.4 %)	1 (0.6%)	0
Disallowed region	0	0	0	0	0	0	0	2 (1%)

APPENDIX B

FASTA SEQUENCES OF GST FROM DIFFERENT CLASSES FOR PHYLOGENETIC TREE

***Cryptosporidium* species GSTs**

>C. andersoni 30847 GST1(cand_012830)

MNDSNYPYSVKSPLKLIYFACRGSCDVIRLLLNDKCIPYEEHNIQGKDFLQPEFQNVLLESDNFPILPYLSDP
NSEMELTGSLTILRYLGRKCNLMGNNYEDELQIENWLEYLQLVLNILWEFDSNSDNFNNIQKNKKRGQFL
LENLHPMLHNIQVRLDNGKKWIMEEYSVADIMLYTVVSAIIRSWGYEILQPYDK

>C. hominis_TU502_2012 GST1(ChTU502y2012_407g2365/Q18145.1)

MIEYLEHNISGKDFLQPEFQQVLVESGNFPMLPYFSDSNNEVELTGSFTILRYLADKCKLMGKSPEERNKI
ENWLEYLQSLLHSVWDFENMSDNYTGIQQAKKKSQFLLETLHPMLKCIDEKIEQGVWALESYSVVDIVLY
SAISVIIRSWGSDLLKPYIRILTHKKNMEKLRKQIDSFKDDPRRF

>C. hominis_30976 GST1(GY17_00002363)

MIEYLEHNISGKDFLQPEFQQVLVESGNFPMLPYFSDSNNEVELTGSFTILRYLADKCKLMGKSPEERNKI
ENWLEYLQSLLHSVWDFENMSDNYTGIQQAKKKSQFLLETLHPMLKCIDEKIEQGVWALESYSVVDIVLY
SAISVIIRSWGSDLLKPYIRILTHKKNMEKLRKQIDSFKDDPRRF

>C. hominis TU502 GST1(XP_667744.1)

MIEYLEHNISGKDFLQPEFQQVLVESGNFPMLPYFSDSNNEVELTGSFTILRYLADKCKLMGKSPEERNKI
ENWLEYLQSLLHSVWDFENMSDNYTGIQQAKKKSQFLLETLHPMLKCIDEKIEQGVWALESYSVVDIVLY
SAISVIIRSWGSDLLKPYIR

>C. hominis_UdeA01 GST1(CUV07467.1)

MIEYLEHNISGKDFLQPEFQQVLVESGNFPMLPYFSDSNNEVELTGSFTILRYLADKCKLMGKSPEERNK

IENWLEYLQSLLHSVWDFENMSDNYTGIQQAKKKSQFLLETLHPMLKCIDEKIEQGVWALESYSVVDIVL

YSAISVIIRSWGSDLLKPYIR

>C. meleagridis_UKMEL1 GST1(CmeUKMEL1_03350)

MEYIGSLDNPLRLIYFSCRGTCDVIRLLLVDQEIPYEEHNISGKDFLQPEFQQVLLESGNFPMLPYFSDSNN
EVELTGSFTILRYLADKCKLMGKSPEERNKIENWLEYLQSLLHSLWDFENMSDNYTGIQQTKKKRQFLLET
LHPMLKCIDEKIEQGVWALESYSVVDIVLYSAISVVIRSWGSDLLKPYIR

>C. parvum_Iowa II GST1(cgd7_4780)

MIEYLEHNISGKDFLQPEFQQVLVESGNFPMLPYLSDSNNEVELTGSFTILRYLADKCKLMGKSPEERNKIE
NWLEYLQSLLHSVWDFENMSDNYTGIQQAKKKSQFLLETLHPMLKCIDEKIEQGIWALESYSVVDIVLYS
AISVIIRSWGSDLLKPYIRILTHKKNMEKLRKQIDSFKDDPRRF

>C. tyzzeri_UGA55 GST1(CTYZ_00001095)

MIEYLEHNISGKDFLQPEFQQVLVESGNFPMLPYFSDSNNEVELTGSFTILRYLADKCKLMGKSPEERNKV
ENWLEYLQSLLHSVWDFENMSDNYTGIQQAKKKSQFLLETLHPMLKCIDEKIEQGVWALEYYSVVDIVLY
SAISVIIRSWGSDLLKPYIRILTHKKNMEKLRKQIDSFKDDPRRF

>C. ubiquitum_39726 GST1(cubi_03151)

MEYIGSLDNPLRLIYFSCRGTCDAIRLLLVDQEIPYEGKDFLQPEFQHVLAESGNFPMLPYLSDSNNEVELT
GSFTILRYLADKCKLMGNNSEERNRVENWLEFLQSLLHSVWDFENISENYTGVQQTKKKSKFLLDTLHPM
LKCIDDKIEQGFWALESYSVVDIVLYSTISVVIRSWGSDLLKPYIRILSHKKNMEKLRKQIDSFKDDPRRF

>C. muris RN66 GST1(XP_002141168.1)

MINVFHMKGKDFLQSEFQNVLLESDNFPILPYLSDPNSEIELTGSLTILRYLGRKCNLMGNNYEDELQIEN
WFEYLQLVLNILWEFDSNLDSFNNIQKNKKRGQFLLENLHPMLHNIQARLDNGKKWILEEYSVADIMLYT
VVSAIIRSWGYEILQPYDK

>C. viatorum isolate UKVIA1 GST1(QZWW01000010.1)

ILKYIEHNISGKDFLQPEFQQVLVESGNFPMLPYLSDSNNEVELTGSFTILRYLADKCKLMGKSPEERNKIE
NWLEYLQSLLHSVWDFENMSDNYTGIQHTKKKSQFLLETLHPMLKCIDEKIEQGVWALDSYSVVDVVLY
SAISVVIRSWGIDLLKPYIK

>C. baileyi strain TAMU-09Q1 GST1(JIBL01000090.1)

EHNISGKDFLQEEFQQVLIESGNFPMLPYLSDSNNEVELTGSFTILRYLGEKCKLMGNNVTERNKIENWLE
FLQSLLHSIWDFENNITNYTESQKKKRKSQFLLENLHPMLRSINDKIESNLWALNDYSIIDIVLYSTISVVIKL
WSIDLLKPYEK

>Cryptosporidium sp. chipmunk LX-2015 GST1 (JXRN01000042.1)

ILYFRGTCDVIRLLLVDQEIPYEGKLKRYCQLIIYTVFSSGIQKFLNILEHNISGKDFLQPEFQQVLVESGNFP
MLPYLSDSNNEVELTGSFTILRYLADKCKLMGKSPEERNKIENWLEYLQSLLHSVWDFENRSDNYTGVQQ
TKKRSQFLLETLHPMLKCIDEKIEHGVWVLDSYSVVDIVLYSAISVIIRSWGGDLLKPYTR

>C. baileyi strain TAMU-09Q1 GST2(JIBL01000106.1)

PMRSNVPCRLTNNRIMMPSKSTVKVVLPVRDIGELSVVTYEHDIFVGNGGSIRFFLLGKQVRHRFINVPL
DEENPIPSFIDSSRVPLGDLPIIKLGDLVLFDEIPCLRYLAKKLGEYGRNYYIDFVIDDIILRCSRWRDVIMEIIS
KGAAIVPNKNSVQDHINSLSNYKVLREKFYSEFETLITCIGERGPFIADKNKAMICDFALFSILFDDVSLMEI
NEHDQLNRTVLLPENCIIHKFPRLKLLFESIAALPLIDQWIKGKYFIIEVENDNNDQVNLLTSNLLQESNQSLI
NGYNNAFVNYPPMLRFQPFPFQHGNQIFAQANAGVRFFSPSTMPTNNYPIIPSSNSFISHSFVNYHPHL
WSNQFMGNYCSSGFSSPVQQKMSPAQSF

>C. viatorum isolate UKVIA1 GST2(QZWW01000018.1)

MNNIEASTTASPKIIAAKVSSELNEIYSPKMSKLVRNNIPCRLTTNRVMAPSRSTYRVILPVRDIGDLSVITY
EHEVYVGNGGSLRFFLLGKQVRHRFINVHLDEEDPIPSYIDPNKVPLGDLPVVKLGDLVIFDEIPCLRYLAKK
LGEYGRNYYIDFVIDDVIFRCSRWRDILMELISKSHKEFLNSNINANNELEKSISNYKLLREQLYCEFETLIASI
GDKGPFIAEKNKPMICDFILFSILFDDISLIEFNETEKFNRMKLLPEESIIHKFPRLKMLFESVATLPLIDQWV

KGKYFDIQIEGENSELLTPPASLSTQDYATNFVSGSNSFNGYQHSLGYQPPVFQQLPNQIFAHVNAGVRF
FPQKMSLPINQPIFSPNNSFISQPIANYHHHFLNNQIQSHGYLGGVSSPFIQKASPRSFKLKF

>C. ryanae isolate 45019 GST2(VHLK01000046.1)

MSKTIIPFRLTSNRIMIPSRSAVRVVLPVRDIGELTVITFEHDIFVGNGGSIRFFLLGKQVRHRFVNVPLDEE
KPIPSYIDSSRVPLGDLPIIKLGDLVLFDEIPCLRYLAKKLGEYGRNYYVDFVIDDIILRCSRWRDIIMELIPGSN
TGVSAKSYIEETNSLSNYKLLREQFYYEFETLITCIGDSGIFIADSNRPMICDFILFSILFDDISLVEIDENNQFN
RTTMIPENSIIHRFPRLKILFESISSLPLIEQWIKGKYFLVNIESESNIGDLITQKNSFPIMEHTNHFFKKEVGN
NQSSRFLPPIFQQPPNQIFGQATAGIRFLHQPINHINNLHRISPSNSFGLQPQFHAPQAWMHQINGNCY
FNNI

>C. andersoni 30847 GST2(cand_023790)

MIGVNSSISTGVASFSRDLSSLPGTSFIPAKAGSPQKSPSNLYGVIQAPRATSIRVMLPVRDIGDLTVVTYE
HQAFVGCGGSLRFFLLGKQVKHKFINVPVDKDNPIPDYIESSKVPLGELPIIKLGDLVIFDEIPCLRFLAKKLG
EYGRNYYIDFVIDDVVMRCNRWRDILMDLILSSNNCMLAASTNLDKSEAPYSLNNSENSGGSAISSLEGY
KQLREQLYTEFEVLITSIGEKEGSYIADKDKPMICDFALFSVLFDDINLSDISPDSMFQRIELLPDNCLIHQFP
RLKSLFLVMSELPLVNQWIKGKYFIQSNIDKATNKDSNENAAASTFPLQSSIVGNQQPSHSLYSLGAGNGI
ASPGVFSIYQSSHPLNPPIPRFQYPMIPYMPNQGLVQASAGVRFAFPGAGLPINNQQIPVIQANSSFINP
HFAPQLNPSLIHPFPIYQTNLGSPCNRMSPSQSFT

>C. hominis_TU502_2012 GST2(ChTU502y2012_421g0615)

MNNKETSTIPSPKTIASKISSELSEIYSPKMSTLVRNNIPCRLTSNRVMAPSKSTYRVILPVRDIGDLSVITYE
HEVYVGNGGSLRFFLLGKQVRHRFINVHLDEESPIPSYIDPNKVPLGDLPVVKLGDLVIFDEIPCLRYLAKKL
GEYGRNYYIDFVIDDVIFRCSKWRDILMELISKSHKEFLINDINAKKELERSISNYKLLREQLYCEFETLISTIG
DKGPFIAEKNKPMICDFILFSILFDDISLIEFNEGEKINRTSLLPEESLIHKFPRLKMLFESVAVLPLIDQWVKG
KYFSIQIEGESGELVTPPASLSTQDHVKNSVLGSNSFNVYQHSFGYQPPVLQQLPNQIFTHVNAGVRFFP
QKMSLPINPIFPTNNSFISQPITNNYHHFLNSQVQGHRYLGGVSSPFMQRVSPSQSFKLEF

>C. hominis_30976 GST2(GY17_00000733)

MNNKETSTIPSPKTIASKISSELSEIYSPKMSTLVRNNIPCRLTSNRVMAPSKSTYRVILPVRDIGDLSVITYE
HEVYVGNGGSLRFFLLGKQVRHRFINVHLDEESPIPSYIDPNKVPLGDLPVVKLGDLVIFDEIPCLRYLAKKL
GEYGRNYYIDFVIDDVIFRCSKWRDILMELISKSHKEFLINDINAKKELERSISNYKLLREQLYCEFETLISTIG
DKGPFIAEKNKPMICDFILFSILFDDISLIEFNEGEKFNRTSLLPEESLIHKFPRLKMLFESVAVLPLIDQWVKG
KYFSIQIEGESGELVTPPASLSTQDHVKNSVLGSNSFNVYQHSFGYQPPVLQQLPNQIFTHVNAGVRFFP
QKMSLPINPIFPTNNSFISQPITNNYHHFLNSQVQGHRYLGGVSSPFMQRVSPSQSFKLEF

>C. hominis_TU502 GST2(Chro.80347)

MNNKETSTIPSPKTIASKISSELSEIYSPKMSTLVRNNIPCRLTSNRVMAPSKSTYRVILPVRDIGDLSVITYE
HEVYVGNGGSLRFFLLGKQVRHRFINVHLDEESPIPSYIDPNKVPLGDLPVVKLGDLVIFDEIPCLRYLAKKL
GEYGRNYYIDFVIDDVIFRCSKWRDILMELISKSHKEFLINDINAKKELERSISNYKLLREQLYCEFETLISTIG
DKGPFIAEKNKPMICDFILFSILFDDISLIEFNEGEKFNRTSLLPEESLIHKFPRLKMLFESVAVLPLIDQWVKG
KYFSIQIEGESGELVTPPASLSTQDHVKNSVLGSNSFNVYQHSFGYQPPVLQQLPNQIFTHVNAGVRFFP
QKMSLPINPIFPTNNSFISQPITNNYHHFLNSQVQGHRYLGGVSSPFMQRVSPSQSFKLEF

>C. hominis_UdeA01 GST2(CHUDEA8_2970)

MNNKETSTIPSPKTIASKISSELSEIYSPKMSTLVRNNIPCRLTSNRVMAPSKSTYRVILPVRDIGDLSVITYE
HEVYVGNGGSLRFFLLGKQVRHRFINVHLDEESPIPSYIDPNKVPLGDLPVVKLGDLVIFDEIPCLRYLAKKL
GEYGRNYYIDFVIDDVIFRCSKWRDILMELISKSHKEFLINDINAKKELERSISNYKLLREQLYCEFETLISTIG
DKGPFIAEKNKPMICDFILFSILFDDISLIEFNEGEKINRTSLLPEESLIHKFPRLKMLFESVAVLPLIDQWVKG
KYFSIQIEGESGELVTPPASLSTQDHVKNSVLGSNSFNVYQHSFGYQPPVLQQLPNQIFTHVNAGVRFFP
QKMSLPINPIFPTNNSFISQPITNNYHHFLNSQVQGHRYLGGVSSPFMQRVSPSQSFKLEF

>C. meleagridis_UKMEL1 GST2(CmeUKMEL1_14570)

MNSKETSTISSPKIIASKISSESSEIYSPKISTLTRNSIPCRLTSNRVMASSKSTYRVILPVRDIGDLSVITYEHEV
YVGNGGSLRFFLLGKQVRHRFINVHLDEESPIPSYIDPNKVPLGDLPIVKLGDLVIFDEIPCLRYLAKKLGEY
GRNYYIDFVIDDVIFRCSKWRDILMELILKSHKEFLINDINTNKELERLISNYKLLREQLYCEFETLISSIGDKGP
FIAEKNKPMICDFILFSILFDDISLIEFNEGEKLNRTSLLPEESIIHKFPRLKMLFESVVMLPLIDQWVKGKYFSI
QIEGESGELVTPPTSLSTQDHVRNSVLGSNSFNWYQHSFGYQPPVPQQLPNQIFTHVSAGVRFFPQKVP
LPINPTFPTNNSFISQPITNNYHHFLNSQIQDHRYLGRASSPFIQRVSPSQSFKLEF

>C. parvum_Iowa II GST2(cgd8_2970)

MNNKETSTIPSPNIIASKISSELSEIYSPKMSTLVRNNIPCRLTSNRVMAPSKSTYRVILPVRDIGDLSVITYEH
EVYVGNGGSLRFFLLGKQVRHRFINVHLDEESPIPSYIDPNKVPLGDLPVVKLGDLVIFDEIPCLRYLAKKLG
EYGRNYYIDFVIDDVIFRCSKWRDILMELISKSRKEFLINDINANKELERSTSNYKLLREQLYCEFETLILSIGD
KGPFIAEKNKPMICDFILFSILFDDISLIEFNEGEKFNRTSLLPEESLIHKFPRLKMLFESVAVLPLIDQWVKGK
YFSIQIEGESGELVTPPASLSTQDHVKNSVLGSNSFNAYQHSFGYQPPVLQQLPNQIFTHVNAGVRFFPQ
KMSLPINPSIFPTNNSFISQPITNNYHHFFNSQVQGHRYLGGVSSPFIQRVSPSQSFKLEF

>C. tyzzeri_UGA55 GST2(CTYZ_0000322)

MNNKETSTIPSPKIIASKISSELSEIYSPKMSTLVRNNIPCRLTSNRVMAPSKSTYRVILPVRDIGDLSVITYEH
EVYVGNGGSLRFFLLGKQVRHRFINVHLDEESPIPSYIDPNKVPLGDLPVVKLGDLVIFDEIPCLRYLAKKLG
EYGRNYYIDFVIDDVIFRCSKWRDILMELISKSRKEFLINEINANKELERSISNYKLLREQLYCEFETLISSIGDK
GPFIAEKNKPMICDFILFSILFDDISLIEFNEGEKFNRTSLLPEESLIHKFPRLKMLFESVAVLPLIDQWVKGKY
FSIQIEGESGELVTPPASLSTQDYVKSSVLGSNSFNVYQHSFGYQPPVLQQLPNQIFTHVNAGVRFFPQK
MSLPINQSIFPTNNSFISQPITNNYHHFLNRQVQGHRYLGGVSSPFIQRVSPSQSFKLEF

>C. ubiquitum_39726 GST2(cubi_03523)

MNNIGAGTTASPKNIATKVSSELNEIYSPKMSNLIRSNAPCRLTSNRVMIPSKSTYRVILPVRDIGDLSVITY
EHEIYLGNGGSLRFFLLGKQVRHRFINVPLDEENPIPSYIDSDKVPLGDLPIVKLGDLVIFDEIPCLRYLAKKL
GEYGRNYYIDFVIDDVIFRCSKWRDVLMDLISRNYSELSNGNINTNKELESSISNYKLLREQLYCEFETLIASI
GDKGPFIAEKNKPMICDFILFSILFDDISLIEFSETEKFNRVTLLPERSIIHKFPRLKMLFESVAILPLIDQWIKG
KYFAIQIEGESSELVTPPTSLTTQDHGTNFVVGTNSFIGCPNSFGYQPPVFQQLPNQLFAHVNAGIRFFPQ
NMAMPINQPIFSPNNSFVSQPITNYYPFLNNQIQNHGYLGGVSSPFVQRISPSQSFKLKF

>*C. bovis* isolate 42482 GST2(VHIT01000012.1)

MIPSRSAVRVVLPVRDIGELTVVTFEHNVFVGNGGSIRFFLLGKQVRHRFVNVPLDEEKPIPSYIDSSRVPL
GDLPIVKLGDLVLFDEIPCLRYLAKKLGEYGRNYYVDFIIDDIILRCSRWRDIIMELITENNTGSYSKNHVKGE

EINPISNYKLLREQFYYEFETLITCIGESGIFIADGNKPMICDFILFSILFDDISLVEIDDNNQFNRTAMLPENSI
IHRFPRLKILFESISSLPLIEQWIKGKYFIVNVEGEINAKDLTAQKNSFPIMEHSNNFYRQDLEHSPLPRFSPPI
FQQFPGQVFAQTTAGIRFVPQPFNHMNSFHRISPSNSFNLQPRSYAPQTWMHQINGSCYFNNITSPM
QFRASPNPSFRM

>C. meleagridis_UKMEL1 GST3(CmeUKMEL1_05845)

MKSISLLASVFAFLALFSTSVESVKAKARIIPITFYSTKELDSNHLIRTVLVYSGLAFAETRFKKDSESQAKLFKE
ITKSGFLQPSIPMISDTGKNVQYLSTDEAVLNYIILSYNKELFSKNLLLHTISIQLSSIARSYIKKTTKILDSSKTL
TCSKLLTNENIHQTLKVLNDTFASTEHKFLIGNKVSFNDLIAYNLILFIENVASGCVISNFKGLRELAFNISSIP
QIAKFESSSYFMSLLVPGTHTFAQRINFAHSSPMFLSLTS

>C. baileyi strain TAMU-09Q1 GST3(JIBL01000138.1)

PMYLYTTKELDNTQILRSLMVVSSLPFYEVRFTKDSEAKKFFFDKIKSLGYLTPSIPVLSDPETFNSYISTEEAI
SQYILLSYYKELYPSTISEYIYSIQAASLMTSYMKKLTNILSESITLPCTKILTLNDIKHLLNVLEKKRSESKSKYFY
GEKYTYIDVSLYNLILFIENVSPGCVIRRYPSLTKLAFEFSQIPQVLAYERSPHFLSLTIPGTRAFAKPINFVLM
SKAFDTLS

>C. ryanae isolate 45019 GST3(VHLK01000056.1)

MFYTSKTIDNSHLIRTLLVLSGIPFNEFRFKKNSPSLEEMFNSVVESGFLVPSIPMITDNEYSVKNISQEEAII
HYLILSYYPDLFPKVISDYAISLQIGSAVRSYIQKVHKIIELSQKLVCEKLLTIDNINITLKLLDDKFIETGSRFYFG
GRYSYFDASVYTLILFVENISSGCITSNYEGLKAFSKEFSSISQISKFEKSSYFLSLIVPGTTRFVKPIDFVSQAH
ES

>*C.* bovis isolate 42482 GST3(PRJNA545579)

MFYTSKTLDSTHLIRTLLILSSLPFNEFRFKKNSSSMEEMFSSIIESGFLNPTIPMISDNEYSVRNLSQDEAIV
HYLVLSYYGELFQKSISDHAISLQIGSTVRSYISKVSGLLELSESLKCEKLLEIENVNVTLRLVNDRFTDTEYKFF
YGGKYSYIDTVVYTLILFIENISNGCIISNFDGLRSFSKEFSSIPQISKFEKSSYFLSLLIPGTKEFVKPIDFVTQS

Alpha Class GSTs

>Bos Taurus(Q28035)

MAGKPTLHYFNGRGRMECIRWLLAAAGVEFEEKFIEKPEDLDKLKNDGSLMFQQVPMVEIDGMKLVQ
TRAILNYIATKYNLYGKDMKERALIDMYSEGVADLGEMIMHFPLCPPAEKDAKLTLIREKTTNRYLPAFEN
VLKSHGQDYLVGNKLSRADIHLVELLYYVEELDPSLLANFPLLKALKARVSNIPAVKKFLQPGSQRKPPTDE
KKIEEARKVFKF

>Gallus gallus(P26697)

AKPVLYYFNGRGKMESIRWLLAAAGVEFEEVFLETREQYEKLLQSGILMFQQVPMVEIDGMKLVQTRAIL
NYIAGKYNLYGKDLKERALIDMYVGGTDDLMGFLLSFPFLSAEDKVKQCAFVVEKATSRYFPAYEKVLKD
HGQDFLVGNRLSWADIHLLEAILMVEEKKSDALSGFPLLQAFKKRISSIPTIKKFLAPGSKRKPI

>Homo sapiens(P08263)

EKPKLHYFNARGRMESTRWLLAAAGVEFEEKFIKSAEDLDKLRNDGYLMFQQVPMVEIDGMKLVQTRA
ILNYIASKYNLYGKDIKERALIDMYIEGIADLGEMILLLPVCPPEEKDAKLALIKEKIKNRYFPAFEKVLKSHGQ
DYLVGNKLSRADIHLVELLYYVEELDSSLISSFPLLKALKTRISNLPTVKKFLQPGSPRKPP

>Rattus norvegicus(P00502)

GKPVLHYFNARGRMECIRWLLAAAGVEFDEKFIQSPEDLEKLKKDGNLMFDQVPMVEIDGMKLAQTRA
ILNYIATKYDLYGKDMKERALIDMYTEGILDLTEMIMQLVICPPDQKEAKTALAKDRTKNRYLPAFEKVLKS
HGQDYLVGNRLTRVDIHLLELLLYVEEFDASLLTSFPLLKAFKSRISSLPNVKKFLQPGSQRKLPV

>Sus scrofa(P51781)

MAGKPILHYFNGRGRMECIRWLLAAAGVEFEEKFIKTPEDLDKLTNDGSLLFQQVPMVEIDGMKLVQTR
AILNYIATKYNLYGKDAKERALIDMYTEGVADLGEMILLLPLCPPNEKDAKVASIKEKSTNRYLPAFEKVLKS
HGQDYLVGNKLSRADIQLVELLYYVEELDPSLLANFPLLKALKTRVSNLPTVKKFLQPGSQRKPPMDAKKI
RRSQEYFPD

>Gallus gallus(Q9W6J2)

MSGKPRLTYVNGRGRMESIRWLLSAAGVEFEEIFLETREQLLKLCQDGSLLFHQLPLVEIDGMKLVQCRAI
LSYIAGKYNLYGKDLKERALIDMYVEGISDLMQLILVFPFSPPEAKEKNLATIAEKATERYFPVFEKVLKQHG
QDFLVGNRFSWADVQLMEAILAVEEKVPSVLSGFPQLQAFKTKMSNMPTIKKFLQPGSPRKPPPDEHYV
ATVKKIFKLN

>Homo sapiens(O15217)

MAARPKLHYPNGRGRMESVRWVLAAAGVEFDEEFLETKEQLYKLQDGNHLLFQQVPMVEIDGMKLV
QTRSILHYIADKHNLFGKNLKERTLIDMYVEGTLDLLELLIMHPFLKPDDQQKEVVNMAQKAIIRYFPVFEK
ILRGHGQSFLVGNQLSLADVILLQTILALEEKIPNILSAFPFLQEYTVKLSNIPTIKRFLEPGSKKKPPPDEIYVR
TVYNIFRP

>Homo sapiens(P09210)

MAEKPKLHYSNIRGRMESIRWLLAAAGVEFEEKFIKSAEDLDKLRNDGYLMFQQVPMVEIDGMKLVQT
RAILNYIASKYNLYGKDIKEKALIDMYIEGIADLGEMILLLPFSQPEEQDAKLALIQEKTKNRYFPAFEKVLKS
HGQDYLVGNKLSRADIHLVELLYYVEELDSSLISSFPLLKALKTRISNLPTVKKFLQPGSPRKPPMDEKSLEES
RKIFRF

>Homo sapiens(Q7RTV2)

MAEKPKLHYSNARGSMESIRWLLAAAGVELEEKFLESAEDLDKLRNDGSLLFQQVPMVEIDGMKLVQTR
AILNYIASKYNLYGKDMKERALIDMYTEGIVDLTEMILLLLICQPEERDAKTALVKEKIKNRYFPAFEKVLKS
HRQDYLVGNKLSWADIHLVELFYYVEELDSSLISSFPLLKALKTRISNLPTVKKFLQPGSQRKPPMDEKSLEE
ARKIFRF

>Rattus norvegicus(P04903)

MSGKPVLHYFNARGRMECIRWLLAAAGVEFEEKLIQSPEDLEKLKKDGNLMFDQVPMVEIDGMKLAQT
RAILNYIATKYDLYGKDMKERALIDMYSEGILDLTEMIIQLVICPPDQREAKTALAKDRTKNRYLPAFEKVLK
SHGQDYLVGNRLTRVDIHLLELLLYVEEFDASLLTSFPLLKAFKSRISSLPNVKKFLQPGSQRKPAMDAKQIE
EARKVFKF

>Rattus norvegicus(P04904)

MPGKPVLHYFDGRGRMEPIRWLLAAAGVEFEEQFLKTRDDLARLRNDGSLMFQQVPMVEIDGMKLVQ
TRAILNYIATKYNLYGKDMKERALIDMYAEGVADLDEIVLHYPYIPPGEKEASLAKIKDKARNRYFPAFEKV
LKSHGQDYLVGNRLSRADVYLVQVLYHVEELDPSALANFPLLKALRTRVSNLPTVKKFLQPGSQRKPLEDE
KCVESAVKIFS

>Mus musculus(P10648)

MAGKPVLHYFNARGRMECIRWLLAAAGVEFEEKFIQSPEDLEKLKKDGNLMFDQVPMVEIDGMKLVQ
TRAILNYIATKYDLYGKDMKERALIDMYTEGILDLTEMIGQLVLCPPDQREAKTALAKDRTKNRYLPAFEK
VLKSHGQDYLVGNRLTRVDVHLLELLLYVEELDASLLTPFPLLKAFKSRISSLPNVKKFLHPGSQRKPPLDAK
QIEEARKVFKF

>Mus musculus(P13745)

MAGKPVLHYFNARGRMECIRWLLAAAGVEFEEKFIQSPEDLEKLKKDGNLMFDQVPMVEIDGMKLAQ
TRAILNYIATKYDLYGKDMKERALIDMYSEGILDLTEMIGQLVLCPPDQREAKTALAKDRTKNRYLPAFEK
VLKSHGQDYLVGNRLTRVDIHLLEVLLYVEEFDASLLTPFPLLKAFKSRISSLPNVKKFLQPGSQRKPPMDA
KQIQEARKAFKIQ

>Rattus norvegicus(P14942)

MEVKPKLYYFQGRGRMESIRWLLATAGVEFEEEFLETREQYEKLQKDGCLLFGQVPLVEIDGMLLTQTRA
ILSYLAAKYNLYGKDLKERVRIDMYADGTQDLMMMIIGAPFKAPQEKEESLALAVKRAKNRYFPVFEKILK
DHGEAFLVGNQLSWADIQLLEAILMVEEVSAPVLSDFPLLQAFKTRISNIPTIKKFLQPGSQRKPPPDGHY
VDVVRTVLKF

>Antechinus stuartii(P80894)

MAGEQNIKYFNIKGRMEAIRWLLAVAGVEFEEKFFETKEQLQKLKETVLLFQQVPMVEIDGMKLVQTRA
ILHYIAEKYNLLGKDMKEHAQIIMYSEGTMDLMELIMIYPFLKGEEKKQRLVEIANKAKGRYFPAFENVLK
THGQNFLVGNQLSMADVQLFEAILMVEEKVPDALSGFPLLQAFKTRISNIPTVKTFLAPGSKRKPVPDAK
YVEDIIKIFYF

>Cavia porcellus(P81706)

SGKPVLHYFNVQGRMESIRWLLAAAGVEFEEKLIMCQEDLDKLKNDGLLMFQQVPMVEMDGMKMV
QSRAILNYIATKYNLYGKDTKERLLIDMYTEGMTDLYELFFKVILAPPEEKDAAKSLIKDRAKNRFLPAFEKV
LKSHGQGYLVGNKLSKADILLTELLYMVEEFDASLLANFTLLQALKTRVSNLPNVKKFLQPGSQRKPFPTQ
EMFEEMRKF

>Gallus gallus(Q08393)

MAGKPKLHYTRGRGKMESIRWLLAAAGVEFEEEFIEKKEDLEKLRNDGSLLFQQVPMVEIDGMKMVQS
RAILCYIAGKYNLYGKDLKERAWIDMYVEGTTDLMGMIMALPFQAADVKEKNIALITERATTRYFPVYEK
ALKDHGQDYLVGNKLSWADIHLLEAILMTEELKSDILSAFPLLQAFKGRMSNVPTIKKFLQPGSQRKPPLD
EKSIANVRKIFSF

>Oryctolagus cuniculus(Q08863)

MARKPLLHYFNGRGRMESIRWLLAAAGEEFDEKFMETAEDLDKLRNDGSLMYQQVPMVEIDGMKLV
QTRAILNYVANKHNLYGKDMKERALIDMYTEGVADLYELVLLLPLCPPEQKDAKVDFIKEKIRTRYFPAFE
KVLKSHGQDYLVGNRLSKADILLVELLYNVEELDPSAIASFPLLKALKTRISSLPTVKKFLQPGSQRKPPMDE
KNLEKAKKIFKIP

>Homo sapiens(Q16772)

MAGKPKLHYFNGRGRMEPIRWLLAAAGVEFEEKFIGSAEDLGKLRNDGSLMFQQVPMVEIDGMKLVQ
TRAILNYIASKYNLYGKDIKERALIDMYTEGMADLNEMILLLPLCRPEEKDAKIALIKEKTKSRYFPAFEKVLQ
SHGQDYLVGNKLSRADISLVELLYYVEELDSSLISNFPLLKALKTRISNLPTVKKFLQPGSPRKPPADAKALEE
ARKIFRF

>Bos Taurus(O18879)

GKPKLHYFNGRGRMECIRWLLAAAGVEFEEKFIEQPEDLDKLRNDGSLMFQQVPMVEIDGMKLVQTRA
ILNYIATKYNLYGKDMKERALIDMYSEGVEDLGEMIMHLPLCPPDQKDAKIAQIKERTTNRYFPAFEKVLK
NHGQDYLVGNKLSKADIHLVELLYYVEELDPSLLANFPLLKGLKARVSSLPAVKKFLQPGSQRKPPM

>Rattus norvegicus(P46418)

GKPVLHYFDGRGRMEPIRWLLAAAGVEFEENFLKTRDDLARLRSDGSLMFEQVPMVEIDGMKLVQTKA
ILNYIATKYNLYGKDMKERALIDMYAEGVADLELMVLYYPYMPPGEKEASLAKIKDKARNRYFPAYEKVLK
SHGQDYLVGNKLSRADVSLVELLYHVEEMDPGIVDNFPLLKALRTRVSNLPTVKKFLQPGSQRKPF

>Mus musculus(P24472)

AKPKLYYFNGRGRMESIRWLLAAAGVEFEEEFLETREQYEKMQKDGHLLFGQVPLVEIDGMMLTQTRAI
LSYLAAKYNLYGKDLKERVRIDMYADGTQDLMMMIAVAPFKTPKEKEESYDLILSRAKTRYFPVFEKILKD
HGEAFLVGNQLSWADIQLLEAILMVEELSAPVLSDFPLLQAFKTRISNIPTIKKFLQPGSQRKPPP

>Rattus norvegicus(Q6AXY0)

MAEKPLFHYDEARGRMESVRWLLAAAGVEYEEKFIHTNEDLEKLRSDGVLMFQQVPMVEVDGMKLVQ
TRAIMNYFSSKYNLYGKDMKERALIDMYSEGLADLNEMFILYPFDPPGVKEANIALMKEKATNRYFPAFE
KVFESHGQDYLVGNKLSKADVHLVEMIYNMEELDTNILANFPLLQALKTRISDMPTIKKFLQPGSQRQPP
VDEKSIQKTRKIFKF

>Macaca mulatta(A0A023JCQ7)

MAGKPKLHYFNGRGRMEPIRWLLAAAGVEFEEKFIESAEDLGKLRNDGSLMFQQVPMVEIDGMKLVQ
TRAILNYIASKYNLYGKDIKERALIDMYTEGMADLNEMILLLPLCRPEEKDAKIALIKEKTKNRYFPAFEKVL
QSHRQDYLVGNKLSRADISLVELLYYVEELDSSLISSFPLLKALKTRISNMPTVKKFLQPGSPRKPPPDAKAL
EEARKIFRF

>Canis lupus familiaris(A0A059V712)

MAVKPMLHYFNGRGRMESIRWLLASAGVEFEEKFINTPEDLDKLKNDGSLMFQQVPMVEIDGMKLVQ
TRAILNYIATKYNLYGKDIKERALIDMYTEGIVDLNEMIMVLPLCPPDQKDAKITLIRERTTDRYLPVFEKVL
KSHGQDYLVGNKLSRADIHLVELLYYVEELDSSLLANFPLLKALKTRVSNLPTVKKFLQPGSPRKPPLDEKSL
EQAKKIFRIN

>Aptenodytes forsteri(A0A087RJW0)

MAGKLKLYYFDGRGKMESIRWLLAAAGVEFEEEFLETREQYEKLLQGGSLLFQQLPMVEMDGMKMVQ
PRAILSYIAAKYNLYGKDLKERALIDMYVGGTDDLMGFILMFPFLSDEDKEKQRAVIVQKATSRYFPAYEK
VLKDHGQDFLVGNNFSWADVHLLEAILMVEEKKSDVLSGFPQLQAFKARISSIPTIKKFLEPGSQRKPVPD
DKYVETVKRVLRVYYDIKA

>Aptenodytes forsteri(A0A087RJW2)

MSGKPKLYYFNARGRMESIRWLLAAAGVEFEECFLETKDDLTKLRKDGSLLFQQVPMVEIDGMKMVQS
RAIGNYIAMKYNLYGKDLKERALIDMYVEAVIDLNELLMTHPFQPADKKEQHFATIVDKATNRYFPVYEK
VLKDHGQDFLVGNQFSRADVQLLETLLMAEECKPDILAKFPLLKSFKARISNIPTIKKFLQPGSQRKPPLQE
KDVAKLMKIFH

>Poecilia formosa(A0A087XSY0)

MSGKVILHYFNGRGKMESIRWLLTVAEAEFDEYYLTDREQYLKLLNEGSLMFQQVPLVEIDGLKLVQTKAI
LHYIAEKYNLYGKDIKERAMINMYAEGLIDHMEMIMVLPFVTDTKPKLDNIQSKAKERYLPVFEKALTGP
VYLVGGKLSLADVLLVECTLMLEEKFPDILKEFPNIKSFQGRMIRIPAISRFLQPGSKRKPQPDEDYVKTVKE
VFNITGPFP

>Corvus brachyrhynchos(A0A091F0Y3)

MSGKPRLTYLNGRGRMEPIRWLLAAAGVEFEEVYLETKEQYDKLIKDGFLLFQQVPLVEIDGMKMVQTR
AILSYIAGKYNLYGKDLKERALIDMYVEGIADLMQMILMFPFSPPDAKEKNLDSVKERATNRYFPVFEKVL
KQHGQDFLVGNKFSWADVQLMEAIAFKTKMSNMPTIKKFLQPGSPRKPPPDAHYVETVLKVFKK

>Cuculus canorus(A0A091GFV5)

MAGKPKLYYFNGRGKMESVRWLLAAAGVEFEEEFLETQEQYEKLLQGGSLLFQQVPMVEIDGMKMVQ
TRAILSYIAAKYNLYGKDLKERALIDMYVGGTDDLMGFIMMFPFLSAEDKEKQRATIVQKATSRYFPAYEK
ILKDHGQNFLVGGSFSWADVHLLEAILMVEEKKSDVLSGFPQLQAFKARISSIPTIKKFLEPGSQRKPLPDD
KYVETVRRVLRMYYDVKAN

>Cuculus canorus(A0A091GJB1)

MAAKPKLHYPKGRGKMESIRWLLAAAGVEFDEQFIEKKEDLEKIRNDGSLLFQQVPMVEIDGMKMVQT
RAILSYIAAKYNLYGKDLKERAWIDMYVEGTTDLMGMIMSLPLQTAETKDKHLALIIERATTRYFPVYEKA
FKTHGQDYLVGNKLSWADIQLLEAILMAEECKADILSAFPQLQAFKGRISNVPTIKKFLQPGSKRKPRPDE
KYIAEMRKIFNF

>Buceros rhinoceros silvestris(A0A091H7X1)

MSGKPKLHYFNGRGRMESIRWLLAAAGVEFEECFLETKDDLIKLQKGGSLLFQQVPMVEIDGMKMVQT
RAILNYIAAKYNLYGKDLKERALIDMYVEAIADLNELLMSHLFQPADKKEEHFATVVDKATNRYFPVYEKV
LKDHGQDFLVGNQFSRADVQLLETLLMAEECKPDILARFPLLKSFKARISNIPTIKKFLQPGSQRKPPLQEE
DIPKVMKIFH

>Calypte anna(A0A091I458)

MSGKPRLTYLNGRGRMEPVRWLLAAAGVEFEEVFLETREQYEKLIKDGVLMFQQVPLVEIDGMKMVQ
TRAILSYIAGKYNLYGKDLKERALIDMYVEGISDLMHMILMFHFSPPDAKEKNIDSVKDRATNRYFPVFEK
VLKQHGQDFLVGNKFSWADVQLTEAILAVEEKIPAVLSEFPQLQAFKVRMTNMPTIKKFLQPGSPRKPP
PDDHYVETVIKIFK

>Egretta garzetta(A0A091J816)

MSGKPKLHYFNGRGRMESIRWLLAAAGVEFEECFLETKDDLVKLQKDGSLLFQQVPMVEIDGMKMVQ
TRAISNYIATKYNLYGKDLKERALIDMYVEAVLDLNELLMTHAFQPADKKEQHFATIVDKATNRYFPVYEK
VLKDHGQDFLVGNRFSRADVQLLETLLMAEECKPDILANFPLLQSFKARISNMPTIKKFLQPGSQRKPPLQ
EKDVPKLMKIFH

>Nestor notabilis(A0A091S0F7)

MSGKPRLTYLNGRGRMESIRWLLAAAGVEFEEIFLETKEQYEKIIKDGILMFQQVPLVEIDGMKMVQTRA
ILSYVAGKYNLYGKDLKERALIDMYVEGIIDLMQMILMFPFSPPEAKEKNLDSIKERATNRYFPVFEKVLKQ
HGQDFLVGNKFSWADVQLIEAILAVEEKVPAVLSGFPQLQVILLLSINMPTIKKFLQPGSPRKPPPDEHYV
ETVLKIFSK

>Mesitornis unicolor(A0A091SHE9)

MAGKPKLHYTKGRGKMESIRWLLAAAGVEFEEEFIEKKEDLDKLLKGGVLMFQQVPMVEIDGMKMVQ
TRAILSYIAGKYNLYGKDLKERAWIDMYVEGTTDLMGMIMVIPLQAADAKEKQLALIIERATTRYFPVYEK
ALKDHGHDYLVGNKLSWADIQLLEAILMTEECKPDILSAFPLLQAFKGRISNIPTIKKFLQPGSQRKPPPDD
KYIASVRKIF

>Pelecanus crispus(A0A091SJF8)

MSGKPRLTYCNGRGRMEPVRWLLAAAGVEFEEIFLETREQYEKLIKDGVLMFQQVPLVEIDGMKMVQT
RAILSYIAGKYNLYGKDLKERALIDMYVEGITDLMQMILMFPFSPPEAKEKNLDSIKERATNRYFPVFEKVL
KQHGQEFLVGNKFSWADVQLIEAILAVEEKIPAVLSEFPQLQVI

Beta Class GSTs

>Escherichia coli(strain K12)(P0A9D2)

MKLFYKPGACSLASHITLRESGKDFTLVSVDLMKKRLENGDDYFAVNPKGQVPALLLDDGTLLTEGVAIM
QYLADSVPDRQLLAPVNSISRYKTIEWLNYIATELHKGFTPLFRPDTPEEYKPTVRAQLEKKLQYVNEALKD
EHWICGQRFTIADAYLFTVLRWAYAVKLNLEGLEHIAAFMQRMAERPEVQDALSAEGLK

>Proteus mirabilis(P15214)

MKLYYTPGSCSLSPHIVLRETGLDFSIERIDLRTKKTESGKDFLAINPKGQVPVLQLDNGDILTEGVAIVQYL
ADLKPDRNLIAPPKALERYHQIEWLNFLASEVHKGYSPLFSSDTPESYLPVVKNKLKSKFVYINDVLSKQKC
VCGDHFTVADAYLFTLSQWAPHVALDLTDLSHLQDYLARIAQRPNVHSALVTEGLIKE

>Haemophilus influenzae(P44521)

MKLYGLIGACSFVPHVALEWVKIRENADYEFEPVTRELIKSPEFLSLNPRGAVPVLVDGDLVLSQNQAILH
YLDELYPNSKLFGSKTVRDKAKAARWLAFFNSDVHKSFVPLFRLPNYAKDNETLAHTIRQQAVEQILDQL
AVANEHLESHIYFGENISVADAYLYIMLNWCKAVKIDFSHLTQLSAFMQRVETDQAVENVRKSEELKV

>Xylella fastidiosa(Q9PE18)

MKLYIMPGACSLADHILLRWSGSSFDLQFLDHQSMKAPEYLALNPSGAVPALQVGDWVLTQNAAILNYI
TDIAPAERGLSGDGSLKARAEINRWIAFSNSDVHPMYWALFGGTAYLQDPQMIARSQDNARQKLRVLY
QRADAHLKHHNWLANGQRSGADAYLYVTLRWAKKVGVDLSSLDALSAFFERMEADPGVQAALQAEG
LI

>Xanthomonas campestris pv campestris(P45875)

MKLYTKPGACSLADHIVLRWSCLPFELTVVDAATMKSPDYLRLNPAGAVPLLVVDQWALTQNAAILNYI
ADTAPLTGLGGDGTARSRAEINRWIAFVNADLHPTFKPLFGSTAYLQEDALIQRSHEDARTKLRTLYTRVD
AHLQGRNWLAGDTHTGADAYLFVTLRWAHKAGVDLSGLSALDAFFQRMLADADVQAALQAEGLN

>Ochrobactrum anthropi(P81065)

MKLYYKVGACSLAPHIILSEAGLPYELEAVDLKAKKTADGGDYFAVNPRGAVPALEVKPGTVITQNAAILQ
YIGDHSDVAAFKPAYGSIERARLQEALGFCSDLHAAFSGLFAPNLSEEARAGVIANINRRLGQLEAMLSDK
NAYWLGDDFTQPDAYASVIIGWGVGQKLDLSAYPKALKLRERVLARPNVQKAFKEEGLN

CLIC GSTs

>Homo sapiens(O00299)

MAEEQPQVELFVKAGSDGAKIGNCPFSQRLFMVLWLKGVTFNVTTVDTKRRTETVQKLCPGGQLPFLLY
GTEVHTDTNKIEEFLEAVLCPPRYPKLAALNPESNTAGLDIFAKFSAYIKNSNPALNDNLEKGLLKALKVLD
NYLTSPLPEEVDETSAEDEGVSQRKFLDGNELTLADCNLLPKLHIVQVVCKKYRGFTIPEAFRGVHRYLSNA
YAREEFASTCPDDEEIELAYEQVAKALK

>Homo sapiens(Q9NZA1)

CPFSQRLFMILWLKGVVFNVTTVDLKRKPADLHNLAPGTHPPFLTFNGDVKTDVNKIEEFLEETLTPEKYP
KLAAKHRESNTAGIDIFSKFSAYIKNTKQQNNAALERGLTKALKKLDDYLNTPLPEEIDANTCGEDKGSRRK
FLDGDELTLADCNLLPKLHVVKIVAKKYRNYDIPAEMTGLWRYLKNAYARDEFTNTCAADSEIELAYADV
AKRL

>Rattus norvegicus(Q9EPT8)

CPFSQRLFMILWLKGVVFNVTTVDLKRKPADLHNLAPGTHPPFLTFNGDVKTDVNKIEEFLEETLTPEKYP
KLAARHRESNTAGIDIFSKFSAYIKNTKQQNNAALERGLTKALRKLDDYLNTPLPEEIDTNTHGDEKGSQR
KFLDGDELTLADCNLLPKLHVVKIVAKKYRNYDIPAEMTGLWRYLKNAYARDEFTNTCAADSEIELAYADV
ARRL

>Bos Taurus(P35526)

QRLFMILWLKGVVFNVTTVDLKRKPADLHNLAPGTHPPFLTFNGDVKTDVNKIEEFLEETLTPEKYPRLAA
KHRESNTAGIDIFVKFSAYIKNTKQQSNAALERGLTKALKKLDDYLNTPLPEEIDADTRGDDEKGSRRKFLD
GDELTLADCNLLPKLHVVKIVAKKYRNYDFPAEMTGLWRYLKNAYARDEFTNTCAADSEIELAYADVAKR
L

>Homo sapiens(O15247)

MSGLRPGTQVDPEIELFVKAGSDGESIGNCPFCQRLFMILWLKGVKFNVTTVDMTRKPEELKDLAPGTN
PPFLVYNKELKTDFIKIEEFLEQTLAPPRYPHLSPKYKESFDVGCNLFAKFSAYIKNTQKEANKNFEKSLLKEF
KRLDDYLNTPLLDEIDPDSAEEPPVSRRLFLDGDQLTLADCSLLPKLNIIKVAAKKYRDFDIPAEFSGVWRYL
HNAYAREEFTHTCPEDKEIENTYANVAKQKS

>Homo sapiens(O95833)

MAETKLQLFVKASEDGESVGHCPSCQRLFMVLLLKGVPFTLTTVDTRRSPDVLKDFAPGSQLPILLYDSDA
KTDTLQIEDFLEETLGPPDFPSLAPRYRESNTAGNDVFHKFSAFIKNPVPAQDEALYQQLLRALARLDSYLR
APLEHELAGEPQLRESRRRFLDGDRLTLADCSLLPKLHIVDTVCAHFRQAPIPAELRGVRRYLDSAMQEKE
FKYTCPHSAEILAAYRPAVHPR

>Rattus norvegicus(Q5M883)

MASLALNTQADPEIELFVKAGSDGESIGNCPFCQRLFMILWLKGVKFNVTTIDTARKPEELKDLAPGTNP
PFLIYNKELKTDFIKIEEFLEKTLAPPRYPHLSPKYKESFDVGCNLFAKFSAYIKNTQKEANKNFEKSLLREFKR
LDDYLNTPLLDEIDPDSTEERTLSRRLFLDGDQLTLADCSLLPKLNIIKVAAKKYRDFDIPAEFSGVWRYLHN
AYAREEFAHTCPEDKEIENTYASVAKQ

>Bos taurus(Q5E9B7)

MAEEQPQVELFVKAGSDGAKIGNCPFSQRLFMVLWLKGVTFNVTTVDTKRRTETVQKLCPGGQLPFLLY
GTEVHTDTNKIEEFLEAVLCPPRYPKLAALNPESNTAGLDIFAKFSAYIKNSNPALNDNLEKGLLKALKVLD
NYLTSPLPDEVDETSAEDEGISQRKFLDGNELTLADCNLLPKLHIVQVVCKKYRGFSIPDVFRGVHRYLRNA
YAREEFASTCPDDEEIELAYEQVAKALK

>Rattus norvegicus(Q6MG61)

MAEEQPQVELFVKAGSDGAKIGNCPFSQRLFMVLWLKGVTFNVTTVDTKRRTETVQKLCPGGQLPFLLY
GTEVHTDTNKIEEFLEAVLCPPRYPKLAALNPESNTSGLDIFAKFSAYIKNSNPALNDNLEKGLLKALKVLD
NYLTSPLPEEVDETSAEDEGISQRKFLDGNELTLADCNLLPKLHIVQVVCKKYRGFTIPEAFRGVHRYLSNA
YAREEFASTCPDDEEIELAYEQVARALK

>Mus musculus(Q8BXK9)

MTDSATTNGDDRDPEIELFVKAGIDGESIGNCPFSQRLFMILWLKGVVFNVTTVDLKRKPADLHNLAPGT
HPPFLTFNGDVKTDVNKIEEFLEETLTPEKYPKLAAKHRESNTAGIDIFSKFSAYIKNTKQQNNAALERGLT
KALRKLDDYLNSPLPEEIDTNTHGDEKGSQRKFLDGDELTLADCNLLPKLHVVKIVAKKYRNYDIPAEMTG
LWRYLKNAYARDEFTNTCAADSEIELAYADVARRLSRS

>Oryctolagus cuniculus(Q95MF9)

MAEEQPQVELFVKAGSDGAKIGNCPFSQRLFMVLWLKGVTFNVTTVDTKRRTETVHKLCPGGQLPFLLY
GTEVHTDTNKIEEFLEAVLCPPRYPKLAALNPESNTAGVDIFAKFSAYIKNSNPALNDNLEKGLLKALKILDN
YLTSPLPEEVDETSAEDEGISQRKFLDGNELTLADCNLLPKLHIVQVVCKKNRGFTIPEVFRGVHRYLSNAY
AREEFASTCPDDEEIELAYEQVAKALK

>Mus musculus(Q9D7P7)

MAETTKLQLFVKASEDGESVGHCPSCQRLFMVLLLKGVPFTLTTVDTRRALDVLKDFAPGSQLPILLYDGD
VKTDTLQIEEFLEETLGPPDFPSLAPRYRESNTAGNDIFHKFSAFIKNPVPTQDNALYQQLLRALTRLDSYLR

APLDHELAQEPHLRESHRRFLDGDQFTLADCSLLPKLHIVDTVCAHFRQLPIPAELSCVRRYLDSALQKKEF
KYTCPHSAEILAAYQPAVHPR

>Mus musculus(Q9Z1Q5)

MAEEQPQVELFVKAGSDGAKIGNCPFSQRLFMVLWLKGVTFNVTTVDTKRRTETVQKLCPGGQLPFLLY
GTEVHTDTNKIEEFLEAMLCPPRYPKLAALNPESNTSGLDIFAKFSAYIKNSNPALNDNLEKGLLKALKVLD
NYLTSPLPEEVDETSAEDEGISQRKFLDGNELTLADCNLLPKLHIVQVVCKKYRGFTIPEAFRGVHRYLSNA
YAREEFASTCPDDEEIELAYEQVARALK

>Oncorhynchus mykiss(A0A060WAT0)

MSLSVPQNGIKADNEPVIELFVKAGSDGESIGNCPFSQRLFMILWLKGVVFNVTTVDLKRKPADLQNLAP
GTHPPFITFNGEVKTDVNKIEEFLEDVLSPPKFTKLGTRHPESNTAGMDIFAKFSAFIKNSKPDANEGLERG
LLKTLQKLDEYLRSPLPDEIDHNSIEDIKISTRKFLDGDEMTLADCNLLPKLHIVKVVTKKYRGFDIPKDMTGI
WQYLQNVYTREEFTNTCPSDKEIEIAYQDVAKRLVK

>Oncorhynchus mykiss(A0A060XCT7)

MSLSVPHNGIKADNEPVIELFVKAGSDGESIGNCPFSQRLFMILWLKGVVFNVTTVDLKRKPADLQNLAP
GTHPPFITFNGEVKTDVNKIEEFLEDVLSPPKFTKLSARHPESNTAGMDIFAKFSAFIKNSKPNANEGLERG
LLKTLQKLDEYLRSPLPDEIDHNSIEDVKNSTRKFLDGDNMTLADCNLLPKLHIVKVVTKKYRGFDIPKDMI
GIWQYLQNAYTHEEFTNTCPSDKEIEIAYQDVAKRLIK

>Oncorhynchus mykiss(A0A060XHL8)

MAQRQNSDKDPTIELFIKAGHDGENMGNCPFCQRLFMVLWLKGVKFTVTTVDMRKKPAELKDLAPGT
NPPFLLYNGTLKTDFIKIEEFLEQTLAPPRYPHLSPLSKESFDVGADIFAKFSAFIKNRPANSTFHEKALLREFK
RLDLYLTSPIPEEINQNSRENILVSKRKFLDGNHLTLADCNLLPKLHVIKIAAKKYCDFDIPVQFTGVWRYLN
NAYEREEFRQTCPANIEIEKAYLDVANKRL

>Aptenodytes forsteri(A0A087R950)

AGSDGESIGNCPFSQRLFMILWLKGVIFNVTTVDLKRKPADLQNLAPGTNPPFMTFDGEVKTDVNKIEEF
LEEKLAPPRYPKLAPNHPESNSAGNDVFAKFSAFIKNPRKDANENLEKSLLKALRKLDNYLNSPLPDEIDAY
STEEITVSSRKFLDGDELTLADCNLLPKLHIIKVVAKKYRNFDFPPEMTGISRYLNNAYARDEFTNTCPADQ
EIEYAYLDVAKRMK

>Aptenodytes forsteri(A0A087QW98)

LLQAGSDGESIGNCPFSQRLFMILWLKGVVFSVTTVDLKRKPADLQNLAPGTHPPFITYNGEVKTDVNKIE
EFLEDVLAPPKYLKLSPKHPESNTAGMDIFAKFSAFIKNSRPEANEALERGLLKTLQKLDEYLNSPLPDEIDE
NSMEDITISTRKFLDGNEMTLADCNLLPKLHIVKVVAKKYRDFEIPKEMTGIWRYLTNAYSRDEFTNTCPG
DKEIEIAYSDVAKRLTK

>Balearica regulorum gibbericeps(A0A087V2T6)

QAGIDGESIGNCPFSQRLFMILWLKGVVFNVTTVDLKRKPADLHNLAPGTHPPFLTFNGEVKTDVNKIEE
FLEEILAPPKYPTLAAKHRESNTAGIDIFSKFSAYIKNTKQQDNAALERGLVKALKKLDDYLRTPLPEEIDAN
STEEEKVSKRKFLDGDDLTLADCNLLPKLHVVKIVAKKYRNFEFPTEMTGLWRYLKNAYARDEFTNTCAA
DKEIEQAYADVAKRLSKS

>Balearica regulorum gibbericeps(A0A087VFB6)

AGSDGESIGNCPFSQRLFMILWLKGVIFNVTTVDLKRKPADLQNLAPGTNPPFMTFDGEVKTDVNKIEEF
LEEKLAPPRYPKLAPNHPESNSAGNDVFAKFSAFIKNPRKDANENLEKSLLKALRKLDNYLNSPLPDEIDAY
STEEITVSSRKFLDGDELTLADCNLLPKLHIIKVVAKKYRNFDFPPEMTGISRYLNNAYARDEFTNTCPADQ
EIEYAYLDVAKRMK

>Poecilia formosa(A0A087X883)

MSDANEPKIELFVKAGSDGLCIGNCPFSQRLFMVLWLKGVVFDVTTVDMRKKPDILNDLAPGAQPPFLQ
YGSEVKTDTNKIEEFIESTLCPPKYPRLAARNPESNTAGVDIFSKFSAYIKNSNPQMNDNLEKGLMKALQK
LDNYLGSPLPDEIDQDAAEEATSSSRPFLDGQQLTLADCNLLPKLHILKVVSLKYRNFTIPESLTNVWRYLN
AAYAREEFSATCPVDTEILMAYSTVAKALK

>Poecilia formosa(A0A087XDI2)

MSLSVPQNGVKADNEPVIELFVKAGSDGESIGNCPFSQRLFMILWLKGVVFNVTTVDLKRKPADLQNLA
PGTHPPFITFNGEVKTDVNKIEEFLEDVLSPPKYVKLGAKHPESNTAGMDIFAKFSAYIKNSKPDANEALER
GLLKTLQKLDDYLRSPLPDEIDHNSIEDIKFSSRKFLDGDEMTLADCNLLPKLHIVKVVTKKYRGFDIPKEMT
SIWKYLNNAYTREEFTNTCPSDKEIEIAYGDVAKRLVK

>Fukomys damarensis(A0A091CT16)

MVLWLKGVTFNVTTVDTKRRTETVQKLCPGGQLPFLLYGTEVHTDTNKIEEFLEAMLCPPRYPKLAALNP
ESNTAGLDVFAKFSAYIKNSNPALNDNLEKGLLKALKVLDNYLISPLPEEVDETSAEDEGISRRKFLDGNELT
LADCNLLPKLHIVQVVCKKYRGFTIPEAFQGVHRYLSNAYAREEFASTCPDDEEIELAYEQVAKALK

>Fukomys damarensis(A0A091EKZ0)

MTDTVTANGDDRDPEIELFVKAGIDGESIGNCPFSQRLFMILWLKGVVFNVTTVDLKRKPADLHNLAPG
THPPFLTFNGEVKTDVNKIEEFLEETLTPEKYPKLAAKHRESNTAGIDIFSKFSAYIKNTKQQNNAALERGLT
KTLRKLDDYLSNPLPEEINTDTCGDSNRGSRRKFLDGDELTLADCNLLPKLHVVKIVAKKYRNYDFPAEMT
GLWRYLKNAYARDEFTNTCAADSEIELAYADVAKRLSRS

>Corvus brachyrhynchos(A0A091F2P6)

AGSDGESIGNCPFSQRLFMILWLKGVIFNVTTVDLKRKPADLQNLAPGTNPPFMTFDGEVKTDVNKIEEF
LEEKLAPPRYPKLAPKHPESNSAGNDVFAKFSAFIKNPRKDANENLEKSLLKALKKLDNYLNSPLPDEIDAY
STEEITVSSRKFLDGDELTLADCNLLPKLHIIKVVAKKYRNFHFPPEMTGISRYLKNAYARDEFTNTCPADQ
EIEYAYLDVAKRMK

>Corvus brachyrhynchos(A0A091F5A3)

QAGIDGESIGNCPFSQRLFMILWLKGVVFNVTTVDLKRKPADLHNLAPGTHPPFLTFNGEVKTDVNKIEE
FLEEILAPPKYPKLAAKHRESNTAGIDIFSKFSAYIKNTKQQDNAALERGLVKALKKLDDYLRTPLPEEIDAD
STEEEKVSKRKFLDGDDLTLADCNLLPKLHVVKIVAKKYRNFEFPAEMTGLWRYLKNAYARDEFTNTCAA
DKEIEQAYADVAKRLSKS

>Cuculus canorus(A0A091FUZ4)

QAGIDGESIGNCPFSQRLFMILWLKGVVFNVTTVDLKRKPADLHNLAPGTHPPFLTFNGEVKTDVNKIEE
FLEETLAPPKYPKLAAKHRESNTAGIDIFSKFSAYIKNTKQQDNAALERGLVKALKKLDDYLRTPLPEEIDAD
STEEEKVSKRKFLDGDDLTLADCNLLPKLHVVKIVTKKYRNFEFPTEMTGLWRYLKNAYTRDEFTNTCAAD
KEIEQAYADVAKRLSKS

>Buceros rhinoceros silvestris(A0A091GQ71)

QAGLDGENIGNCPFCQRLFMVLWLKGVKFNVTTVDMTRKPEELKDLAPGTNPPFLLFNKELKTDFIKIEE
FLEQTLGPPMYPHLSPKYKESFDVGSDIFAKFSAYIKNPRKEANINFEKALLREFQRLDDSMEDITVSKRKFL
DGDHLTLADCNLLPKLHIIKIAAKKYRDFEIPADMTGVWRYLNNAYACDEFSHTCPADEEI

>Egretta garzetta(A0A091JA73)

QAGIDGESIGNCPFSQRLFMILWLKGVVFNVTTVDLKRKPADLHNLAPGTHPPFLTFNGEVKTDVNKIEE
FLEEILAPPKYPKLAAKHRESNTAGIDIFSKFSAYIKNTKQQDNAALERGLVKALKKLDDYLRTPLPEEIDAD
STEEEKVSKRKFLDGDDLTLADCNLLPKLHVVKIVAKKYRNFEFPTEMTGLWRYLKNAYARDEFTNTCAA
DKEIEQAYADVAKRLSKS

>Calypte anna(A0A091J0Y9)

QAGIDGESIGNCPFSQRLFMILWLKGVVFNVTTVDLKRKPADLHNLAPGTHPPFLTFNGEVKTDVNKIEE
FLEETLAPPKYPKLAAKHREANTAGIDIFSKFSAYIKNTKQQDNAALERGLVKALKKLDDYLRTPLPEEIDAD
STEEEKVSKRKFLDGDDLTLADCNLLPKLHVVKIVTKKYRNFEFPTEMTGLWRYLRNAYARDEFTNTCAA
DKEIEQAYADVAKRLSKS

Delta Class GSTs

>Sarcoptes scabiei(Q5ISV1)

MASEKPTIYWMPESAPCRSVYLVAKILGIDCEWKVFNLFKKEQMNPEFLAINPLHCVPTLVESDGFTLWE
SRVICSYLIESRDPDSALYPKDLKKRAIIDRCLHFDLGTLYRALADVVYDILFFGKPNLTKLPRLEEVLQLMED
GFAKIDSDYLAGGDGPTLADIVSYFTLQMLDILQELDLTKYSKLYAWRERMEEFVKSNDDGSLAKGLQNF
VGFAQQMQQQHSS

>Drosophila melanogaster(P20432)

MVDFYYLPGSSPCRSVIMTAKAVGVELNKKLLNLQAGEHLKPEFLKINPQHTIPTLVDNGFALWESRAIQ
VYLVEKYGKTDSLYPKCPKKRAVINQRLYFDMGTLYQSFANYYYPQVFAKAPADPEAFKKIEAAFEFLNTF
LEGQDYAAGDSLTVADIALVATVSTFEVAKFEISKYANVNRWYENAKKVTPGWEENWAGCLEFKKYF

>Blattella germanica(A9QUN5)

MTIDFYYLPGSAPCRSVLLAAKAFGVNLNLKVTNLMAGEHLTPEFLKMNPQHTIPTLNDNGFCLWESRAI
LSYLADQYGKDDSLYPKDPKKRALVDQRLYFDLGTLYQRFGDYYYPIMFAKASPDAEKMKKLEEAYQFLD
KFLEGQKFVAGNSLTIADIATIASVSTAAILGFDITRYPNVNKWFENAKKVIPGYDELNHSGCLEFRKMWD
NLTQK

>Agrotis ipsilon(A0A2S0RQT2)

MPIDLYYVPGSAPCRAVLLTAKALNLNLNLKLVDLHHGEHLKPEYLKLNPQHTVPTLVDDGLSIWESRAIIT
YLVNKYGKGSSLYPEEPRARALVDQRLYFDIGTLYQRFADYFYPQVFGGAPADKEKLAKIEDALQLLNTFLE
GQKFCAGPNLTVADLSLIAGVSSFEASDIDFKKYPNIKRWYETVKTTAPGYQEANEKGLDAFKGL

>Aedes aegypti(J9HHL7)

MDFYYLPGSAPCRAVQMTAAAVGVELNLKLTNLMAGEHMKPEFLKLNPQHCIPTLVDNGFSLWESRAII
AYLVEKYGKDDKLYPKDPQKRAVVNQRLYFDQGTLYQRFADYFYPQVFAKQAPVPDNEKKMLDALDFL
NTFLKDSKYVAGDELTIADLSILATVSTFDVAKVDLSKYPNVASWYERLRKEAPGADINEAGCKEFAKYF

>Sarcoptes scabiei(Q2YFE6)

MAKPTFYYMPESPPCRTVMAVARMIGLDMEMKKLNLRNKEHLTPEFLKINPMHKVPTLVEPDGFALGE
SRAISTYIIQKYKPSSPLYPVDDLRRRAHIDGWLQYDCSTLGPALRAVIMDRMYGGGLNENRLNQTKETL
KTLNEVLKALEGRYLLDDQITVADISMYFSCNMIEVLPDLEMSDYEHLCKWYKNMTEAMNAVDHDGLF
AEAIQSAKKYIAEKL

>Danaus plexippus(A0A212EIY8)

MPSQPIKLYYLPPSPPCRAVMMTARALGLDLDLVLTNIMEGQHMTPEFLKMNPQHTIPTMDDSGFILW
ESRAIMTYLANAYGRDDSLYPKNPRSRALVDQRLNFDLGTLFNRFFNLYGRMLFQGEKYDDEAAKKLKE
AIGWMNTMLDGRAFVAGDNMTLADISIIVTFSNLE

>Aedes aegypti(Q17MB7)

MDLYYHIIPPPSRAVLVLAKKLNITLNLISIDTRDANEMAILTEVNPLQSLPTLIDDGQVIGESHTVLIHLTSLF
DKEGMLYPADLKIRSAINELLFFDTNMYKCFVLFAMPTVIKRQDPNHDMLEKLLVCVKALDNYLRARVYA
AGDHFTLADLSLAHTISSLDVIKVKLSDYPNVERWMTKVLPEMPQFEEFQVRAEEALSTFLAKQYGAKCI

>Aedes aegypti(Q17MB8)

MDLYYMPISPPCWSILLLGRQLDLTFNLKEIDFKAEEHKKPEFLKINPAHTVPTLAVGDGYALSESRAILVYL
VESLKTEGQENSLYPRDAKTRGLIHNRLDFDLGTLYQRIIAYCSPQWKSGSMGTEENRTKVQDAFELLEVF
LSKTKYVAADQLTIADISLFVSVSLLDLCYFDRSGYGKVAAWHDVLKKELVGYEDVIAKGFPEWRKH

>Aedes aegypti(Q17MB1)

MELYYSHASAPCRAVQMTALALGVQLNLKEIHLMNGKDHQRPDYGRITPQHSIPTLKDKDLILWESRAI
QMYLVQQYGKDDSLYPKDPSKQAKVNERLFFDACILYHRFTEYYHEQVYGGLEGDDKKLAALEDAVKML
DLFLEGQPYVTGEAMTIVDLSMLATVATMNCLGFELKPYHNVFEWYKHMKDVAPGSKFNETGAKEFAA
FK

>Aedes aegypti(Q16SH6)

MPLDLYCHIVAPFCRSVILLADALEVELNFIEVNVLKKEQFKPEFIAMNPQHCIPTLVDGDVVVWESNAILI
YLAEKYGKVSKRFYPTDIAERAKVNRILFFQLGTLHRALSTYYYPILAGLGEGKPEDFRKIQDAVGVMDKLL
DGNKWLAGEDLSIADFSVVISVASLEGVIKFDLTVYKNVYRWYQQCKKEFRKFEELTQEANDKSQEMIAA
LRQYKLEEINSAKEPCCSAPPGAKTPPKPPCPDSS

>Aedes aegypti(Q17MA8)

MTPVLYFLPASPPCRAVMLLAKMIGVDLEYKTLNVMEGEQLRPEFVELNPQHTIPTLDDHGLVLWESRVI
LSYLVSAYGKDESLYPKDFRSRAMVDQRLHFDLGTLYQRVVDYYFPTIMVGAHLDQTKKAKLAEALGWF
DAMLKQYQWAAANHFTIADVTLCVTVSQIEAFEFDLHPYPKVRAWLAKCKEELEPHGYKDINQTGAEAL
AGLFRAKLKQ

>Plutella xylostella(219AA)

MAIDLYLTPGSAPCRLVLLTAAALNIQLNLNHVDLRAGEQFSPDFLKLNPQHTIPTIVDQGFALWESRAIS
RYLVNKYGHGSTLYPEDPQSRALVDQRLDFDLGTLYPKFADYFYPQVFGSAPADPEKLKKLHEVLGFLDIF
LGDEKYAAGSDLTIADLSLVATVSTIDAAGISLDDFPNIHRWFELVKSTAPDYENANGKGIKAFKELVAQL
NAKTEL

>Plutella xylostella(223AA)

MPAIELYEMQGSAPCRAVRLTARALGKPLTVHHLDLMAGEHLKPEFVKINPQHTIPTIVDDGFALWESRT
IMRYLVNKYGKGSSLYPEEPKARALVDQRLDFDLGTLYAKYAVYFYPQVFGTAPENAEDLKKLNEALAHL
NTFLGESKYAAGSNLTIADFSLVATISTIDVSDIVDLKQYPNIVKWYEHLQSSVEGYEEENLAGLENFRSFIK
EFKAKKAAAK

>Drosophila melanogaster(Q9VG98)

MDFYYMPGGGGCRTVIMVAKALGLELNKKLLNTMEGEQLKPEFVKLNPQHTIPTLVDNGFSIWESRAIA
VYLVEKYGKDDYLLPNDPKKRAVINQRLYFDMGTLYESFAKYYYPLFRTGKPGSDEDLKRIETAFGFLDTFL
EGQEYVAGDQLTVADIAILSTVSTFEVSEFDFSKYSNVSRWYDNAKKVTPGWDENWEGLMAMKALFDA
RKLAAK

>Drosophila melanogaster(Q9VG97)

MVGKALGLEFNKKIINTLKGEQMNPDFIKINPQHSIPTLVDNGFTIWESRAILVYLVEKYGKDDALYPKDIQ
KQAVINQRLYFDMALMYPTLANYYYKAFTTGQFGSEEDYKKVQETFDFLNTFLEGQDYVAGDQYTVADI
AILANVSNFDVVGFDISKYPNVARWYDHVKKITPGWEENWAGALDVKKRIEEKQNAAK

>Drosophila melanogaster(Q9VG96)

MDFYYSPRSSGSRTIIMVAKALGLELNKKQLRITEGEHLKPEFLKLNPQHTIPTLVDNGFAIWESRAIAVYL
VEKYGKDDSLFPNDPQKRALINQRLYFDMGTLHDSFMKYYYPFIRTGQLGNAENYKKVEAAFEFLDIFLE
GQDYVAGSQLTVADIAILSSVSTFEVVEFDISKYPNVARWYANAKKITPGWDENWKGLLQMKTMYEAQ
KASLK

>Drosophila melanogaster(Q9VG95)

MDFYYSPRGSGCRTVIMVAKALGVKLNMKLLNTLEKDQLKPEFVKLNPQHTIPTLVDNGFSIWESRAIAV
YLVEKYGKDDTLFPKDPKKQALVNQRLYFDMGTLYDSFAKYYYPLFHTGKPGSDEDFKKIESSFEYLNIFLE
GQNYVAGDHLTVADIAILSTVSTFEIFDFDLNKYPNVARWYANAKKVTPGWEENWKGAVELKGVFDAR
QAAAKQ

>Drosophila melanogaster(Q9VG94)

MDLYNMSGSPSTRAVMMTAKAVGVEFNSIQVNTFVGEQLEPWFVKINPQHTIPTLVDNLFVIWETRAI
VVYLVEQYGKDDSLYPKDPQKQALINQRLYFDMGTLYDGIAKYFFPLLRTGKPGTQENLEKLNAAFDLLN

NFLDGQDYVAGNQLSVADIVILATVSTTEMVDFDLKKFPNVDRWYKNAQKVTPGWDENLARIQSAKKF
LAENLIEKL

>Drosophila melanogaster(Q9VG92)

MDFYYHPCSAPCRSVIMTAKALGVDLNMKLLKVMDGEQLKPEFVKLNPQHCIPTLVDDGFSIWESRAILI
YLVEKYGADDSLYPSDPQKKAVVNQRLYFDMGTLFQSFVEAIYPQIRNNHPADPEAMQKVDSAFGHLD
TFLEDQEYVAGDCLTIADIALLASVSTFEVVDFDIAQYPNVARWYENAKEVTPGWEENWDGVQLIKKLV
QERNE

>Drosophila melanogaster(Q9VGA0)

MLDFYYMLYSAPCRSILMTARALGLELNKKQVDLDAGEHLKPEFVKINPQHTIPTLVDDGFAIWESRAILI
YLAEKYDKDGSLYPKDPQQRAVINQRLFFDLSTLYQSYVYYYYPQLFEDVKKPADPDNLKKIDDAFAMFN
TLLKGQQYAALNKLTLADFALLATVSTFEISEYDFGKYPEVVRWYDNAKKVIPGWEENWEGCEYYKKLYL
GAILNKQ

>Anopheles gambiae(Q8MUS3)

MDYYCNFVSPPSQSVILVAKKLGIKLNLRKINIYDPVAMDTLSKLNPHHILPMLVDNGTVVFEPCAIVLYLV
EMYAKNDALYPKDALVRCVVNQRLFFDVSTLFKQIYENVHVQMRNSQPSEKQVQRLQKAVDVLESFLYE
RSYTAADQLTVADICLLVTVNALTLWLGYELAPYPRIRDWLGRVVAEIPGCAEFQREVEDATRAYVVNRKI

>Anopheles gambiae(Q8MUS4)

MELYSDIVSPSCQNVLLVAKKLGIALNIKKTNIMDATDVAELTKVNPQHLIPTFVEDDGHVIWESYAIAIYL
VEKYGQDDALYPKDPKVRSIVNQRLFFDIGTLYKNILANVDVLIEKQQPSAELRGKLEQALDLTEKFVTECR
FVAADHLTLADIFMLGSITALEWFRYDLERYPGIRGWVERVTAQFPDYSNFHKEIREATKQYVATHCPHL
EY

>Anopheles gambiae(O76483)

MTPVLYYLPPSPPCRSVLLLAKMIGVELELKALNVMEGEQLKPDFVELNPQHCIPTLDDHGLVLWESRVIL
AYLVSAYGKDENLYPKDFRSRAIVDQRLHFDLGTLYQRVVDYYFPTIQLGAHLDQTKKAKLAEALGWFEA
MLKQYQWSAANHFTIADIALCVTVSQIEAFQFDLHPYPRVRAWLQKCKDELQGHGYKEINETGAETLAG
LFRSKLKQ

>Anopheles gambiae(9GPL7)

MDFYYHPASPYCRSVMLVAKALKLSLNLQFVDLMKDEQLRPTFTVLNPFHCVPTLVDNDLTMWESRAIL
VYLVDKYGRTNSRLYPKDAKTRAIINQRLFFDHGTLGTRLEDYYYPLYFEGATPGGEKLEKLEEALAVLNGY
LINNPYAAGPNITLADYSLVSTVTSLEVVQHDLSKYPAISAWYEGCKATMADFQEINESGMQQYRLTSSL
VPHLQLLHMPFAE

>Anopheles gambiae(Q86D84)

MDLYYNILSPPSRAILLLGEALQLKFNLISLDVHRKDYVNPAFKKINPQHTVPTLVVDGVAICEPGAILIYLAE
QYAPAGTTYYPPDPLRRAIVNQRLLFECGTLYKCIFVYYSPVVLERATPVETDRQKLIEAVAVLDGILQHSA
FVAGDCLTVADYSLVCTVSMLVVLKFELAPYAAVRRWYERCKEVIAGYTDLTQRAVTMFQKWMEQENS
KG

>Bombyx mori(Q60GK5)

MTIDLYYVPGSAPCRAVLLTAKALNLNLNLKLVDLHHGEQLKPEYLKLNPQHTVPTLVDDGLSIWESRAIIT
YLVNKYAKGSSLYPEDPKARALVDQRLYFDIGTLYQRFSDYFYPQVFAGAPADKAKNEKVQEALQLLDKFL
EGQKYVAGPNLTVADLSLIASVSSLEASDIDFKKYANVKRWYETVKSTAPGYQEANEKGLEAFKGLVNSM
LKK

>Bombyx mori(Q2I0J5)

MAIDLYFTAGSAPCRVVLLVAAALDLQLNLKPLNLWEREQLQADFLKLNPQHTVPTIVDEGFPLWESRAI
SRYLVNKYGGDSSSLYPKDLMARALVDQRLDFDIGTLYPRFAQYFYPQVFGGAKPDAAALKKLEEALVFL
NAFLEGQKYVTGDVLTIADLSLVATISTIDAAEISLKSYPNVEKWFELMKTTAPDYQNANQKGIDEFKKLIA
QMKAKTEL

>Bombyx mori(H9JAJ3)

MLTASVLGVELELIAVNILDNEHKTPEYLKMNPQHTIPTMDDNGFILWESRAIQAYLVNAYGKNDALYPK
NPRLRAIIDQRLNFDLGTLSRRWIDLYVPMLIKGEPFDDEKGEKLNEALELLNIFLEGHAFVAGENMSIADL
SIVVTISNLDAVEYDLSSYDNVRKWFERMKIALKPYDYEDIDQTGAEILASFINKDDE

Epsilon Class GSTs

>Aedes aegypti(Q170C6)

MSPITLYTTRRTPAGRAVEITAKLIGLELDVKFIDLSKKEHLTEEFLKLNPQHTVPTIVDNGVALYDSHAIIVY
LVQKYGKDDALYPKDLVTQARVNALLHFESGILFARLRGTLEPIFYHGFPEVPQEKLTAIHGAYDLLEATLK
SDYLVGDSLTLADVSCSTSLSTLNALFPIDAEKCPKLVAYLQRLEANMPNYKELNSDRAAEAVAFFKVKLEE
NKKK

>Danaus plexippus(A0A212F3G0)

MMVAEILGVQYSNCEINPVLREQDTPEMTAKNPLRTIPYIEDDGFCLGDSHAIILYLFDKYAKPEHDHLYP
SNVKIRAKINQILFFDCGVLFARLRSVMAPTYMGRLSELSQSMKRNIEDAYRIIEAYLSNTLYIADNNVTLA
DYSVLATMSSLHGLHPIDSNKYPKLLNWYKRMSSLPVCKTINDPGAELHVTGLKKLMEQRKLSKL

>Danaus plexippus(A0A212FES8)

MSKGRLNEVTKLPRLLLYKRNASPPSSAVMILGDMLGLNFDYREPDLIKLEHRSPEFKKINPMATILVLQD
GDVTICESHAIMKYLVNKYGGERCERLYPADLSVRANIDQLMFYDAGVLFVRLKVVALPTMLQGLTGPTK
EQVADIDEGYTVLEAYLNKHSYIATDHLTIADLSVGTTTTALQSVHKLDKNRFPLSAEWLERLKGEKSFKKF
NEPSVKELSTILNVFWKKNKERIR

>Aedes aegypti(Q5PY78)

MGKVQLYTAKLSPPGRAVELTAKAIGLDLDVHPINLIAGDHLKPEFVKMNPQHTIPLIVDEDGTIVYDSHA
IIIYLVSKYAKDDSLYPKDIATRAKINAALHFDSGVLFARLRFYLEPILYYGSPDTPQDKIDYACKAYQLLNDTL
VDEYIVGNRMTLADLSCIASIASYHAIFPIDAAKYPKLAAWVQRLEKLPYYKGTNQEGAEELAAVYRDRLA
QNRAGKK

>Aedes aegypti(Q170C9)

MTKPIVYTLYLSPPSRAVDLCAVALGIELERKVMNLLEREHLDPKFLKMNPQHTIPVLDDGGIIVRDSHAI
MIYLVSKYGKDDSLYPKDLAEQAKVNAALYFDCGVLFARLRFITEQILMGGSEIPAEKAAYVESAYQLLED
ALTDDFIAGNSLTIADLSCGSTVSTAMGLIPMDRDKYPKIYAWLNRLKALPYFEELNDQGAVELPAIMKNL
METNARKA

>Aedes aegypti(Q170C7)

MATNSPRIKLYTNPISPPGRSVELTAKAIDLDIEVIAIDLLGNDTLKPDYLLKNPQHTIPMIDDGGKFIWDS
QAINVYLTTVYSRNEDLYPNDPFVRAKVNAGLHFNSGVLFSRLKLLISPVIRGFKQDLDPEKVEYFNIGLQL
LEDTLHADYYIGNRMTLADLSCVSSVSSFDAVLPLDNSRFPKTVDWLRRMEQLPYYGEANGEGAKKLAK
VVQSFLK

>Aedes aegypti(Q170C8)

MSSKIVLHTTRRTPGGRAVQILSHILGLDLDLKFVDLSKKEQMSEEFLKLNPFHTIPTIDDDGVPVYDSHAIL
VYLVSKYAKDRDLFPEDPVIQARINAWFHFDSGVLFPRLRGAVEPVFYFGLKKIPRDRMEAIEAAYDLFEG
ALKGDFLVGDSLTLADISVTTCLVSLNGVCPMEESKYPKSCAFLKRMEQSMPCYKEFNAEVLEETKVFLKQ
KLDENNKK

>Plutella xylostella(220AA)

MGLTVYKIDWSPPARAVIMTLEALNITDAELVDVSLLDGKHMSEEYLKMNPQHTVPVIKDGDFVLWDS
HAICAYLVDKYGKDDSLYPKDLQKRAVVDQRLHFDTGILFPSVRGAAEPVLFDWEPTFNPEKLKVIQSGY
DFLEKFLDHSYLAGDHLTIADICAGATVSSMNVIVPIAANRYPKISAWLDRLNSIEYFSRINGNGIKIITALFE
SKLNKSKK

>Plutella xylostella(216AA)

MVLTLYKLDASPPARAVMMTLEALGIRDVEMVDVNLFEGAQFTEEYVKMNPQHTIPALKDDDFAIWDS
HAICPYLVSKYGQDDSLYPQDPQRRAVIDQRLHFDSGILFPSLRATVAPVLFLGERSFKPEGLQAIKAGYDF
LEKFLDKPYCAGDQLSIADICTSATVSSMSAALPIDADTYPNITAWLDRLSKEEFYTKVNLPGLQQFSGALK
SKLL

>Drosophila melanogaster(A1ZB68)

MGKLTLYGIDGSPPVRSVLLTLRALNLDFDYKIVNLMEKEHLKPEFLKINPLHTVPALDDNGFYLADSHAIN
SYLVSKYGRNDSLYPKDLKKRAIVDQRLHYDSSVVTSTGRAITFPLFWENKTEIPQARIDALEGVYKSLNLFL
ENGNYLAGDNLTIADFHVIAGLTGFFVFLPVDATKYPELAAWIKRIKELPYYEEANGSRAAQIIEFIKSKKFTI
V

>Drosophila melanogaster(A1ZB69)

MGKISLYGLDASPPTRACLLTLKALDLPFEFVFVNLFEKENFSEDFSKKNPQHTVPLLQDDDACIWDSHAI
MAYLVEKYAPSDELYPKDLLQRAKVDQLMHFESGVIFESALRRLTRPVLFFGEPTLPRNQVDHILQVYDFV
ETFLDDHDFVAGDQLTIADFSIVSTITSIGVFLELDPAKYPKIAAWLERLKELPYYEEANGKGAAQFVELLRS
KNFTIVS

>Drosophila melanogaster(A1ZB70)

MVKLTLYGVNPSPPVRAVKLTLAALQLPYEFVNVNISGQEQLSEEYLKKNPEHTVPTLEDDGNYIWDSHA
IIAYLVSKYADSDALYPRDLLQRAVVDQRLHFETGVVFANGIKAITKPLFFNGLNRIPKERYDAIVEIYDFVE

TFLAGHDYIAGDQLTIADFSLISSITSLVAFVEIDRLKYPRIIEWVRRLEKLPYYEEANAKGARELETILKSTNF
TFAT

>Drosophila melanogaster(A1ZB71)

MVKLTLYGLDPSPPVRAVKLTLAALNLTYEYVNVDIVARAQLSPEYLEKNPQHTVPTLEDDGHYIWDSHAI
IAYLVSKYADSDALYPKDPLKRAVVDQRLHFESGVVFANGIRSISKSVLFQGQTKVPKERYDAIIEIYDFVET
FLKGQDYIAGNQLTIADFSLVSSVASLEAFVALDTTKYPRIGAWIKKLEQLPYYEEANGKGVRQLVAIFKKT
NFTFEA

>Anopheles gambiae(Q8WQJ9)

MAPIVLYSTRRTPAGRAVELTAKMIGIELDVQYIDLAKKENMTEEYLKMNPMHTVPTVNDNGVPLYDSH
AIINYLVQKYAKDDTLYPAKDLVKQANINALLHFESGVLFARLRWILEPVFYWGQTEVPQEKIDSVHKAYD
LLEATLKTSGTDYLVGGTITLADISVSTSLCTLNALFPADASKYPLVLAYLKRLEQTMPHYQEINTDRANDAL
QLYNQKLGKV

>Anopheles gambiae(Q8WQJ8)

MPNIKLYTAKLSPPGRSVELTAKALGLELDIVPINLLAQEHLTEAFRKLNPQHTIPLIDDNGTIVWDSHAIN
VYLVSKYGKPEGDSLYPSDVVQRAKVNAALHFDSGVLFARFRFYLEPILYYGATETPQEKIDNLYRAYELLN
DTLVDEYIVGNEMTLADLSCIASIASMHAIFPIDAGKYPRLAGWVKRLAKLPYYEATNRAGAEELAQLYRA
KLEQNRTNAK

>Anopheles gambiae(Q8WQJ7)

MATNPIIKLYTAKLSPPGRAVELTAKLLGLSLDIVPINLLAGDHRTDEFLRLNPQHTIPVIDDGGVIVRDSHA
IIIYLVQKYGKDGQTLYPEDPIARAKVNAGLHFDSGVLFSRLRFYFEPILYEGSAEVPQDKIDYMKKGYELLN
DALVEDYIAGSSLTLADVSCIATIATMEEFFPMDRSRYPALVAWIERLSRTLPEYDQLNQEGAVEFAEICES
LRLKNGASVAAK

>Anopheles gambiae(Q8MUQ7)

MEPSRLVLYTNRKSPPCRAVKLTARALGIELVEKEMTLLRGDKLMEEFLKVNPQQTIPVLDDGGIVITASH
AITIYLVCKYGRDDGLYPSELVRRARVHTALHLEAGVIFSRLSFLFEPVIYSGKSYFHSDRIEHIRKAYRLLEDS
LVDQYMVGESLTIADFSCISSIATLVGVVPLDESKFPKSTAWMRRMQELPYYEEANGTGALELAEFVLGK
KEANASQFL

>Anopheles gambiae(Q8WQJ5)

MILYYDEVSPPVRGVLLAIAALGVKDRIKLEYIDLFKGGHLSSDYLKINPLHTVPVLRHGELTLTDSHAILVYL
CDTFAPPGHTLALPDALTRAKVFNMLCFNNGCLFQRDAEVMRKIFSGAITDPTQHLKPIEAAIDALEQFL
QRSRYTAHDQLSVADFAIVATLSTVAIFVPLPADRWPRVCEWFAVMEALPYYNDQNRVGLDMLRKHLA
GKIKL

>Bombyx mori(B0LB15)

MSLMLYKLNASPPARTAMMVCELFKVPVKMVDVNLSKGEHFSPEYLKRNPLHTVPTLEDGDLIITDSHAI
AMYLADKYGKDDSLYPKDLKSRAIVNQRLFFDSTVLFSRMRSVTFPVIIEGCKTVTEKQINDIIEAYGYVETY
LSNTKFIATNNLTIADISAYAVVSSLLFIVPLDGAKFPKTQTWLNEMEKKPFAQKYNVNGVAELGALLKEKL
GS

>Bombyx mori(B0LB16)

MVFILYKKDTSPPCRSVQMVLHELGIYDVELIEVNLPERDHLKEEFLRMNPQHTVPTLIDGDFIIWDSHAI
VTYLVNRYAKNDTLYPKEPKQRAIVDQRLHFDTGVLFAILRATAEPVLYNNEKSFKQENLEKMEAAYEFVE
KFLTSDWLAGDQVTLADICCVSTISSMNVIVPIDKKKYPKIISWLQRCSEQEFYKKANEPGLKKFIEMFKNK
IGN

>Bombyx mori(B0LKP4)

MVLTLYKLDASPPVRSVYMVIEALKIRDVEYVDVNLLEGSHLKEEFLKMNPQHTIPLLKDDDFLIWDSHAI
SGYLISVYGADDSLYPNEPKKRALIDQRLHFDSGILFPALRGVAVIIFFNLLCLGQDELIIFRGEKEIRPENLAK
IKSAYDFTEKILSSDWIAGDEFSLADICCVTSISTLNEMVPIDGSLYPKLASWLDRSSQLPIYKKANEPGLLQF
REIFKNKTS

>Bombyx mori(B0LKP5)

MTPILYKTDASPPARAVMMIVDILGLKVDEQELNPILRQQDTPEFKKKNPMRTIPILEEGDFYLADSHAIM
LYLIDKYGKPEHAHLYPSEKRKRATINQRLFFDCGVLFPRLRAVMAPTYAGKLAELNRNMIKNIEDAYSIM
ESYLTENLYLADEVVTVADISAITTISSLNGLYPVDEKSKWINRMNDKEYCRKINTPGSELHVAGLIALMDN
TKHNQQSKL

Kappa Class GSTs

>Macaca fascicularis(A0A023JCQ9)

MGPLLRTVELFYDVLSPYSWLGFEVLCRYQNIWNINLQLRPSLIGGIMKDSGNKPPGLLPRKGQYMANDI
KLLRHHFQIPIQFPKDFFSVIIEKGSLSAMRFLTAVSLEHPEMLEKVSRELWMRVWSRDEDITQPQSILAA
AEKAGMSAEQAQGLLEKISTPKVKNQLKETTEAACRYGAFGLPITVAHVDGQTHMIFGSDRMELLAFLL
GEKWMGPVPPAVNARL

>Rattus norvegicus(P24473)

MGPAPRVLELFYDVLSPYSWLGFEVLCRYQHLWNIKLKLRPALLAGIMKDSGNQPPAMVPHKGQYILKEI
PLLKQLFQVPMSVPKDFFGEHVKKGTVNAMRFLTAVSMEQPEMLEKVSRELWMRIWSRDEDITESQNI
LSAAEKAGMATAQAQHLLNKISTELVKSKLRETTGAACKYGAFGLPTTVAHVDGKTYMLFGSDRMELLA
YLLGEKWMGPVPPTLNARL

>Homo sapiens(Q9Y2Q3)

GPLPRTVELFYDVLSPYSWLGFEILCRYQNIWNINLQLRPSLITGIMKDSGNKPPGLLPRKGLYMANDLKL
LRHHLQIPIHFPKDFLSVMLEKGSLSAMRFLTAVNLEHPEMLEKASRELWMRVWSRNEDITEPQSILAAA
EKAGMSAEQAQGLLEKIATPKVKNQLKETTEAACRYGAFGLPITVAHVDGQTHMLFGSDRMELLAHLLG
EKWMGPIPPAV

>Mus musculus(Q9DCM2)

GPAPRILELFYDVLSPYSWLGFEVLCRYQHLWNIKLQLRPTLIAGIMKDSGNQPPAMVPRKGQYIFKEIPL
LKQFFQVPLNIPKDFFGETVKKGSINAMRFLTTVSMEQPEMLEKVSREIWMRVWSRDEDITEYQSILAAA
VKAGMSTAQAQHFLEKISTQQVKNKLIENTDAACKYGAFGLPTTVAHVDGKTYMLFGSDRLELLAYLLGE
KWMGPVPPTA

>Xenopus tropicalis(Q5M8U8)

MSNRRVLECFYDVVSPYSWLGFEVVCRYKNIWNVDALLRPGFLGGIMQASGNSPPAMVPKKGVYMAQ
DIARLSDFFQVPLRQPSNFFQSVIKKGSLQAMRFVTAVEMQHPEFLEPVSRELWRRIWSEDKDITEPESIL
EAAKKAGMSADQAKKLIESTALPEVKNKLKQNTDEALKYGAFGMPIIVAHVDGKPHMYFGSDRFELLAH
QLGEKWMGPVPQKPRL

Lambda Class GSTs

>Sorghum bicolor(A0A1B6P694)

GTTRLYISYICPYVQRVWIARNFKGLQEKIQLVAIDLQDKPAWFLEKVYPPGKVPVLEHNGNIIAESLDLLS
YLDANFEGPKLFPGDQDPAKQAFADELIANSDSIIIALFRAGRAYAEGQGDDDISKLLAPALDKVEESLGRF
SDGPFLLGQSMSAVDMVYAPFIERFKDFFAAVKHYDMTQERPKLKEWIEELNKIDAYTATWGDRRLQLA
ALMNKFGIQSPV

>Glycine hispida(I1L8Q0)

GTTRLYISYSCPYAQRVWIARNFKGLKDKINLVPINLQDRPAWYKEKVYPENKVPSLEHNGKVLGESLDLI
KYVDENFEGTPLFPRDPAKKEFGEQLISHVDTFSRDLFVSLKGDAVQQASPAFEYLENALGKFDDGPFLLG
QFSLVDIAYIPFAERFQIVFAEVFKHDITEGRPKLATWFEELNKLNAYTETRVDPQEIVDLFKKRF

>Pinus tabuliformis(L7S317)

GTTRLYISVACPYAQRVWSARNIKGLSQIQLVPIDLQDRPAWYKEKVYPPNKVPAIEHNGKVTGESLDLLE
YLENNFEGPKLFPTDPAKKEAANELLKYTDTFTKNSFIALTKPDSETAQEAGPALDYLENALGKFSDGPFLL
GQFSVVDIAYGPFVERFHVAFPALKNYDITAGRPKLSKWIQELHKIEGYAKTVSDPESIVETYKRI

>Arabidopsis thaliana(Q6NLB0)

TTRLYISYTCPFAQRVWITRNLKGLQDEIKLVPIDLPNRPAWLKEKVNPANKVPALEHNGKITGESLDLIKY
VDSNFDGPSLYPEDSAKREFGEELLKYVDETFVKTVFGSFKGDPVKETASAFDHVENALKKFDDGPFFLGE
LSLVDIAYIPFIERFQVFLDEVFKYEIIIGRPNLAAWIEQMNKMVAYTQTKTDSEYVVNY>

>Arabidopsis thaliana(Q9M2W2)

MSVGLKVSAFLHPTLALSSRDVSLSSSSSSLYLDRKILRPGSGRRWCKSRRTEPILAVVESSRVPELDSSSEP
VQVFDGSTRLYISYTCPFAQRAWIARNYKGLQNKIELVPIDLKNRPAWYKEKVYSANKVPALEHNNRVLG
ESLDLIKYIDTNFEGPSLTPDGLEKQVVADELLSYTDSFSKAVRSTLNGTDTNAADVAFDYIEQALSKFNEG
PFFLGQFSLVDVAYAPFIERFRLILSDVMNVDITSGRPNLALWIQEMNKIEAYTETRQDPQELVERYKRRV
QAEARL

>Arabidopsis thaliana(Q9LZ06)

MAPSFIFVEDRPAPLDATSDPPSLFDGTTRLYTSYVCPFAQRVWITRNFKGLQEKIKLVPLDLGNRPAWYK
EKVYPENKVPALEHNGKIIGESLDLIKYLDNTFEGPSLYPEDHAKREFGDELLKYTDTFVKTMYVSLKGDPS
KETAPVLDYLENALYKFDDGPFFLGQLSLVDIAYIPFIERFQTVLNELFKCDITAERPKLSAWIEEINKSDGYA
QTKMDPKEIVEVFKKKFM

>Glycine max(I1LYZ1)

MLMATLGLQAVRPPPLTSTSDPPPLFDGTTRLYISYSCPYAQRVWITRNYKGLQDKIKLVPIDLQDRPAW
YKEKVYPENKVPSLEHNGKVLGESLDLIKYVDVNFEGTPLVPSDPAKKEFGEHLISHVDTFNKDLNSSLKG
DPVQQASPSFEYLENALGKFDDGPFLLGQFSLVDIAYIPFIERYQIVFAELFKQDIAEGRPKLAAWIEEVNKI
DAYTQTKNDPQEIADKYKKRLLSQQ

>Pinus tabuliformis(L7S6H2)

MAALYAVPPPLTSKSEPPPLFDGTIRLYINVLCPYAQRAWSARNIKGLSEIQIVSIDLQDRPAWYKEKVYPP
NKVPALEHNGKVTGESIALLEYLENNFEGPKLFPTDPAKTEAATELLKHTDTFTQNLFGALTKPEPKAAQE
AGPALDYLENALGKFADGPFFLGQLSVVDIAYGPFFERFQVVFPALKNSAGRPKLLKWIQELHKIEGYAKT
VADPEIIVEIFKTRQANSAK

>Glycine max(I1NA39)

MATAGVQEVRVPPLTSTSEPPSLFDGTTRLYISYICPYAQRVWITRNYKGLQDKIKLVPIDLQNRPAWYKE
KVYPENKVPSLEHNGKVLGESLDLVKYIDDNFEGPSLVPSDPAKKEFGEELISHVDTFTKELYSALKGDPIH
QAGPAFDYLENALGKFGDGPFFLGQFSWVDIAYVPFVERFQLVFADVFKHDITEGRPKLATWIEEVNKIS
AYTQTRADPKEIVDLFKKRFLAQQ

>Pinus tabuliformis(L7S0Z7)

MAALYSIPPALTSKSEPPPLFDGTTRLYICVKCPYAQRAWSARNIKGLSQIQIVPIDLQDRPVWYKEKVYPP
NKVPALEHNGKVTGESMDLLEYLDNNFEGPKLFPTDPAEKEAANELLKHTDAFTKTVFVALTKPDSEAAQ
EAGPALDYLENALGKFSDGPFLLGQFSVADIAYGPFVERFQVAFPALKNYDITAGRPKLLKWIQELHKIEG
YAKTTVADPEIILETYKRILAIFTKRL

>Capsella rubella(A0A140EH50)

MSVGVNVSVCSYSSLALSTKDFSSISSSSLSLDRNILRPVSGRLRCKSSGKRRIEPILAVKESSRVPELDSSSEP
VQVFDGSTRLYISYSCPFAQRAWIARNYKGLQNKIELVPIDLKNRPAWYKEKVYAANKVPALEHNNRVIG
ESLDLIKYIDTNFEGPSLTPDGLEKQAAADELLSYTDSFSKAVRSTLNGTDSNAADATFDYIEQALSKFNEG
PFFLGQFTLVDVAYAPFIERFQLILRDVTNVDITSGRPNLTLWIQEMYKIEAFTETRQDPKELVERYKKRVQ
AEARL

>Populus trichocarpa(D2WL72)

MELPRLYTCYTCPFAHRVWITRNFKGLQDEIKLVPLILQNRPAWYSEKVYPPNKVPSLEHNGKITGESLDLI
KYLESNFQGPSLLPEDPAKKEFAEELFSYTDTFNRTVFTSFKGDPAKEAGPAFDHLENALHKFGDGPFFLG
QEFSLVDIAYIPFVERFCIFLSEVFKYDITAGRPKLAAWIEELNKIEAYKQTKTDPKEMVEVYKKRFMA

>Physcomitrella patens subsp. patens(K9Y419)

MATLVTSYLHACNATFATTLPRHSRLIQAPSVQFSQVCGKNLGQSFSSPSARILRRNYDFRRELVSVTRSM
ASSFENREVLDSKSASPAIFDGTTRLYFSSRCPYAQRVWVAVKYKGLDEIECVEISLSDKPTWYKEKVYPVG
KVPALEHNGTVTGESMDLLTYLDDHFGGPKLAPTEESKKQAAAELLQYADTFNKLGFTGLSMKSSTPDEI
AAAVAPAFDFLENALAKFSSEGPLFLGNFGLVDIVYAPFIERFEIAFGGIRNYDIRAGRPRLAKWIEAMDN
VEAYSSTKVPRATLLELYKKMLENDYFIRVGVAANQNNSSGSSVAVN

>Larix kaempferi(V5L6L6)

MAALDLIPPVLNSKSEPPPLFDGTTRLYISVTCPYAQRVWSARNIKGLSEIQIVPIDLQDRPAWYKEKVYPP
NKVPSLEHNGKIIGESLDLLEYLENNFEGPKLFPTDPAKKEAANELLKYTDTFTKSLFIGLTKPEPEVAQEAG
PALDYLENALGKFADGPFFLGEFSVVDIAYGPFVERFQIVYPALKNNDITSDRPKLLKWIQELHKIDAYAKT
KVDPETIVETYKRILAKKGDPNH

>Larix kaempferi(V5L7R3)

MAAIDLIPPVLNSKSEPPPLFDGTTRLYISVICPYAQRVWSARNIKGLSQIQIVPIDLQDRPAWYKEKVYPP
NKVPSLEHNGKVIGESLDLLEYLENNFEGPKLFPTDSEKIEAANVLLKYTDTFTKDLFIGLTKPEPEAAQEAG
PALDYLENALGKFADGPFLLGEFSLVDIAYGPFVERFQIVYPVLKNYDITADRPKLLKWIQELHKIDAYAKT
VTDSETTVEIYKRLLANYAK

>Populus trichocarpa(D2X9U2)

MGSLRLYTCYTCPFAQRVWITRNFKGLQDEIKLVPLILQNRPAWYPEKVYPPNKVPSLEHNGKITGESLDL
IKYLESNFEGPSLLPQDPAKKEFAEELFSYTDKFNGTVYTAFKGDLAKSGPAFDYLENALHKFDDGPFFLGK
CCQVDIAYIPFVERLNIFLLEVFKYDIAAGRQKLAAWIEEVNKIEAYKQTKTDPKELVEFYKKRFVSHS

>Populus trichocarpa(D2WL71)

MEQQRLYISYTCPYAQRVWITRNCKGLQDKIKLVPIDLQDRPAWYKEKVYPPNKVPSLEHNNEVKGESL
DLIKYIDSHFDGPSLFPDDPAKKEFAEDLFSYTGSFSKANNSTFKGEADEAGAAFDYIETALSKFDDGPFFL
GQFSLVDIAYAPFIERFQPALLEFKKYDITAGRPKLAAWIEEMNKIEAYNQTRREPKQHVGTYKKRFEAHL

MAPEG GSTs

>Homo sapiens(O14880)

MAVLSKEYGFVLLTGAASFIMVAHLAINVSKARKKYKVEYPIMYSTDPENGHIFNCIQRAHQNTLEVYPPF
LFFLAVGGVYHPRIASGLGLAWIVGRVLYAYGYYTGEPSKRSRGALGSIALLGLVGTTVCSAFQHLGWVKS
GLGSGPKCCH

>Homo sapiens(Q99735)

MAGNSILLAAVSILSACQQSYFALQVGKARLKYKVTPPAVTGSPEFERVFRAQQNCVEFYPIFIITLWMAG
WYFNQVFATCLGLVYIYGRHLYFWGYSEAAKKRITGFRLSLGILALLTLLGALGIANSFLDEYLDLNIAKKLR
RQF

>Mus musculus(Q9CPU4)

MAVLSKEYGFVLLTGAASFVMVLHLAINVGKARKKYKVEYPVMYSTDPENGHMFNCIQRAHQNTLEVY
PPFLFFLTVGGVYHPRIASGLGLAWIIGRVLYAYGYYTGDPSKRYRGAVGSLALFALMGTTVCSAFQHLG
WIRPGLGYGSRSCHH

>Bos taurus(Q3T100)

MAVLSKEYGFVILTGAASFLMVTHLAINVSKARKKYKVEYPTMYSTDPENGHIFNCIQRAHQNTLEVYPP
FLFFLAVGGVYHPRIVSGLGLAWIVGRVLYAYGYYTGEPRKRQRGALSFIALIGLMGTTVCSAFQHLGWV
RTGLNSGCKSCH

>Bos taurus(Q64L89)

MANLSQLMENEVFMAFASYTTIVLSKMNFMSTATAFYRLTKKVFANPEDCAGFGKGENAKKYLRTDDR
VERVRRAHLNDLENIVPFLGIGLLYSLSGPDLSTAILHFRLFVRARIYHTIAYLTPLPQPNRALAFFIGYGVTLS
MAYRLLKSKLYL

>Bos taurus(Q2KJG4)

MAGNSILLAALSVLSACQQSYFAMQVGKARSKYKVTPPSVSGSPDFERIFRAQQNCVEFYPIFIITLWMA
GWYFNQVFATCLGLVYIYSRHQYFWGYAEAAKKRVTGFRLSLGVLALLTVLGAVGILNSFLDEYLDIDIAK
KLRHF

>Rattus norvegicus(A0A0G2JU12)

GYFALQVGRVRLKYKIAPPAVTGSLEFERIFRAQQNSLEFYSVFIISLWMAGWYFNQVFATCLGLLYIYAR
HKYFWGYAEAAEKRIIGFRLSLGILALLTVLAVLGVASRFLDEYLDFHVAKKLKRPF

>Drosophila melanogaster(Q8SY19)

MASPVELLSLSNPVFKSFTFWVGVLVIKMLLMSLLTAIQRFKTKTFANPEDLMSPKLKVKFDDPNVERVR
RAHRNDLENILPFFAIGLLYVLTDPAAFLAINLFRAVGIARIVHTLVYAVVVVPQPSRALAFFVALGATVYM
ALQVIASAAF

>Macaca mulatta(F6RJ20)

MAGNSILLAAVSILSACQQSYFALQVGKARLKYKVTPPAVTGSPEFERVFRAQQNCVEFYPIFIITLWMAG
WYFNQVFATCLGLMYIYGRHLYFWGYSEAAKKRITGFRLSLGILALLTLLGALGIANSFLDEYLDLNIAKKLR
RQF

>Danio rerio(Q5XJJ0)

MAEVVHMIDSEVFLAFSTYATIVILKMMLMSLMTSYLRLTKQVFSNLEDTAMAIAEDKKKLVRTDPDVER
VRRCHLNDLESIVPFVVIGLLYALTGPVLSTALLHFRVFVVSRFIHTVAYIMALPQPTRGVAFGVGLLTTLS
MAYRVLTTALFL

>Danio rerio(B0R1F0)

MAEVVHMIDSEVFLAFSTYATIVILKMMLMSLMTSYLRLTKQVFSNLEDTAMAIAEDKKKLVRTDPDVER
VRRCHLNDLESIVPFVVIGLLYALTGPVLSTALLHFRVFV

>Pongo abelii(A0A2J8XP37)

MAGNSILLAAVSILSACQQSYFALQVGKARLKYKVTPPAVTGSPEFERVFRAQQNCVEFYPIFIITLWMAG
WYFNQVFATCLGLVYIYGRHLYFWGYSEAAKKRITGFRLSLGILALLTLLGALGIANSFLDEYLDLNIAKKLR
RQF

>Canis lupus familiaris(F1PFR5)

IFSCRQNCVEFYPIFLVTLWMAGWYFNQVFATCLGLVYIYARHQYFWGYSEAAKKRITGFRLSLGCLALLT
VLGALGIANSFLDEYLDLNVIKKLR

>Equus caballus(F7AUV5)

MAGNSILLAAVSLLSACQQSYFAFQVGRARLKYKIMPPAVSGSPEFDRIFRAQQNSVEFYPAFMITLWM
AGWYFNQVFATCLGLLYIYARHQYFWGYSEAANKRMTGFRLGLGILALLAILGALGIANSFLDEYLDLNV
AKKLRHF

>Canis lupus familiari(E2RHK1)

MVDLTELMENEVFMAFASYTTIILSKMMFMSTATAFFRLTRKVFANPEDCASFGKGENAKKYLRTDDRV
ERVRRAHLNDLENIVPFLGIGLLYSLSGPDLSTALLHFRLFVGARIYHTIAYLTPLPQPNRALAFFIGYGVTFS
MAYRLLKSRLYL

>Ornithorhynchus anatinus(F6QVY1)

MADDLILLAVVSVLSACQQTYFAWQVGKARFKYKIMPPAVSGSPEFERIYRAHQNCVECYPVFLTTFWIA
GWYFNQELVAILGLGYMYARHQYFYGYSEAVKRRIKGFRLTVGILTLLVVLSAVGIANRFLDEYVDFSLSKK
IRRLF

>Homo sapiens(P10620)

MVDLTQVMDDEVFMAFASYATIILSKMMLMSTATAFYRLTRKVFANPEDCVAFGKGENAKKYLRTDDR
VERVRRAHLNDLENIIPFLGIGLLYSLSGPDPSTAILHFRLFVGARIYHTIAYLTPLPQPNRALSFFVGYGVTLS
MAYRLLKSKLYL

>Rattus norvegicus(P08011)

MADLKQLMDNEVLMAFTSYATIILAKMMFLSSATAFQRLTNKVFANPEDCAGFGKGENAKKFLRTDEK
VERVRRAHLNDLENIVPFLGIGLLYSLSGPDLSTALIHFRIFVGARIYHTIAYLTPLPQPNRGLAFFVGYGVTL
SMAYRLLRSRLYL

>Mus musculus(Q91VS7)

MADLRQLMDNEVLMAFTSYATIILTKMMFMSSATAFQRITNKVFANPEDCAGFGKGENAKKFVRTDEK
VERVRRAHLNDLENIVPFLGIGLLYSLSGPDLSTALMHFRIFVGARIYHTIAYLTPLPQPNRGLAFFVGYGVT
LSMAYRLLRSRLYL

>Sus scrofa(P79382)

MADLTELMKNEVFMAFASYATIVLSKMMFMSTATAFYRLTRKVFANPEDCSSFGKGENAKKYLRTDER
VERVRRAHLNDLENIVPFLGIGLLYSLSGPDLSTAILHFRLFVGARIYHTIAYLTPLPQPNRGLAFFLGYGVTL
SMAYRLLKSRLYL

Mu Class GSTs

>Chlamys islandica(Q0P7I5)

AQQLRLMLQYGGVEYEDKRYELQKGTDGSYKCPEWFEQDKKTLKLDLPNLPYLIDGSTELTETDAIALYLA
EKLKLTGSSEKEKHLAHMTNLRIHDFRLAIIKVVYSPEHEALKGELFASFPERLALFSDFLGPKKKWLVGDSI
TFADFNFYDLLDILEVYVPTCLDEFPPLQRFIERFEALPKIKKYLASEQHQAVKNQPNNKSAYMGNSYVK

>Gallus gallus(P20136)

MVVTLGYWDIRGLAHAIRLLLEYTETPYQERRYKAGPAPDFDPSDWTNEKEKLGLDFPNLPYLIDGDVKL
TQSNAILRYIARKHNMCGETEVEKQRVDVLENHLMDLRMAFARLCYSPDFEKLKPAYLEQLPGKLRQLSR
FLGSRSWFVGDKLTFVDFLAYDVLDQQRMFVPDCPELQGNLSQFLQRFEALEKISAYMRSGRFMKAPI

>Mus musculus(O35660)

MPVTLGYWDIRGLGHAIRLLLEYTETGYEEKRYAMGDAPDYDRSQWLNDKFKLDLDFPNLPYLIDGSHK
VTQSNAILRYLGRKHNLCGETEEERIRVDILEKQVMDTRIQMGMLCYSADFEKRKPEFLKGLPDQLKLYSE
FLGKQPWFAGDKITFADFLVYDVLDQHRMFEPTCLDAFPNLKDFMARFEGLRKISAYMKTSRFLPSPVYL
KQATWGNE

>Rattus norvegicus(P08010)

MPMTLGYWDIRGLAHAIRLFLEYTDTSYEDKKYSMGDAPDYDRSQWLSEKFKLGLDFPNLPYLIDGSHKI
TQSNAILRYLGRKHNLCGETEEERIRVDVLENQAMDTRLQLAMVCYSPDFERKKPEYLEGLPEKMKLYSE
FLGKQPWFAGNKITYVDFLVYDVLDQHRIFEPKCLDAFPNLKDFVARFEGLKKISDYMKSGRFLSKPIFAK
MAFWNPK

>Mesocricetus auratus(P86214)

SMVLGYWDIRRMLLEFTDTSYEEKRYICGEAPDYDRSQWLDVKFKLDLDFPNLPYLMDGKNKITQSNAIL
RIRVDIMENQIMDFRQFSLFLGKKLTFVDFLTYDVLDQNRMFEPKCLDEFPNLKAFMCRCFKMPINNK

>Pongo abelii(Q5R8E8)

MPIILGYWNIRGLAHSIRLLLEYTDSSYEEKKYMMGDAPDYDRSQWLNEKFKLGLDFPNLPYLIDGTHKIT
QSNAILRYIARKHNLCGETEKEKIQEDILENQLMDNRMQLARLCYNPDFEKLKPEYLEGLPEMLKLYSQFL
GKQPWFLGDKITFVDFIAYDVLERNQVFEPSCLDAFPNLKDFISRFEGLEKISAYMKSSRFLPRPVFTKMA
VWGNK

>Macaca fuscata fuscata(Q9BEB0)

MPMTLGYWNIRGLAHSIRLLLEYTGSSYEEKKYTMGDAPDYDRSQWLNEKFKLGLDFPNLPYLIDGTHKI
TQSNAILRYIARKHNLCGETEKEKIREDILENQLMDNRMQLARLCYDPDFEKLKPEYLEGLPEMLKLYSQFL
GKQPWFLGDKITFVDFIAYDVLERNQVFEPSCLDAFPNLKDFISRFEGLEKISAYMKSSRFLPRPVFTKMA
VWGNK

>Macaca fascicularis(Q9TSM5)

MPMTLGYWDIRGLAHAIRLLLEYTDSSYEEKKYTMGDAPDYDRSQWLNEKFKLGLDFPNLPYLIDGTHKI
TQSNAILRYIARKHNLCGETEEEKIRVDILENQAMDVSNQLARVCYSPDFEKLKPEYLEGLPTMMQHFSQ
FLGKRPWFVGDKITFVDFLAYDVLDLHRIFEPKCLDAFPNLKDFISHFEGLEKISAYMKSSRFLPKPLYTRVA
VWGNK

>Schistosoma mansoni(P15964)

MAPKFGYWKVKGLVQPTRLLLEHLEETYEERAYDRNEIDAWSNDKFKLGLEFPNLPYYIDGDFKLTQSM
AIIRYIADKHNMLGACPKERAEISMLEGAVLDIRMGVLRIAYNKEYETLKVDFLNKLPGRLKMFEDRLSNK
TYLNGNCVTHPDFMLYDALDVVLYMDSQCLNEFPKLVSFKKCIEDLPQIKNYLNSSRYIKWPLQGWDAT
FGGGDTPPK

>Schistosoma japonicum(P26624)

VKLIYFNGRGRAEPIRMILVAAGVEFEDERIEFQDWPKIKPTIPGGRLPIVKITDKRGDVKTMSESLAIARFI
ARKHNMMGDTDDEYYIIEKMIGQVEDVESEYHKTLIKPPEEKEKISKEILNGKVPILLQAICETLKESTGNLT
VGDKVTLADVVLIASIDHITDLDKEFLTGKYPEIHKHRKHLLATSPKLAKYLSERHATAF

>Schistosoma haematobium(P30114)

MTGDHIKVIYFNGRGRAESIRMTLVAAGVNYEDERISFQDWPKIKPTIPGGRLPAVKITDNHGHVKWML
ESLAIARYMAKKHHMMGETDEEYYNVEKLIGQVEDLEHEYHKTLMKPEEEKQKITKEILNGKVPVLLDIIC
ESLKASTGKLAVGDKVTLADLVLIAVIDHVTDLDKEFLTGKYPEIHKHRENLLASSPRLAKYLSDRAATPF

>Schistosoma japonicum(P08515)

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAII
RYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYL
NGDHVTHPDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFG
GGDHPPK

>Sarcoptes scabiei(Q2YFE9)

MATRTTDSNSNDSKLPILAYWNIRGNAQPIRLLLRYTKTPYKEKSYNFGKYEQDKAIWRADKPHLGLDFP
NLPYYIDGDLRLTQSLTILRYLAKKHHLAGINETERIRIDLMEQQLRDFRNQFIDATNDANFEKARVIYLARL
PEKLQSLSNFLKDRPFFAGNSISYVDFMAYEFIDQHYYLNPDLFGQNQQWRNLIDFLHRIESFPTIKEYQYS
EDYIRHPSGLLIAWYEAKFFSTFNRSLGDQPSEQLRKEFIRSEMVSN

>Rattus norvegicus(P04905)

MPMILGYWNVRGLTHPIRLLLEYTDSSYEEKRYAMGDAPDYDRSQWLNEKFKLGLDFPNLPYLIDGSRKI
TQSNAIMRYLARKHHLCGETEEERIRADIVENQVMDNRMQLIMLCYNPDFEKQKPEFLKTIPEKMKLYSE
FLGKRPWFAGDKVTYVDFLAYDILDQYHIFEPKCLDAFPNLKDFLARFEGLKKISAYMKSSRYLSTPIFSKLA
QWSNK

>Conorchis sinensis(Q25595)

MAPVLGYWKIRGLAQPIRLLLEYVGDSYEEHSYGRCDGEKWQNDKHNLGLELPNLPYYKDGNFSLTQSL
AILRYIADKHNMIGNTPVERAKISMIEGGLVDLRAGVSRIAYQETFEQLKVPYLQQLPSTLRMWSQFLGN
NSYLHGSTPTHLDFMFYEALDVIRYLDPTSVEAFPNLMQFIHRIEALPNIKAFMESDRFIKWPLNGWSAYF
GGGDAPPK

>Cyphoma gibbosum(A7LFK1)

MPTLAYWKIRGLAQPMRLLLNYAGEKFEDKQYEQGDAPGFSVEEWTKEKFSLGLDFPNLPYYIDGDIKIT
QSNAILRYIANKHNLMGKTPKEKAQVDMMLENAMDLRNGVVKMAYNKDYEKMIKEYEPKCKEILAGY
EKWLSSHKWFGCDTVTVADFPMYELLDQHRLMIKDVLVPYPNITKFMAAFEALPAIKAYMASDKFMKR
PINNKIASFK

>Cyphoma gibbosum(A7LFK0)

MPTLGYWKTRGIGQPIRLLLNYVKEEFDDVYYEMGDAPDYSRDAWLNVKYTLGMNFPNLPYYIDGDVK
VSQSNAILRYIARKHDLLGKTDEEKTNVDMMLDIAMDMRNGVIKMVYNKDYEKLIKEYEPKCKEILAGYE

KWLSSHKWFGGDDVTVADFHMYELLDQHRLMIKDVLVPYPNITKFMAAFEALPAIKAYMASDKFMKQ
PINNKFASFI

>Dermatophagoides pteronyssinus(P46419)

MSQPILGYWDIRGYAQPIRLLLTYSGVDFVDKRYQIGPAPDFDRSEWLNEKFNLGLDFPNLPYYIDGDMK
MTQTFAILRYLGRKYKLNGSNDHEEIRISMAEQQTEDMMAAMIRVCYDANCDKLKPDYLKSLPDCLKL
MSKFVGEHAFIAGANISYVDFNLYEYLCHVKVMVPEVFGQFENLKRYVERMESLPRVSDYIKKQQPKTFN
APTSKWNASYA

>Dermatophagoides pteronyssinus(Q2YFE4)

MNKPTLGYWDLRGLGQPIRLMLAYAGVDYVDKRYTLGPDMDRSEWLKDKFNLGLDFPNLPYYIDGDVK
MTQSMAILRYLARKYNMDGSNEQERVRISMAEQQVYDMFMAMVRVCYDPNMEKLRVDYLKTLPDSL
KLMSKFMANHDFIAGSKISYADFYLYEYMCRIKVMVPEVYGQFENLKKFVERFESLPRVSDYIKKQTPKTF
NAAMAKWNGSYP

>Dermatophagoides pteronyssinus(Q2YFE5)

MSQPILGYWDIRGYAQPIRLLLTYSGVDFVDKRYQIGPAPDFDRSQWLNEKFNLGLDFPNLPYYIDGDM
KMTQTFAILRYLGRKYKLNGSNDHEEIRISMAEQQTKDMMAAMIRVCYDANCDKLKPDYLKSLPDCLKL
MSKFVGEHPFVAGANISYVDFYLYEYLCRVKVMVPEVFGQFENLKRYVERMESLPRVSDYIKKQQPKTFN
APTSKWNASYA

>Echinococcus granulosus(O16058)

MAPTLAYWDIRGLAEQSRLLLKYLEVEYDDKRYKIGSTPTFDRSAWLSEKFSLGLDFPNLPYYIDGDFKLTQ
SGAILEYIADRHGMIPDCKKRRAVLHMLQCEVVDLRMAFTRTCYSPDFEKLKPGLFETLAQKLPNFEAYLG
EKEWLTGDKINYPDFSLCELLNQLMKFEPTCLEKYPRLKAYLSRFENLPALRDYMASKEFKTRPCNGASAK
WRGDC

>Fasciola hepatica(P56598)

MPAKLGYWKIRGLQQPVRLLLEYGEKYEEQIYERDDGEKWFSKKFELGLDLPNLPYYIDDKCKLTQSLAILR
YIADKHGMIGSTPEERARVSMIEGAAVDLRQGLSRISYDPKFEQLKEGYLKDLPTTMKMWSDFLGKNPY
LRGTSVSHVDFMVYEALDAIRYLEPHCLDHFPNLQQFMSRIEALPSIKAYMESNRFIKWPLNGWHAQFG
GGDAPPSHEKK

>Homo sapiens(P09488)

MPMILGYWDIRGLAHAIRLLLEYTDSSYEEKKYTMGDAPDYDRSQWLNEKFKLGLDFPNLPYLIDGAHKI
TQSNAILCYIARKHNLCGETEEEKIRVDILENQTMDNHMQLGMICYNPEFEKLKPKYLEELPEKLKLYSEFL
GKRPWFAGNKITFVDFLVYDVLDLHRIFEPKCLDAFPNLKDFISRFEGLEKISAYMKSSRFLPRPVFSKMAV
WGNK

>Homo sapiens(P28161)

MPMTLGYWNIRGLAHSIRLLLEYTDSSYEEKKYTMGDAPDYDRSQWLNEKFKLGLDFPNLPYLIDGTHKI
TQSNAILRYIARKHNLCGESEKEQIREDILENQFMDSRMQLAKLCYDPDFEKLKPEYLQALPEMLKLYSQF
LGKQPWFLGDKITFVDFIAYDVLERNQVFEPSCLDAFPNLKDFISRFEGLEKISAYMKSSRFLPRPVFTKMA
VWGNK

>Homo sapiens(P46439)

MPMTLGYWDIRGLAHAIRLLLEYTDSSYVEKKYTLGDAPDYDRSQWLNEKFKLGLDFPNLPYLIDGAHKIT
QSNAILRYIARKHNLCGETEEEKIRVDILENQVMDNHMELVRLCYDPDFEKLKPKYLEELPEKLKLYSEFLG
KRPWFAGDKITFVDFLAYDVLDMKRIFEPKCLDAFLNLKDFISRFEGLKKISAYMKSSQFLRGLLFGKSAT
WNSK

>Homo sapiens(Q03013)

MSMTLGYWDIRGLAHAIRLLLEYTDSSYEEKKYTMGDAPDYDRSQWLNEKFKLGLDFPNLPYLIDGAHKI
TQSNAILCYIARKHNLCGETEEEKIRVDILENQAMDVSNQLARVCYSPDFEKLKPEYLEELPTMMQHFSQF
LGKRPWFVGDKITFVDFLAYDVLDLHRIFEPNCLDAFPNLKDFISRFEGLEKISAYMKSSRFLPKPLYTRVAV
WGNK

>Homo sapiens(P21266)

MSCESSMVLGYWDIRGLAHAIRLLLEFTDTSYEEKRYTCGEAPDYDRSQWLDVKFKLDLDFPNLPYLLDG
KNKITQSNAILRYIARKHNMCGETEEEKIRVDIIENQVMDFRTQLIRLCYSSDHEKLKPQYLEELPGQLKQF
SMFLGKFSWFAGEKLTFVDFLTYDILDQNRIFDPKCLDEFPNLKAFMCRFEALEKIAAYLQSDQFCKMPIN
NKMAQWGNKPVC

Omega Class GSTs

>Bombyx mori(A8R5V3)

MSAIKDSRNINFNTKHLRKGDPLPPFNGKLRVYNMRYCPYAQRTILALNAKQIDYEVVNIDLIDKPEWLTT
KSAFAKVPAIEIAEDVTIYESLVTVEYLDEVYPKRPLLPQDPLKKALDKIIVEASAPIQSLFIKILKFSDTVNEEH
VAAYHKALDFIQEQLKNRGTVFLDGSEPGYADYMIWPWFERLRAFAHDERVRLEPSKYSLLLEYIDNMLK
DSAVSQYLIPLEILAKFHEAYTKKERPNYELLNECLKSF

>Saccharomyces cerevisiae(P48239)

CPFTHRAILARSLKKLEPVLGLVLSHWQLDSKGARFLPAPHRPEKYKERFFTATGGIASAKLDESEELGDVN
NDSARLFVDGAFDPVENISRLSELYYLNDPKYPGTKFTVPVLWDSKTRKIVNNESGDIIRILNSGVFDEFIQS
EETNVIDLVPHDLIDEIDKNIKWVVHPKINLGVYKVGLAENGKIYETEVKTLFENLQKMECVLKENYKRLEEQ
FSGNKQKILAKYFVLGQRLTEADIRLYPSIIRFDVVYVQHFKCNLKTIRDGFPYLHLWLINLYWNYAEFRFTT
DF

>Schizosaccharomyces pombe(O94524)

DRYHLYVSYACPWAHRTLIVRKLKGLENVIPVHVVGWLMGPNGWNFDKENDSTGDPLYNSPYLRNLYF
RADPNYNMRFTVPVLWDSKYNTIVNNESAEIIRMFNDAFNEVIEDEEKRVVDLYPSSLRTKIDELNDYFYD
TVNNGVYKTGFATTAEAYEKNVRVVFQGLDRLEQVLKESKGPFLLGDHLTETDVRLYTTIVRFDPVYVQH
FKCNIGTIRHNYPHINQWLKRLYWKHPAFHETTDFKHIKCHYT

>Aspergillus ruber(A0A017SRP4)

MPPPDADLYPEASGAAKALVEEHSVEQPLKLYAGWFCPFVQRVWLALEEKQIPYQYIEVNPYHKSQSLLS
LNPRGLVPTLSVSHSGISKPLYESTVILEYLEEAYPDHKPCLLPEDPYERARVRIWVDYVTSRIIPAFHRFLQY
QEGQSSSSIDTLRNEFLNHLKEWTKEAHPDGPFFLGKDVSIPDLVLSPWAIRLWIFDEFKDGGLGIPFEGQ
GGEDESVWSRWRKWLAAIEARPSIQQTTSEKEFYIPIYKRYADNTAESELAKATRTGRGVP

>Aedes albopictus(A0A023EL34)

MSNGKHLAKGSTPPVLGNDGKLRLYSMRFCPYAQRVHLILDAKNIAYHTIYINLSEKPEWYFDKNPLGKV
PALEVPGKENITLYESLVVADYIEEAFPDKQRKLYPSDPFKKAQDRILIERFNGAVISPYYRILFSSDGIPPGAI
TEFGTGLDIFETELKNRGTSYYGGDKPGMLDYMIWPWCERVDLLKFALGDKYELDKQRFGKLLQWRDL
MEKDDAVQKSFLSTENHTKFLQSRKSGENNYDILS

>Amblyomma triste(A0A023GGV7)

MSAWALETGSKLPPLVPGKLRLYSMRFCPYAQRALLMLKAKGVDHEVVNVSLRNRPEWYNEVLPAGTV
PVLYQDEKVISGSMPIAEYLEEAYPEPHLLPTDPYLKALDRSFLDVALPCAGLISSISLNKGVKEEHWANFLK
KIEAFDKELAKRKTKFFGGEKPGLVDYVVWPAFPAALAYSKLYPDLKMPAAEQFPHFSRWLQAMREQPV
VKAVVNEDHVLLYAKSGIDGDRDFNAGLK

>Amblyomma triste(A0A023GK27)

MSAWALETGSKLPPLVPGKLRLYSMRFCPYAQRALLMLKAKGVDHEVVNVSLRNRPEWYNEVLPAGTV
PVLYQDEKVISGSMPIAEYLEEAYPEPHLLPTDPYLKALDRSFLDVALPCAALISSISLNKGVKEEHWANFLK
KIEAFDKELANRKTTFFGGEKPGLVDYVVWPAFPAALAYSKLYPDLKMPAAEQFPHFSRWLQAMREQP
VVKAVVNEDHVLLYAKSGIDGDRDFNAGLQ

>Macaca fascicularis(A0A023JBX8)

MSQDATRTLGKGSQPPGPVPEGLIRIYSMRFCPYSHRTRLVLKAKDIRHEVVNINLRNKPEWYYTKHPFG
HIPVLETSQCQLIYESVIACEYLDDAYPGRKLFPHDPYERARQKMLLELFCKVPHLTKECLVALRCGRECTD
LKASLRQEFCNLEEILEYQNTTFFGGTCTSMIDYLLWPWFERLDVYGIADCVSHTPALRLWISAMKWDPT
VCALLTDKSIFQGFLNLYFQNNPNAFDFGLC

>Rattus norvegicus(Q9Z339)

MSGASARSLGKGSAPPGPVPEGQIRVYSMRFCPFAQRTLMVLKAKGIRHEIININLKNKPEWFFEKNPFG
LVPVLENTQGHLITESVITCEYLDEAYPEKKLFPDDPYEKACQKMTFELFSKVPSLVTSFIRAKRKEDHPGIK
EELKKEFSKLEEAMANKRTAFFGGNSLSMIDYLIWPWFQRLEALELNECIDHTPKLKLWMATMQEDPVA
SSHFIDAKTYRDYLSLYLQDSPEACDYGL

>Sus scrofa(Q9N1F5)

MSGGSARSLGKGSAPPGPVPEGLIRVYSMRFCPFAQRTLLVLNAKGIRHQVININLKNKPEWFFQKNPS
GLVPVLENSQGQLIYESAITCEYLDEAYPGKKLLPDDPYEKACQKMVFELSSKVPPLLIRFIRRENEADCSGL
KEELRKEFSKLEEVLTKKKTTYFGGSSLSMIDYLIWPWFERLEALELNECIDHTPKLKLWMAAMMKDPAV
SALHIEPRDLRAFNDLYLQNSPEACDYGL

>Phanerochaete chrysosporium(J7JYU5)

MAQKEQITFYTHFYSPYCDRVHLALEEVKADYTVYTVDVMNKPKWYTEQINPVGKIPAITYGGPKVKPEE
PSPESCKLRESLVILEFLADLFPEAGLLPTDPVLRAKARLFASDVDAHVFEGFKAYFFMREPASKLLDALDRF
QQQLPARGFAVGDKWTLADMAAAPFLVRTYLLLEHDLGVYPAGEGPKTLALLRGERFARLNQYLADLR
AQPSFKATWDEAAQVAIWKSNPMFKRE

>Oncorhynchus kisutch(L7QHV5)

MASEKCFAKGSSAPGLVAKGQIRLYSMRFCPFAHRTRLVLHAKGIKHDTVNINLKDKPEWFLKKNPLGLV
PTLETSSGQVIYESPITCDYLDEVYTDKKLLPADPFQKAQQKMMLENFSKVTPYFYKIPMGKQNGEDISVL
EGELKEKFVKLNEDLVNKKSKFFGGNAITMIDYMMWPWFERLEIFELKHCLDGTPELKKWTEHMSEDQ
TVKATMFPTETYKAFYKTYADGKPNYDYGL

>Apis mellifera(A0A088A0S0)

MSSKHLTIGSVAPPIVPGKIRLYSMRFCPYAQRIHLVLDAKHIPHDVVYVNLTHKPDWLLEKSPLGKVPCIE
LEGGEILYESLVIAEYLDDTYPQNKLYPNDPLARAKDKLLIGRFNSVINTMCKLFINTSIDQDIFDEALSELEL
FERELASRGTPFFHGNSPGMLDFMIWPWWERSNTIKMLRGDQFTIPHDRFKRLLEWRSAMKENPAIRS
NYLDTEIHAKYMQSRRAGTPQYDLITD

>Drosophila melanogaster(Q9VSL5)

MALPQKHFKRGSTKPELPEDGVPRFFSMAFCPFSHRVRLMLAAKHIEHHKIYVDLIEKPEWYKDFSPLGK
VPALQLTGVKDQPTLVESLIIAEYLDQQYPQTRLFPTDPLQKALDKILIERFAPVVSAIYPVLTCNPNAPKDA
IPNFENALDVFEVELGKRGTPYFAGQHIGIVDYMIWPWFERFPSMKINTEQKYELDTKRFEKLLKWRDL
MTQDEVVQKTALDVQLHAEFQKSKTLGNPQYDIAFKGTP

>Homo sapiens(P78417)

MSGESARSLGKGSAPPGPVPEGSIRIYSMRFCPFAERTRLVLKAKGIRHEVININLKNKPEWFFKKNPFGL
VPVLENSQGQLIYESAITCEYLDEAYPGKKLLPDDPYEKACQKMILELFSKVPSLVGSFIRSQNKEDYAGLKE
EFRKEFTKLEEVLTNKKTTFFGGNSISMIDYLIWPWFERLEAMKLNECVDHTPKLKLWMAAMKEDPTVS
ALLTSEKDWQGFLELYLQNSPEACDYGL

>Homo sapiens(Q9H4Y5)

MSGDATRTLGKGSQPPGPVPEGLIRIYSMRFCPYSHRTRLVLKAKDIRHEVVNINLRNKPEWYYTKHPFG
HIPVLETSQCQLIYESVIACEYLDDAYPGRKLFPYDPYERARQKMLLELFCKVPHLTKECLVALRCGRECTNL
KAALRQEFSNLEEILEYQNTTFFGGTCISMIDYLLWPWFERLDVYGILDCVSHTPALRLWISAMKWDPTV
CALLMDKSIFQGFLNLYFQNNPNAFDFGLC

>Danaus plexippus(A0A212ENR8)

MSEKHLQSGDSLPPFTGKLRLFAMRFCPYAERSILCLNAKKLQYDLVFINLDHKPEWIFQFNPKGAVPALE
YEEGKAIFDSNVINVYLDEKYPEIPLQNSDPLRRAQDKLLVEMFAGAQSAYYTAAFNPQAVEPSMVENFH
KGLDLLQKEIESRGTKFLNGDEPGLVDYTIWPFLERFEALPILGQQEFAIDKSKYEILITYMAAMRDSPAVK
AYALAPDTHAKFTESRIKGDANYNMLDTSAVCCMRPRKKKE

>Danaus plexippus(A0A212ENT1)

MSSKAITGKINFNTKHLKRGDPLPPYNGKLRVYNMRYCPFAQRTILALNAKQMDYEVVNINLMDKPEW
LTRKSAFGKVPAIEINEDVCIFESLVTVEYLDEAYPQRPLLPKDPLRKALDKILIEASGPIHTMMFKTVKMPD
SITEDNLKAYESSLQYIQNELINRKTKFLSGNEPGYVDYMIWPWFERIGALKKFDERAGIDSSKFGLLLEYCS
NMAKDPAVSDYLLPDDILFKYFEGYKAGAPNYELITEE

>Danaus plexippus(A0A212ENC3)

MAYYPHRTAGPTPPPPLTDKLRLYHVDMNPYGHRVLLILEAKRAKYEVYRLDPLNLPQWFKTKNPRLKIP
VLEIPTDQGDKYLFESVVICDYLDEKYPRNPLHSRDPYVKAQDRLLIERFNELIKGSLECFDTNFAFGGEQIF

QTLDIFEKELASRGTNYFGGHSPGMLDYMIWPWVERLYLLRCVNERKFDEKRTLFPNFADWGDQMQL
DDVVKRHSNSPSEYFAYYRNARTHSMGYYL

>Plutella xylostella(255AA)

MSEKHLQTGDALPPFGGKLRLFAMRFCPYAERSVLVLNAKNIPYDLVFINLDQKPEWIFNFSPRGAVPAL
EYEQGKGIFDSNVINVYLDEKYPEVPLQAADPLRRAQDKLIVENFSAAQSAYYTAAFNAQALQPSHLENY
HKGLELLQKELETRGTKFLHGDQAGLVDYTLWPFLERFEALPLLGKSEYAIDKSKYDILLTYMESMKQVPA
VKTYYLSADMHAKFTESRVKGDPNYNMLDSSAEVCCFRPRKKKE

>Plutella xylostella(249AA)

MSTRGIKFNTKHLRKGDPLPPYNGKLRLYNMRYCPFAQRTVLALNAKDIDYEVVNINLFEKPEWLTSKSA
FGKVPSLEIKEGLSIYESLVTVEYLDEVYPQRPLLPKDPVQRALDKIIVEACTPIQGLFIKLIKFPESISEDTVAA
YHKALHFLQEQLQSRGTRFFGGDQPGFVDYMIWPWFERVLPYQKVESRVQIDAGKFKLLLEYLQNLKQ
DPVVKQYLIEDEVLFKFLEPYKTGGEPNYDLLLEA

>Plutella xylostella(241AA)

MSFYYQERPAGPTPPGPLSNKLRLYHVDMNPYGHRVLLILDAKKVPYEVCKLDPLRLPEWFREKNPRLKI
PVLEIPTDQGDKYLFESIVICDYLDERYTRNPLHSRDPFVKAQDRLLIERFNELIKGSLECFDTNFAFGNEQII
QTVNIFEKELESRGTVYFGGDRPGMLDYMIWPWIERLYMLRCLNPTKFDEKRHIFPNFADWGDQMQL
DEVVKKHASSPEDNFEYYKNARAHSMGYYL

>Drosophila melanogaster(Q9VSL4)

MALPQKHFKRGSPKPEIPEDGVLRYYSMRFCPYSQRAGLVLAAKKIPHHTVYIDLSEKPEWYIDYSPLGKV
PAIQLPNLPGQPALVESLVIAEYLDEQYPGEGSLFPKDPLQKALDRILIERLSPAVSAIYPVLFTKNPPADAIK
NFETALDVFEQEITKRGTPYFGGNKIGIADYMIWPWFERFPALKYTLDEPYELDKTRYQNLLKWRDLVAQ
DEAVKATALDARIHAKFMKTRHENKPDYDVAFQPL

>Drosophila melanogaster(Q9VSL6)

MSNTQHLTIGSPKPVFPDDGILKLYSMRFCPYAHRVHLVLDAKKIPYHAIYINLRDKPEWFSLVSSSTKVPA
LELVKEQGNPVLIESLIICDYLDEKYPEVPLYPKDLLKKAQEKILIERFGQFINAFYYLLLHDNPEQLVDTDHY
AGLVVYEEELKRRCTKFFGGDSPGMLDYMMWPWCERFDSLKYTFEQKFELSPERFPTLIKWRDLMIQD
RAVKCFYLDGQTHAKYMNSRRSGQADYNMLYNEAKRVKLG

>Bombyx mori(Q2F689)

MSEKHLQTGDVLPPYSGKLRVFAMRFCPYAERTVLTLNAKNIPYDLVFINLDQKPEWIFNFSPKGTVPALE
YEPGKALFDSNIINVYLDEKYPEIPLQASDPLRRAQDKILVESFAPAQSAYYTAAFNAQALEPSMVETYHK
GLEGLQKELETRSTKYLHGDEPGWVDYTLWPFLERFEALPLIGKAEFAIDQTKYERLVTYIEAMKNVPAVK
SYFLAAETHAKFIESRAQGDANYNMLDTSAVCCMRPRKKKE

>Bombyx mori(Q2I0J6)

MSAIKDSRNINFNIKHLRKGDPLPPFNGKLRVYNMRYCPYAQRTILALNAKQIDYEVVNIDLIDKPEWLTT
KSAFAKVPAIEIAEDVTIYESLVTVEYLDEVYPKRPLLPQDPLKKALDKIIVEASAPIQSLFIKILKFSDTVNEEH
VAAYHKALDFIQEQLKNRGTVFLDGSEPGYADYMIWPWFERLRAFAHDERVRLEPSKYSLLLEYIDNMLK
DSAVSQYLIPLEILAKFHEAYTKKERPNYELLNECLKSF

>Bombyx mori(Q1HPV9)

MTYFHSVNAGVIPPPALTDKLRLYHVDMNPYGHRVLLVLEAKRIKYEVYRLDPLRLPEWFRAKNPRLKIPV
LEIPTDQGDRFLFESVVICDYLDEKYTRHTLHSHDPYVKAQDRLLIERFNELIKGSLECFDTNFAFGSEQIIQT
LEIFEKELTNRGTNYFGGNRPGMLDYMVWPWVERLYLLRCVNDRKFVEKKSLFPNFADWGDQMQLD
DIVKKHAHSPQEYFDYYKNARAHSMGYYL

>Bombyx mori(B0LKP6)

MVSPKINFNTKHLGKGDPLPPWSGKLRVYNMRLCPFAQRTILTLNAKQIDYEVINIDLVNKPEWLPTKSIF
GKVPTIEVEDGVCICESLIIAEYLEEVYPEIPLISKDPIKKAYEKIIIEASEPIFVMYFKVMRTPDTINDETLMSY
HKALTFFEGQLRNRGTRFLGGEKPGFADYMIWPWFERIQSMNDEKLKIKSAKFDLLVAYIENMYKDPAV
SQYLLPKDVMDKLHAEYKTGKFEVQSIEDLL

>Caenorhabditis briggsae(A8XT16)

MPVLAGINSKVLKNGDSEPSPPPAGIYRIYNMRFCPWAQRALIYASVKNVPSEVINIHLKEKPDWYFSKH
YKGQVPALELDEGKKHVIESAHIPEYLDDLFPESRILPSDPYEKVQQKLLLERLAAVAPAFYAAAQAANNPE
GRDEKYAALVKAFEDAEKLLTGDFFSGKAKPGFADYLIFPNYQRVFWLSHILPNSPFSSESFPGPNFPKLAK
WYRTLDSIPEVAAASQPTEMGVGFFNDYLKGTPNYDYGL

>Mus musculus(O09131)

MSGESSRSLGKGSAPPGPVPEGQIRVYSMRFCPFAQRTLMVLKAKGIRHEVININLKNKPEWFFEKNPLG
LVPVLENSQGHLVTESVITCEYLDEAYPEKKLFPDDPYKKARQKMTLESFSKVPPLIASFVRSKRKEDSPNL
REALENEFKKLEEGMDNYKSFLGGDSPSMVDYLTWPWFQRLEALELKECLAHTPKLKLWMAAMQQDP
VASSHKIDAKTYREYLNLYLQDSPEACDYGL

>Caenorhabditis elegans(P34345)

MVLTGVTSKAIRKGDAEPPLSKGSFRVYNMRFCPWAERAMLYVAAKGIEAEVVNLNVTDKLEWYWTK
HYQGKAPAVEHNGKVVIESGFIPEYLDDAFPETRILPTDPYEKVQQKLLADRLTAVAHAVPLLFAVMRDR
TLKDEKQRKVFEVLKQAENLLANDFYAGSQPGYPDYLSFPFFEKIWWSASLDGVVDLPTIEFPGEEEYPKL
TKWFQKMISSDVVQSVTQSLEHGAAFMNAYATHQELNYDLGL

>Rattus norvegicus(Q6AXV9)

MSGDLTRCLGKGSCPPGPVPEGVIRIYSMRFCPYSHRTRLVLKAKSIRHEIININLKNKPDWYYTKHPFGQ
VPVLENSQCQLIYESVIACEYLDDVFPGRKLFPYDPYERARQKMLLELFCKVPQLSKECLVALRCGRDCTDL
KVALRQELCNLEEILEYQNTTFFGGDSISMIDYLVWPWFERLDVYGLADCVNHTPMLRLWISSMKQDPA
VCALHIDKNIFLGFLNLYFQNNPCAFDFGLCGPIVR

Pi Class GSTs

>Homo sapiens(P09211)

MPPYTVVYFPVRGRCAALRMLLADQGQSWKEEVVTVETWQEGSLKASCLYGQLPKFQDGDLTLYQSN
TILRHLGRTLGLYGKDQQEAALVDMVNDGVEDLRCKYISLIYTNYEAGKDDYVKALPGQLKPFETLLSQN
QGGKTFIVGDQISFADYNLLDLLLIHEVLAPGCLDAFPLLSAYVGRLSARPKLKAFLASPEYVNLPINGNGK
Q

>Mus musculus(P19157)

PPYTIVYFPVRGRCEAMRMLLADQGQSWKEEVVTIDTWMQGLLKPTCLYGQLPKFEDGDLTLYQSNAIL
RHLGRSLGLYGKNQREAAQMDMVNDGVEDLRGKYVTLIYTNYENGKNDYVKALPGHLKPFETLLSQNQ
GGKAFIVGDQISFADYNLLDLLLIHQVLAPGCLDNFPLLSAYVARLSARPKIKAFLSSPEHVNRPI

>Dirofilaria immitis(P46426)

MSYKLTYFPIRGLAEPIRLLLVDQGIKFTDEHIPKDDFVSIKSQFQFGQLPCFYDGDQQIVQSGAILRHLARK
FNLNGENNAETSYVDMFYEGIRDLHSKYTRMIYEAYETQKDPFIKNILPQELAKLEKLLATRDNGKNFILG
DKISFADYVLFEELDVQQILDPHCLEKFPLLKAFHQRLGDKPKIKEYCAKRNASKM

>Rattus norvegicus(P04906)

PPYTIVYFPVRGRCEATRMLLADQGQSWKEEVVTIDVWLQGSLKSTCLYGQLPKFEDGDLTLYQSNAILR
HLGRSLGLYGKDQKEAALVDMVNDGVEDLRCKYGTLIYTNYENGKDDYVKALPGHLKPFETLLSQNQGG
KAFIVGNQISFADYNLLDLLLVHQVLAPGCLDNFPLLSAYVARLSARPKIKAFLSSPDHLNRPI

>Wuchereria bancrofti(Q86LL8)

MSYKLTYFPIRGLAEPIRLVLVDQGIKFTDDRINASDWPSMKSHFHFGQLPCLYDGDHQIVQSGAILRHLA
RKHNLNGGNELETTHIDMFCEGIRDLHTKYAKMIYQAYDTEKDSYIKDILPVELAKFEKLLATRDDGKNFIL
GEKISYVDFVLFEELDIHQILDPHCLDKFPLLKAYHQRMEDRPGLKEYCKQRNRAKIPV

>Caenorhabditis elegans(P10299)

MTLKLTYFDIHGLAEPIRLLLADKQVAYEDHRVTYEQWADIKPKMIFGQVPCLLSGDEEIVQSGAIIRHLAR
LNGLNGSNETETTFIDMFYEGLRDLHTKYTTMIYRNYEDGKAPYIKDVLPGELARLEKLFHTYKNGEHYVI
GDKESYADYVLFEELDIHLILTPNALDGVPALKKFHERFAERPNIKAYLNKRAAINPPVNGNGKQ

>Bos Taurus(P28801)

MPPYTIVYFPVQGRCEAMRMLLADQGQSWKEEVVAMQSWLQGPLKASCLYGQLPKFQDGDLTLYQS
NAILRHLGRTLGLYGKDQQEAALVDMVNDGVEDLRCKYVSLIYTNYEAGKEDYVKALPQHLKPFETLLSQ
NKGGQAFIVGDQISFADYNLLDLLRIHQVLAPSCLDSFPLLSAYVARLNSRPKLKAFLASPEHMNRPINGN
GKQ

>Cricetulus longicaudatus(P46424)

MPPYTIVYFPVRGRCEAMRILLADQGQSWKEEVVTVETWRKGSLKSTCLYGQLPKFEDGDLTLYQSNAIL
RHLGRSLGLYGKDQREAALVDMVNDGVEDLRCKYITLIYTKYEEGKDDYVKALPGHLKPFETLLSQNQGG
KAFIVGDQISFVDYNLLDLLLIHQVLAPGCLDNFPLLSAYVARLSARPKIKAFLSSPDHVNRPINGNGKQ

>Mus musculus(P46425)

MPPYTIVYFPSPGRCEAMRMLLADQGQSWKEEVVTIDTWMQGLLKPTCLYGQLPKFEDGDLTLYQSNA
ILRHLGRSLGLYGKNQREAAQVDMVNDGVEDLRGKYGTMIYRNYENGKNDYVKALPGHLKPFETLLSQ
NQGGKAFIVGDQISFADYNLLDLLLIHQVLAPGCLDNFPLLSAYVARLSARPKIKAFLSSPEHVNRPINGNG
KQ

>Cricetulus migratorius(P47954)

MPPYTIVYFPVRGRCEAMRILLADQGQSWKEEVITGETWGKGSLKSTCLYGQLPKFEDGDLTLYQSNAIL
RHLGRSLGLYGKDQREAALVDMVNDGVEDLRCKYVTLIYTKYEEGKDDYVKALPGHLKPFETLLSKNQGG
KAFIVGDQISFADYNLLDLLLIHQVLAPGCLDNFPLLSAYVARLSARPKIKAFLSSPDHVNRPINGNGKQ

>Bufo bufo(P81942)

PEYTIIYFNARGRCEAMRMLMADQGAQWKEEVVTSDDWQKGDLKKAAVYGQLPGFKDGDFTLYQSN
AMLRLLARNHDLYGKNPREASLIDMVNDGVEDLRLKYLKMIYQNYENGKDDYVKALPTNLGHFERLLAS
NNEGKGFVVGAHISFADYNLVDLLHNHLVLAPDCLSGFPLLCAYVKRISSRPKLEAYLSSDAHKKRPINGN
GKQQ

>Bufo bufo(P83325)

SGYTLTYFPLRGRAEAMRLLLGDQGVSWTDDEVQMQDWAAGIRDLKKNAVFGQIPRFQEGDFVLYQS
QTILRLLARYGLSGSNEREIAINEMMNDGVEDLRLKYYKFIFWDNEANKEKFLEELATQLGYFERILTNNA
GKTFVLVGDKISYADYNLLDTLFCVLDLSPTCLSGFPLLSDYVERLGKRPKLQQYLKSEGRKRRPINGNGKQ

>Macaca mulatta(Q28514)

MPPYTVVYFPVRGRCAALRMLLADQGQSWKEEVVTMETWQEGSLKASCLYGQLPKFQDGDLTLYQSN
TFLRHLGRTLGLYGKDQREAALVDMVNDGVEDLRCKYLSLIYTNYEAGKDDYVKALPGQLKPFETLLSQN
QGGKTFIVGDQISFADYNLLDLLLIHEVLAPGCLDAFPLLSAYVARLSARPKLKAFLASPEHVNLPINGNGK
Q

>Pongo abelii(Q5R8R5)

MPPYTVVYFPVRGRCAALRMLLADQGQSWKEEVVSMETWQEGSLKASCLYGQLPKFQDGDLTLYQSN
TILRHLGRTLGLYGKDQREAALVDMVNDGVEDLRCKYLSLIYTNYEAGKDDYVKALPGQLKPFETLLSQN
QGGKTFIVGDQISFADYNLLDLLLIHEVLAPGCLDAFPLLSAYVARLSARPKLKAFLASPEHVNLPINGNGK
Q

>Mesocricetus auratus(Q60550)

MPPYTIVYFPVRGRCEAMRLLLADQGQSWKEEVVTGDSWVKGSLKSTCLYGQLPKFEDGDLILYQSNAIL
RHLGRSLGLYGKDQKEAALVDMANDGVEDLRCKYVTLIYTKYEEGKDDYVKALPGHLKPFETLLSQNQG
GKAFIVGDQISFADYNLLDLLLIHQVLAPGCLDNFPLLSAYVARLSARPKIKAFLSSPDHVNRPINGNGKQ

>Xenopus laevis(Q8JFZ2)

MPGYVLTYFPVRGRAEPIRLLLADQGISWKEDEVQIPDWFSGKDARKKEAVFGQLPQFQDGDYVLYQS
NSILRYLGNKHGLTGANDEERGHIDMVNDGVEDLRQKYGRLIFFEYETGKDKYLKELPSQLDFFERILSKN
ANGSKFVVGQKISFADYNLLDILQCHLDLCSKSLSAYPLLTAYVERLVARPKISEYLKSDARNKRPITPKHKK

>Capra hircus(Q9TTY8)

MASYTIVYFPVQGRCEAMRMLLADQDQSWKEEVVAMQSWLQGPLKASCLYGQLPKFQDGDLTLYQS
NAILRHLGRTLGLYGKDQREAALVDMVNDGVEDLRCKYVSLIYTNYQAGKEDYVKALPQHLKPFETLLSQ
NKGGQAFIVGDQISFADYNLLDLLRIHQVLAPSCLDSFPLLSAYVARLNSRPKLKAFLASPEHVNRPINGN
GKQ

>Macaca fascicularis(A0A023JCR5)

MPPYTVVYFPVRGRCAALRMLLADQGQSWKEEVVTMETWQEGSLKASCLYGQLPKFQDGDLTLYQSN
TFLRHLGRTLGLYGKDQREAALVDMVNDGVEDLRCKYLSLIYTNYEAGKDDYVKALPGQLKPFETLLSQN
QGGKTFIVGDQISFADYNLLDLLLIHEVLAPGCLDAFPLLSAYVARLSARPKLKAFLASPEHVNLPINGNGK
Q

>Fukomys damarensis(A0A091DCX2)

MTSELQSGVSASLGPSPLPLWTRALEKTDLERRVPQTLKIEVRAKGAAGGGGPRNDSALGAERGGIRFIR
SRGLQACTRVPPPTEVHATRCSIQAAGYAASMPPYTIVYFPVRGRGEAMRMLLTDQGQSWKEEVVTVE
SWTQGPLKASCLYGQLPKFQDGDLTLYQSNAILRHLGRSLGLYGKDQREAALVDMVNDGVEDLRGKYV
TLIYTNYEAGKNDYVKALPGHLKPFETLLSKNQGGKAFIVGDQISFADYNLLDLLLNHQVLAPSCLDAFPLL
SAYVARLSARPKLKAFLASPDHVNRPINGNGKQ

>Toxocara canis(A0A0B2VSH0)

MKPIYKLTYFDIRGLGEQIRLLLIDNKIPFEEVRIKSEDEWKGIKDSFVFGQLPCLKDDDVKIVQSGAIMRHL
ARRHDLYGRTEMDRTFADMFYEGIRDIQQRYIRMIYNEYEKKNEFIVDYLHDALNKLDALLESHEEGNGFI
LGENICFADYSLFELLDVLLILSPTCLLQCPKLKSFHQRFNERPSLQNYLMKRASANVRVNWNGKE

>Toxocara canis(A0A0B2VSM5)

MGYKVTYFAIRGLAEPIRLLLTDQSIPFEDSRIADKNEWQTMKHQFQFGQVPCLHDDDEQIVQSGAILRH
LARKHNLNGANEKETTYADMFYEGIRDLHNKYTRMIYFEYETEKDNFIKDVLPVELAKFEKLLQTRGGGT
GFVLGDKICFADYVLFEELDIMQILDAHALDKFPTLKAYHQRVHDRPLIKAYYKKREDAKVPVNGNGKQ

>Chlorocebus sabaeus(A0A0D9R8K4)

MWSEVRVLKHSRKSRLPAPAETVPSGNCWVFFWESFLAMKSRLNSRPPRWKTLDRVRVGPLQFQMGF
SAANKVRNMALDDADAPLGTWGRRKGFPGQLRGDSGSSEAPLCGRRPGCSGRRAWGRREPAGPSRG
AAGAVTQHWGGAGRDHPYKARRPQGLRWSFVVVAAAACAIMPPYTVVYFPVRGRCAALRMLLADQ
GQSWKEEVVTVETWQEGSLKASCLYGQLPKFQDGDLTLYQSNTFLRHLGRTLGLYGKDQREAALVDMV
NDGVEDLRCKYLSLIYTNYEAGKDDYVKALPGQLKPFETLLSQNQGGKTFIVGDQISFADYNLLDLLLIHEV
LAPGCLDAFPLLSAYVARLSARPKLKAFLASPEHVNLPINGNGKQ

>Heterocephalus glaber(A0A0N8ESH7)

MAPYTIVYFPVRGRCEAIRMLLADQGQSWKEEVVTMESWMQGQLRASCLYGQLPKFQDGDLTLYQSN
AILRHLGRSLGLYGKDQREAALVDMVNDGVEDLRCKYVTFIYTNYEAGKNEYVKALPGHLKPFETLLSKN
QGGKAFIVGDQISFADYNLLGLLLNHQVLAPGCLDAFPLLSAYVARLSARPKLKAFLASPDHVNRPINGNG
KQ

>Fundulus heteroclitus(A0A146NIX1)

MGYKLTYFAIRGLAEPIRLLFTDQGITFEDSRIKDKSEWPAMKQQFQFGQVPCLYDDDEQIVQSGAILRHL
ARKHNLNGANESETTYADMFYEGIRDLHSKYTHMIYGAYETEKDGFIKDILPVELAKFEKLLQTRGGGAAF
ILGDKICFADYVLFEELDIMQILDPHALEKFPTLKAYHRRMRDRPHLKTYCQKRDESKVPVNGNGKQ

>Macaca mulatta(A0A1D5RCG2)

MGPGEGRLRAPRLGVLPRAPANLSLSAAVPPYTVVYFPVRGRCAALRMLLADQGQSWKEEVVTMETW
QEGSLKASCLYGQLPKFQDGDLTLYQSNTFLRHLGRTLGLYGKDQREAALVDMVNDGVEDLRCKYLSLIY
TNYISFADYNLLDLLLIHEVLAPGCLDAFPLLSAYVARLSARPKLKAFLASPEHVNLPINGNGKQ

>Dipodomys ordii(A0A1S3FJI9)

MPPYTIVYFPVRGRCEAMRMLLADQGQSWKEEVVTGEIWQKGPLKASCLYGQLPKFQDGDLTLYQSN
AILRHLARTFGLYGKNQEEAALVDVVNDGVEDLRCKYVTLIYTNYEAGKNDYVKALPTHLKPFETLLSQNQ
GGQAFIVGDQISFADYNLLDLLLIHQVLDPGCLENFPLLTGYTKRLVARPKIKAFLDSPEHVKRPINGNGKQ

>Erinaceus europaeus(A0A1S3W606)

MPPYTIVYFPVQGRCEAMRTLLADQGQSWQEEVVTKDSWLQGPLKATCLYGQLPKFQDGDLTLYQSN
AILRHLGRTLGLYGKDQREAALIDMANDGVEDLRSKYAALIYTNYEAGKEAYVKALPGHLKPFETLLSQNQ
GGKAFLVGDQISFADYNLLDLLRIHQTLAPDCLGAFPLLSAYLERLSTRPKIKAFLASPEHLNRPINGNGKQ

>Mesocricetus auratus(A0A1U7Q833)

MPPYTIVYFPVRGRCEAMRLLLADQGQSWKEEVVTGDSWVKGSLKSTCLYGQLPKFEDGDLILYQSNAIL
RHLGRSLGLYGKDQKEAALVDMANDGVEDLRCKYVTLIYTKYEEGKDDYVKALPGHLKPFETLLSQNQG
GKAFIVGDQISFADYNLLDLLLIHQVLAPGCLDNFPLLSAYVARLSARPKIKAFLSSPDHVNRPINGNGKQ

>Scleropages formosus(A0A1W5AH81)

MPPYIITYFPVRGRCGAMRILMADQGVEWKEKVVVFDEWMKGDLKATCVFGQLPKFEDGDLVLYQSN
AILRHLGRKHEAAGSDDKEAALIDMMNDGVEDLRLKYIKLIYQDYDTGKDQYIKDLPGHLSKFEAVLAKN
KSGFLIGGKISFADYSLFEVLLNHLVLCSSCLDTFPALKSFVESMSARPGIKAFLDSDAYKKLPINGNGKQ

>Mizuhopecten yessoensis(A0A210Q711)

MELIYFPVKGRAEVIRLMLIDNGTSYTETSCANDWDSKWKPKMAFGQTPCLKDGDLTLVQSNSIIRHLAR
KFSLYGANEAEACRADIINDSVEDLRSAYVNLIYNNYDAGKEEYINKLPAKLQYFEKYIEGKSPYVLGDHICF
ADYSLFELLDIHLVLAPSCLDKFPALKALHTTVGSRQKVKAHRDSDAVKAMPINGNGKQ

>Castor canadensis(A0A250YCG2)

MPPYTIVYFPVQGRCEAMRMLLADQDQSWKEEVVMLDSWYQSPLKASCLYGQLPKFQDGDLTLYQSN
AILRHLGRSFGLYGKDQREAALVDMVNDGVEDIRSKYIALIYTNYEAGKNDYVKALPTHLKPFETLLSQNK
GGQAFIVGDQISFADYNLLDLLLTHKVLAPSCLDAFPLLSAYVSRLSGRPKVKAFLASPDHVNRPINANGK
Q

>Onchocerca volvulus(P46427)

MSYKLTYFSIRGLAEPIRLFLVDQDIKFIDDRIAKDDFSSIKSQFQFGQLPCLYDGDQQIVQSGAILRHLARK
YNLNGENEMETTYIDMFCEGVRDLHVKYTRMIYMAYETEKDPYIKSILPGELAKFEKLLATRGNGRNLILG
DKISYADYALFEELDVHQILDPHCLDKFPLLKAFHQRMKDRPKLKEYCEKRDAAKVPVNGNGKQ

>Sus scrofa(P80031)

PPYTITYFPVRGRCEAMRMLLADQDQSWKEEVVTMETWPPLKPSCLFRQLPKFQDGDLTLYQSNAILRH
LGRSFGLYGKDQKEAALVDMVNDGVEDLRCKYATLIYTNYEAGKEKYVKELPEHLKPFETLLSQNQGGQ
AFVVGSQISFADYNLLDLLRIHQVLNPSCLDAFPLLSAYVARLSARPKIKAFLASPEHVNRPINGNGKN

Phi Class GSTs

>Triticum aestivum(P30110)

MSPVKVFGHPMLTNVARVLLFLEEVGAEYELVPMDFVAGEHKRPQHVQLNPFAKMPGFQDGDLVLFE
SRAIAKYILRKYGGTAGLDLLGENSGIEELAMVDVWTEVEAQQYYPAISPVVFECIIIPFIIPGGGAAPNQT
VVDESLERLRGVLGIYEARLEKSRYLAGDSITFADLNHIPFTFYFMTTPYAKVFDDYPKVKAWWEMLMAR
PAVQRVCKHMPTEFKLGAQY

>Populus trichocarpa(B9GQ64)

TPVKVYGPPLSTAVSRVLVTLLEKDVPFQIIPVDMSKGEHKKPDYLKIQPFGQVPAFQDESISLFESRSICRY
VCEKYADRGDKGLYGTNPLERASIDQWVEAEGQSFGPSSGALVFQLAFAPRMNIPQDQGVIKQNEEKL
GKVLDIYEQRLGESRFLAGDEFTFADLSHLPNGDYLVNATDKGHLFTSRENVGRWWNEISDRESWKKVI
EMRKSG

>Arabidopsis thaliana(O80852)

MVLKVYGPHFASPKRALVTLIEKGVAFETIPVDLMKGEHKQPAYLALQPFGTVPAVVDGDYKIFESRAVM
RYVAEKYRSQGPDLLGKTVEDRGQVEQWLDVEATTYHPPLLNLTLHIMFASVMGFPSDEKLIKESEEKLA
GVLDVYEAHLSKSKYLAGDFVSLADLAHLPFTDYLVGPIGKAYMIKDRKHVSAWWDDISSRPAWKETVA
KYSFPA

>Zea mays(P04907)

MAPLKLYGMPLSPNVVRVATVLNEKGLDFEIVPVDLTTGAHKQPDFLALNPFGQIPALVDGDEVLFESRA
INRYIASKYASEGTDLLPATASAAKLEVWLEVESHHFHPNASPLVFQLLVRPLLGGAPDAAVVEKHAEQLA
KVLDVYEAHLARNKYLAGDEFTLADANHALLPALTSARPPRPGCVAARPHVKAWWEAIAARPAFQKTV
AAIPLPPPPSSSA

>Zea mays(P12653)

MAPMKLYGAVMSWNLTRCATALEEAGSDYEIVPINFATAEHKSPEHLVRNPFGQVPALQDGDLYLFESR
AICKYAARKNKPELLREGNLEEAAMVDVWIEVEANQYTAALNPILFQVLISPMLGGTTDQKVVDENLEKL
KKVLEVYEARLTKCKYLAGDFLSLADLNHVSVTLCLFATPYASVLDAYPHVKAWWSGLMERPSVQKVAAL
MKPSA

>Triticum aestivum(P30111)

MSPVKVFGHPMLTNVARVLLFLEEVGAEYELVPVDFVAGEHKRPQHVQLNPFAKMPGFQDGESLHIKS
RAIAKYILRKYGGTAGLDLLGENSGIEELAMVDVWTEVEAQQYYPAISPVVFECIIIPFIIPGGGAAPNQTV
VDESLERLRGVLGIYEARLEKSRYLAGDSISFADLNHIPFTFYFMTTPYAKVFDEYPKVKAWWEMLMARP
AVQRVCKHMPTEFKLRARTRCLCTPRGCVPCRAGDDPTQEKRPPSRGWVIICPSTSSIPFQFQQHEESAC
ASARASPDSLL

>Arabidopsis thaliana(P42760)

MAGIKVFGHPASTATRRVLIALHEKNVDFEFVHVELKDGEHKKEPFILRNPFGKVPAFEDGDFKIFESRAIT
QYIAHEFSDKGNNLLSTGKDMAIIAMGIEIESHEFDPVGSKLVWEQVLKPLYGMTTDKTVVEEEEAKLAK
VLDVYEHRLGESKYLASDHFTLVDLHTIPVIQYLLGTPTKKLFDERPHVSAWVADITSRPSAQKVL

>Arabidopsis thaliana(P42761)

MVLTIYAPLFASSKRAVVTLVEKGVSFETVNVDLMKGEQRQPEYLAIQPFGKIPVLVDGDYKIFESRAIMR
YIAEKYRSQGPDLLGKTIEERGQVEQWLDVEATSYHPPLLALTLNIVFAPLMGFPADEKVIKESEEKLAEVL
DVYEAQLSKNEYLAGDFVSLADLAHLPFTEYLVGPIGKAHLIKDRKHVSAWWDKISSRAAWKEVSAKYSL
PV

>Zea mays(P46420)

MATPAVKVYGWAISPFVSRALLALEEAGVDYELVPMSRQDGDHRRPEHLARNPFGKVPVLEDGDLTLFE
SRAIARHVLRKHKPELLGGGRLEQTAMVDVWLEVEAHQLSPPAIAIVVECVFAPFLGRERNQAVVDENV
EKLKKVLEVYEARLATCTYLAGDFLSLADLSPFTIMHCLMATEYAALVHALPHVSAWWQGLAARPAANK
VAQFMPVGAGAPKEQE

>Hyoscyamus muticus(P46423)

MGMKLHGPAMSPAVMRVIATLKEKDLDFELVPVNMQAGDHKKEPFITLNPFGQVPAFEDGDLKLFESR
AITQYIAHTYADKGNQLLANDPKKMAIMSVWMEVESQKFDPVASKLTFEIVIKPMLGMVTDDAAVAEN
EEKLGKVLDVYESRLKDSKYLGGDSFTLADLHHAPAMNYLMGTKVKSLFDSRPHVSAWCADILARPAWS
KAIEYKQ

>Silene vulgaris(Q04522)

MTIKVHGNPRSTATQRVLVALYEKHLEFEFVPIDMGAGGHKQPSYLALNPFGQVPALEDGEIKLFESRAIT
KYLAYTHDHQNEGTSLIHKEKHEMAAQLVWEEVEAHQFDPVASKLAWELVFKGIFGMQTDTTVVEENE
AKLAKVLDVYEARLTESEYLGANDSFTLVDLHHLPLLGYLMGTQVKKLFEERAHVSAWCKKILARPSWEK
TLALQKQA

>Arabidopsis thaliana(Q84TK0)

MDCLQMVFKLFPNWKREAEVKKLVAGYKVHGDPFSTNTRRVLAVLHEKRLSYEPITVKLQTGEHKTEPFL
SLNPFGQVPVFEDGSVKLYESRAITQYIAYVHSSRGTQLLNLRSHETMATLTMWMEIEAHQFDPPASKLT
WEQVIKPIYGLETDQTIVKENEAILEKVLNIYEKRLEESRFLACNSFTLVDLHHLPNIQYLLGTPTKKLFEKRS
KVRKWVDEITSREAWKMACDQEKSWFNKPRN

>Arabidopsis thaliana(Q96324)

MVVKVYGQIKAANPQRVLLCFLEKDIEFEVIHVDLDKLEQKKPQHLLRQPFGQVPAIEDGYLKLFESRAIA
RYYATKYADQGTDLLGKTLEGRAIVDQWVEVENNYFYAVALPLVMNVVFKPKSGKPCDVALVEELKVKF
DKVLDVYENRLATNRYLGGDEFTLADLSHMPGMRYIMNETSLSGLVTSRENLNRWWNEISARPAWKKL
MELAAY

>Arabidopsis thaliana(Q9SRY6)

MGINASHVPETCYHHCNQTFESSRQCFKWCQELARKDEYKIYGYPYSTNTRRVLAVLHEKGLSYDPITVN
LIAGDQKKPSFLAINPFGQVPVFLDGGLKLTESRAISEYIATVHKSRGTQLLNYKSYKTMGTQRMWMAIE

SFEFDPLTSTLTWEQSIKPMYGLKTDYKVVNETEAKLEKVLDIYEERLKNSSFLASNSFTMADLYHLPNIQY
LMDTHTKRMFVNRPSVRRWVAEITARPAWKRACDVKAWYHKKKN

>Chondromyces apiculatus(A0A017SXI2)

MKIYGHPMSTCARKVLTVLAEKGHEAEMVLVDLMKGEQKKPEFLKLQPFGVIPVLDDDGFILYESRAIIRY
LDQKLSGTSLTPSDPKERALMEQWISVETSYLSPPAMKIVAQKLFVPMRGGQTDEAIVEVGRKETVRTLD
IMEQTLSKQEFLAGNSFSLADVSCMPYLGYLFPAGAGDLVTSRPGVAAWWERISSRPSWRKVAG

>Elaeis guineensis var tenera(A0A060IGG8)

MGVKVYGPTMSTCTARVLLCLEEVGAEYELVPINLSTGEHKQPAHLARNPIGQVPAFEDGALMLHESRAI
ARYVSRKYKSSGADLLKEGGLEESAMVDVWLEVESHQFDPAIGPIFFQSFIVPMIGGVPDQTVINTNLEKL
CKVLDIYEARLSKTKYLAGDFFSLADLSHVPLLYYFMGSPHASVVNSRPHVKAWWEAVSSRPACKKVTSA
MPGSA

>Theobroma cacao(A0A061DK50)

MVVKVYGPAYATPKRVLVCLIEKEVEFETVPVDLLKGEHKDPEYLKLQPFGTVPVTQDGDYILYESRAILRY
YAEKYKSQGTDLLGKTVEERGLVENWLEVEAQSYHPPIYTLTVQILFSSKLGFPRDENLIKESEEKLAKVLDI
YEERLSKSKYLGGDFFSLADLSHLPFTQYLVDQMGKEYMIRSRKHVSAWWDDISSRPSWQKVLQLYAAP
FKN

>Theobroma cacao(A0A061DKZ7)

MVVKVYGPAYASPKRVLVCLIEKEVEFETVPVDLLKGEHKDPEYLKLQPFGTVPVTQDGDYTLYESRAIMR
YYAEKYKSQGTDLLGKTVEERGLVENWLEVEAQSYNPPMFTLTVQIMLSSKLGLPRDENLIKESEEKLGKV
LDIYEERLSKSKYLAGNFFSLADLSHLPFTQYLVDHMGKEYMIKSRKHVSAWWDDISSRPSWQKVLQLYA
APFKN

>Theobroma cacao(A0A061EET2)

MVVKVHGSVRAACPQRVLACLLEKDVEFEIFHVDLDAGEHKRPEFLLRQPFGQIPAVEDGDLKLFESRAII
RYYAAKYVDRGPNLLGNSLEERAVVDQWLEVEAHNFNDLVKTLVFQIVILPRMGEHGDLALAHKCEQQL
EKVFDIYEQHLSKSSYLAGDSFTLADLSHLPAIRYLVNDAGMGHLVAERKHVNAWWEDISNRSAWKKL
MELAKY

>Theobroma cacao(A0A061DSD1)

MKNKKVYGSLNSAATLKVLACLFEHDLDFDFVPIDLEAGEHKKKPFLSMNPFGQVPVFEDGDVKQFESR
AIIRSMGHQYGKKGEELIYWDSREQAVVANWIDVEDHHFEPPALKLISELVIKPKKGLTPDEETVAEAEAK
LAKVLDVYEARLSKFKYLASDKYTIADMLHLPNIQALIGTQAKKLFDSRPRLSEWCTAILARPAWIKVVEM
QQKAQA

>Rhizoctonia solani(A0A066VYX0)

MVTVKLHGMPYSACAHRVWATAQEIGVTVELVPVDLAKVEHKTPEYVENYHPFGVIPVLVDEDGTKLFE
SRAICRYLVAKYGKGSPLLPDPSDAKAYGLFEQAASIEYSNFDPSAASLTYERIVAPMRKEKPNEELIKKCID
TLISKMDGYERILSKQKYLAGDTFTLADLFHLPYGQIISVIEPRILASKPHVKAWWDDISARESWKATLKFS
GH

>Medicago truncatula(A0A072UH22)

MFSSICAESRAIMRYYAEKYRSQGVELLGKTIEEKGLVEQWLEVEAHNFHPSAYNLTCHVLCPTLLGGSSP
DPKVIEESEAKLVKVFNIYEERLSKNKYLAGDFFSLADISHLPFMDYVVNNMGKDYLIKDKKHVSAWWND
ISSRPSWNKVLELYKPPI

>Medicago truncatula(A0A072VNL4)

MALKIYGLAMSTNTTRAMICLHEKEVDFELIPVNVFTSEHKQPPFLNKNPFGLIPVLEDDDLTLFESRAITS
YVAEKYKEVGPDLIRHNDTKEASLVKMWTEVESHYYDPAVTPIIYEYFVAPFQGQEPNKSVIESNIEKLKIV
LDVYETKLSTTKYLAGDFYSLADLSHISSTHYFMQTPCASMINERIHVKAWWEDISSRTAFQKVVGGMTF
GQSDQK

Sigma Class GSTs

>Ascaris suum(P46436)

MPQYKLTYFDIRGLGEGARLIFHQAGVKFEDNRLKREDWPALKPKTPFGQLPLLEVDGEVLAQSAAIYRY
LGRQFGLAGKTPMEEAQVDSIFDQFKDFMAELRPCFRVLAGFEEGDKEKVLKEVAVPARDKHLPLLEKFL
AKSGSEYMVGKSVTWADLVITDSLASWESLIPDFLSGHLQLKKYIEHVRELPNIKKWIAERPKTPY

>Homo sapiens(O60760)

PNYKLTYFNMRGRAEIIRYIFAYLDIQYEDHRIEQADWPEIKSTLPFGKIPILEVDGLTLHQSLAIARYLTKNT
DLAGNTEMEQCHVDAIVDTLDDFMSCFPWAEKKQDVKEQMFNELLTYNAPHLMQDLDTYLGGREWL
IGNSVTWADFYWEICSTTLLVFKPDLLDNHPRLVTLRKKVQAIPAVANWIKRRPQTKL

>Hyriopsis cumingii(A0A023I760)

MRTYRLTYSDIRGRAELARLVFVAAGESFEDRRVSREEWAELKKETPFGQIPVLEVDGKPLAQSYAIARFL
GREFGLAGSSNWESAQIDQVMDLVEDLRRELLKVIFEKEPDKKKDLEQKLQDEVFPKFIVFFQKLLENTG
GQYFVGSSLSLADLAVLDLFDTPLQMYPSLLDKSPALQAHRKLLESSPKLDEYLKSRKKTDI

>Danaus plexippus(A0A212FJP1)

MENVKVIYFPLKGMAEGIRLILAHNGQEFEDVRIPHEEWPAMKPNTPFGQLPILEINGKQYAQTSAIVRY
LGRKYGLGGNNVDEDFEIDQNMEFFNDLRSKGSALFYEKDEKRKAALREDLQKNYYPTALAKLNDIIAQN
NGHLALGRLTWADFMFAGIYDSMKYIMQIPDIDEKYPMLVELQQKVLSLPRVKEFCDRAPKSDF

>Bombyx mori(Q5CCJ4)

MPNVKFYYFPVKALGESQRLLLAYGGQEFEDNRISSENWPEFKPKTPFGQMPVLEIDGKQYAQSTAICRY
LGRKYGLAGANDEEAFEIDQNVEFLNDIRASAASVHYEKDEAVKAKKKAELEETKYPFFFEKLNEILTKNN
GHIALGKLTWGDFVYAGMYDYLKAMLQKPDLEQKYPAFRKPIEAVLAIPKVKAYVDAAPRTEL

>Haliotis discus discus(B6RB00)

MPSYKLIYTDNKGRAEVSRLLFALAGQEYEDVRWTRETFQTEKQNLLFGQIPVLEVDGKQYAQSMAIAG
FLAREFGFHGKTSVEQMEVDQVIGIINDIFSALIKQFHEQDEEKKADIIKQNNETTFPKFISFFEEILKNNNT
GFYVGDKLGFADVAAFDTLSKIEAGINMDDFPLVKANKEKVASNERVAKWLQDRPVTQF

>Haliotis discus discus(B6RB01)

MPTYKLRYFNARGFGEVSRLLFALAGQEYEDVRFTQETWPAEKPNTPLGQMPVLDVDGQSFGQSSAISR
FLARRFNFYGQGDVQALQVDQVLGIIQDIINALIKAYYEKDEERKAQALKENKEEKMPLYFGMLEKLLERN
GSTGFFVGNSITLADVSVFDIYDKAKPMVDLDKFPLVKKSVDNVASNPKIKTWIEKRPQTEN

>Haliotis discus discus(B6RB02)

MPTYRFRYFDYKAVGELSRLLFALAGQEYEDVRITYETWPAEKPNTPLGQVPVLEIDGKPFSQSSAISRYLA
RTFGFYGNGDLEALAVDQVLGVVQDVNTFMRDYHKEQDEAKKAELLKEAKDVKIPLYFGMFEKLLKKN
GSTGLFVGKKISIADVSLFDICDKTTDAMLKIEDYPLVKKCCDNVAANPKIKAWVEKRPVTAF

>Aedes aegypti(Q16P79)

MPDYKVYYFNVKALGEPLRFLLSYGNLPFDDIRITREEWPALKPSMPMGQMPVLSVDGKKVHQSVAMS
RYLAKQVGLAGADDWENLMIDTVVDTINDFRLKIAVVSYEPDDDVKEKKLVTLNSEVIPFYLEKLDDIARD
NNGHMANGKLTWADMYFVAILDYLNYMTKSDLVANHPNLQRVVDNVTSIDSIKAWIDKRPQTEI

>Plutella xylostella(204AA)

MPVVKFYYFPIKALGEGPRLLLAYGGQEFQDIRVDKESWPEFKPKTKYGQMPILEIDGKQYAQSAAICRYL
ASRYGLTGADAEQNFEIDEAVDFFNDIRAKAAQVHYEEDEKVKEKRHETYSQTVYPDLLGKLHDIVQRNN
GHLAANKLTWADFYFAGVYDYMKVMLRRPDLDQQYPGFAKVYETVYSLPKVKAFADAAPKTDF

>Drosophila melanogaster(P41043)

MADEAQAPPAEGAPPAEGEAPPPAEGAEGAVEGGEAAPPAEPAEPIKHSYTLFYFNVKALAEPLRYLFAY
GNQEYEDVRVTRDEWPALKPTMPMGQMPVLEVDGKRVHQSISMARFLAKTVGLCGATPWEDLQIDI
VVDTINDFRLKIAVVSYEPEDEIKEKKLVTLNAEVIPFYLEKLEQTVKDNDGHLALGKLTWADVYFAGITDY
MNYMVKRDLLEPYPALRGVVDAVNALEPIKAWIEKRPVTEV

>Anopheles gambiae(P46428)

MPDYKVYYFNVKALGEPLRFLLSYGNLPFDDVRITREEWPALKPTMPMGQMPVLEVDGKKVHQSVAM
SRYLANQVGLAGADDWENLMIDTVVDTVNDFRLKIAIVAYEPDDMVKEKKMVTLNNEVIPFYLTKLNVI
AKENNGHLVLGKPTWADVYFAGILDYLNYLTKTNLLENFPNLQEVVQKVLDNENVKAYIAKRPITEV

>Papilio polytes(I4DMZ0)

MAKVKVTYFNVKALGEGIRMLLAYGGQEFEDIRVERDSWPELKPKTPFGQLPMLEIDGKQYAQSIAICRY
LGRKYGLAGTTPEEDLIIDQNLDFFNDIRLKTVAANYESDEKVKADKLEDLKKNHFPVLFSKLDQMIKENN
GFLAVGRLTWADFVFAGVYDALRMFTHLPDLDEKYPSFSKLRDTVRSLPKVKEFCDSAPKTDL

>Caenorhabditis elegans(Q09596)

VSYKLTYFNGRGAGEVSRQIFAYAGQQYEDNRVTQEQWPALKETCAAPFGQLPFLEVDGKKLAQSHAIA
RFLAREFKLNGKTAWEEAQVNSLADQYKDYSSEARPYFYAVMGFGPGDVETLKKDIFLPAFEKFYGFLVN
FLKASGSGFLVGDSLTWIDLAIAQHSADLIAKGGDFSKFPELKAHAEKIQAIPQIKKWIETRPVTPF

>Nototodarus sloanii(P46088)

PKYTLHYFPLMGRAELCRFVLAAHGEEFTDRVVEMADWPNLKATMYSNAMPVLDIDGTKMSQSMCIA
RHLAREFGLDGKTSLEKYRVDEITETLQDIFNDVVKIKFAPEAAKEAVQQNYEKSCKRLAPFLEGLLVSNGG
GDGFFVGNSMTLADLHCYVALEVPLKHTPELLKDCPKIVALRKRVAECPKIAAYLKKRPVRDF

>Danaus plexippus(A0A212FJP7)

DNVKVLYFPLKAMAEGIRLILAYVGQDFEYVRISEEEWPSVKPNTPFGQVPVIEINGKRHAQTSSILRYLGK
KHGLGGNNLEEDFEIDQVVDFFNDLRLRAASLHYEKDENKKAVLKQELYNNYFPEMFTRLNDIITRNNGY
MAVGKLTWADFMFAGMYERIKVMLAMPDLDEKYPKFKKLEQTVLNLPKVKEYCANEPEYN

>Manduca sexta(P46429)

MPKVVFHYFGAKGWARPTMLLAYGGQEFEDHRVEYEQWPEFKPNTPFGQMPVLEIDGKKYAQSLAIS
RYLGRKYGLAGNDIEEDFEIDQIVDFVNDIRASAASVEYEQDAANKEVKHEENMKNKYPFQLNKLSEIITK
NNGFLALGRLTWADFVFVGMFDYLKKMLRMPDLEEQYPIFKKPIETVLSNPKLKAYLDSAPKKEF

>Fasciola hepatica(Q06A71)

MDKQHFKLWYFQFRGRAEPIRLLLTCAGVKFEDYQFTMDQWPTIKPTLPGGRVPLLDVTGPDGKLRRY
QESMAIARLLARQFKMMGETDEEYYLIERIIGECEDLYREVYTIFRTPQGEKEAKIKEFKENNGPTLLKLVSE
SLESSGGKHVAGNRITLGDLFLFTTLTHVMETVPGFLEQKFPKLHEFHKSLPTSCSRLSEYLKKRAKTPF

>Crassostrea gigas(Q5K4L8)

MASYRLHYFDVRGRGEIVRMLFKLAQAEFGDIRVTQGEWTDVKHDTPTGELPYLEVGEKQLTQSLTIARY
LAREFGLAGDTNWERALVEQVVDTCDDLRAENAKIIHERDPVRLALMKSKMKDQILPKYLNRLTKFLNE
HGDRYFIGSKITSADIAVHEVLTTFLQNDPSCLDKHDVLRKHRQLVEHHPNLSEYLSSRPRFVV

>Spodoptera littoralis(A0A3G1ZLC8)

MPKYVFHYFPIKALGESVRLLLAYGGEGFEDHRIDLDDWPKFKPNTPFGQMPVIEFDGKQYAQSIAIARY
LGNKYGLTGDTLEDNLEIDQNVYLINDLRIKAASAHYEKDEVIKEQKYKELSKGAFPDSLEMLNALFAKNN
GHVALGKLTWADFMFAGLFDYLSAMMRMPDLGQKYPALQQVKDRVYSLPKVKAYADAVPA

>Spodoptera littoralis(A0A3G1ZLD6)

MAKKLHYFNFIAIAEPIRYILHYTKQEFEDVRHDHRFWPNAEFKEKLPFGQFPLYEEGDRMLTQSLAIAKY
VARGTDLIPSDPWTQAVLDAAVYTIYDYWSKVVTFIREKDPEKKKELKRELLDETIDFFFSRLDKDLKENGG
YFSGKLSWVEFVLCGLVEASNYFLDTELEKKYPRVEALIKKIKSLPGVKEYVAARGPNVFKQ

>Haliotis fulgens(A0A346QRM4)

MPTYKLRYFNARGFGEVSRLLFALAGQEYEDVRFTQETWPAEKPNTPLGQMPVLDVDGQSFGQSSAISR
FLARRFNFYGQGDIQALQVDQVLGIIQDIINSMIKAFYEKDEERKAQFMKENKEEKLPLYFGMLEKLLERN
GSTCFFVGNSITLADVSVFDIYDKVKSMVDLDKFPLVKKSIDNVASNPKVKTWIEKRPQTEN

>Argopecten irradians(A0A173DQE4)

MPSYKLIYFTVRGRGELIRLAFAASGQSYDEEKVTFETWPALKPKMPTKQLPVLEVDGKQLTQSLAIARYL
GREFGLAGEGNMDQFLVDQVIDTGADALTAYVKWYFEKEETKKAELKKELVDTTIPKFAEILTNYLENSG
GKNGFFVGSKLSLADLACHETFTDFLQLNPDCLKDYPKLAANRQKVEENANVKQYLSSRPESVI

>Ruditapes philippinarum(H6B8N8)

MATYKVSYFPWKGNGEIIRLVLVIAEQDFKDERLTMDEWLKVKEASPTKHMPLLTVNGTVYGQTAACA
RYLARKYGLMGSTPEEELLIDEVYECIVDFLKEVFKLTYEKDDKTKEELKQKIITENVPKLNDYIKLRSKLGGN
GFIIGKKISLADIHLYNIVDQCEASFPGFFSSAPDIKKHADVVKSDARIQKWIATMPKMPQ

>Azumapecten farreri(I6LKU6)

MPSYKLIYFGVRGRGELIRLAFAASGQTFEEDTITFADWPELKQKMPTGQIPVLEIDGKQLSQSLAIARYLG
REFGLAGKTNMDQCLVDQVIDTAGDCLTEYVKSHFESDETKKAELRKTLVETTIPKFAKIFTTFLENSGGK
NGFFVGSELTLADLACHEAFTDFLQLNADALKDYPQLAANRQKVEENENVKRYLAKRKESPI

>Laternula elliptica(B9VX80)

MAGTVQGDKWVLYYWPGFKGRAEFVRLVFEEAGIPYLESNQGVADSIIKGEIGGYPVMMPPVVAKGDF
RLGQTQMICQYLAGKYGLAPKGEEDKIHAEQVCASMYDYLTEGYGAFHGAKPGVKYADQKEEAQRYID
RVVQQRLPRYLKHFETVLAANTAGTGFLFGDSISHADLALFHIMNATEFQFPEVYKSADYIPLLKAHRDRI
ASRPNIVAYTQSERCKPFSGDSFM

>Spodoptera littoralis(A0A3G1ZLD1)

MPKYVFHYFDGKGLGEPVRLLLAFGDEGFEDHRVAFKDWPDFKPKTPFGQMPLLEFDGKQYAQSLSIAR
YLGKKYGLAGESLEDALEIDQNVDLINDLRAKAAIASYEKDEAVKEKKYAEFNKDVFPNMLEKLNEIITKNN
GHIAIGKLSWGDFVFAGMFDYIKHLLVVPDLEKKYPAFQKVVDAVYSIPKVKTYADAVGPTEF

Tau Class GSTs

>Arabidopsis thaliana(Q9ZRW8)

MANEVILLDFWPSMFGMRTRIALREKGVEFEYREEDLRNKSPLLLQMNPIHKKIPVLIHNGKPVNESIIQV
QYIDEVWSHKNPILPSDPYLRAQARFWADFIDKKLYDAQRKVWATKGEEQEAGKKDFIEILKTLESELGD
KPYFSGDDFGYVDIALIGFYTWFPAYEKFANFSIESEVPKLIAWVKKCLQRESVAKSLPDPEKVTEFVSELRK
KFVPE

>Populus trichocarpa(D2WL48)

MEDRVTLLDFWPSPWATRVKVALAEKGIEYESREQNLIDKSPLLLEMNPVHKTIPVLIHNGKPICESHNIV
QYIDEVWKDKSPLLPSDPYQRSQARFWADYIDKKIYNNAKKLWKEKGEEQEEVKREFIEGLKTLEGELGD
KLYFGGESFGFVDVVLVPVTSWFYSLEICGKFSIEAECPRFTAWIKRCMEKESVSSSLPDPHKVYDYVLLLK
KKMGIE

>Populus trichocarpa(D2WL57)

MAEVKLLGAWGSPFSRRVEMALKLKGVEYEYIDEDLANKSPLLLKYNPIHKKVPVLLHNGKTMAESLVILE
YIDETWKSNPILPEDPYDKAMARFWAKFIDEKCMPAIWQIMLSKENEREKAIEEAIQHLKTLENELKDKKF
FGGETIGLVDIVANFIGFWLGAAQEATGMELVNKERFPVLCKWIDEYANCSVVKENLPPRDKLIAFLRPRL
SASSWKY

>Glycine max(Q9FQE8)

MTDEVVLLDFWPSPFGMRVRIALAEKGIEYEYKEEDLRNKSPLLLQMNPVHKKIPVLIHNGKPISESLIAV
QYIEEVWNDRNPLLPSDPYQRAQARFWADYVDIKIHDLGKKIWTSKGEEKEAAKKEFIEALKLLEEQLGD

KTYFGGDNIGFVDIALVPFYTWFKVYETFGSLNIENECPRFVAWAKRCLQKESVAKSLPDQHKVYEFVVEI
RKKLVIE

>Pinus tabuliformis(Q6DNI8)

MENQVKVLNLWASPFGLRVLVGLEEKGVKYEYQEENLPSKSELLLKMNPIHKKIPVLIHNDKPVLESLIIVE
YIDEAWPNTNPFMPSSAYERARARFWADFVDKKIYDNGSALIMKCKGEAQEEAKRNMLEYLGLLEGAL
DELSGGIKPYFGGEKFGYMDIAFIPFASWFQAWEVMGNWKIPLETQFPRLHEWVNACMERESVKKVLP
HPEKVAEFAMQIRQRFVGSD

>Capsella rubella(R0GRU5)

DEVILLDFWPSMFGMRTRIALEEKNVKFDYREQDLWNKSPILLEMNPVHKKIPVLIHNGKPVCESLIQVEY
IDEVWPSKNPLLPSDPYQRAQAKFWGDFIDKKVYASARLIWGAKGEEQEAGKKEFIEILKTLESELGDKIYF
GGETFGYVDIALIGFYSWFEAYDKFGNFSIEAECPKLIAWAKRCVERESVAKSLPDSNKIVEF

>Medicago truncatula(G7L788)

QDVKLLNFLLSPVGRRVEWALKLKGVEFDYIEEDIFNKSSLLLEMNPVHKKVPVLVHGQKSIAESLIILEYID
ETWKQYPLLPPDPYQRSRARFWAKLSDEKLALGSWIALIKKGNEWEKALKEAREIMEKLEEDIKGKKFFG
GDTIGYLDLTLGWITCFLPIWEEIGSTQILDPLKCPSISSWKINFLSHPIIKECLPPRDEMILYCHRRIKEY

>Glycine soja(K9M8M1)

EEVILLGKWASPFSNRVDLALKLKGVPYKYSEEDLANKSADLLRYNPVHKKVPVLVHNGNPLPESLIIVEYI
DETWKNNPLLPRDPYERALARFWSKTLDDKILPAIWNACWSDENGREKAVEEALEALKILQEALKDKKFF
GGESIGLVDIAANFIGYWVAILQEIAGLELLTIEKFPKLYKWSQEFINHPVIKEGLPPRDELFAFF

>Arabidopsis thaliana(Q9FUS8)

MASSDVKLIGAWASPFVMRPRIALNLKSVPYEFLQETFGSKSELLLKSNPVHKKIPVLLHADKPVSESNIIVE
YIDDTWSSSGPSILPSDPYDRAMARFWAAYIDEKWFVALRGFLKAGGEEEKKAVIAQLEEGNAFLEKAFI
DCSKGKPFFNGDNIGYLDIALGCFLAWLRVTELAVSYKILDEAKTPSLSKWAENFCNDPAVKPVMPETAK
LAEFAKKIFPKPQA

>Arabidopsis thaliana(Q8L7C9)

MANLPILLDYWPSMFGMRARVALREKGVEFEYREEDFSNKSPLLLQSNPIHKKIPVLVHNGKPVCESLNV
VQYVDEAWPEKNPFFPSDPYGRAQARFWADFVDKKFTDAQFKVWGKKGEEQEAGKKEFIEAVKILESE
LGDKPYFGGDSFGYVDISLITFSSWFQAYEKFGNFSIESESPKLIAWAKRCMEKESVSKSLPDSEKIVAYAA
EYRKNNL

>Arabidopsis thaliana(Q9FUS6)

MAQNDTVKLIGSWSSPYSLRARVALHLKSVKYEYLDEPDVLKEKSELLLKSNPIHKKVPVLLHGDLSISESLN
VVQYVDEAWPSVPSILPSDAYDRASARFWAQYIDDKCFAAVDAVVGAKDDEGKMAAVGKLMECLAILE
ETFQKSSKGLGFFGGETIGYLDIACSALLGPISVIEAFSGVKFLRQETTPGLIKWAERFRAHEAVKPYMPTV
EEVVAFAKQKFNVQ

>Glycine max(Q9FQE7)

MASSQEEVTLLGVVGSPFLHRVQIALKLKGVEYKYLEDDLNNKSDLLLKYNPVYKMIPVLVHNEKPISESLV
IVEYIDDTWKNNPILPSDPYQRALARFWAKFIDDKCVVPAWKSAFMTDEKEKEKAKEELFEALSFLENELK
GKFFGGEEFGFVDIAAVLIPIIQEIAGLQLFTSEKFPKLSKWSQDFHNHPVVNEVMPPKDQLFAYFKARAQ
SFVAKRKN

>Glycine max(Q9FQD8)

MGSEEVKLLSFWVSPFGKRIEWALKLKGVEYEYIEEDIFNKSSLLLELNPVHKKVPVLVHAEKSIIAESFIILEY
IDEKWKQYSLLPHHPYQRALARFWAATAEEMFRKVVWIALRSPTSGDEREKALKESREVMERIEEEIRGK
KYFGGDNIGYLDIALGWISYWLPVLEEVGSMQIIDPLKFPATTAWMTNFLSNPVIKDNLPPRDKMLVYLK
DLRSKYIVL

>Glycine max(I1MJ34)

MADEVVLLDFWPSPFGMRVRIALAEKGIKYEYKEEDLRNKSPLLLQMNPVHKKIPVLIHNGKPICESLIAV
QYIEEVWNDRNPLLPSDPYQRAQARFWADYVDKKIYDLGRKIWTSKGEEKEAAKKEFIEALKLLEEQLGD
KTYFGGDNLGFVDIALVPFYTWFKAYETFGTLNIESECPKFIAWAKRCLQKESVAKSLPDQQKVYEFIMDL
RKKLGIE

>Capsella rubella(B2BXV3)

MAERSEDSEEVKLLGMWASPFSRRIEIALTLKGVPYEFSEQDITNKSDLLLRLNPVYKMIPVLVHNGKPISE
SLVILEYIDDTWRNNPILPQDPSERAMARFWAKLIDQQICVAAMKVAGTIGEERDAAVEETRNLLMFLEK
ELVGKDFFGGQSLGLVDIVATLVAFWLIRTEDLLGVKVVPVEKFPEIHRWVKNLSGNDVIKKCIPPEDEHL
EYIRARMDKLNIKSA

>Pinus brutia(D3YLT8)

MENQVKVLNLWASPFGLRVLVGLEEKGVKYEYQEENLASKSELLLKMNPIHKKIPVLIHNDKPVLESLIIVE
YIDEAWPNTNPFMPSSAYERARARFWADFVDKKIYDNGGALIMKCKGEAQEEAKRNMMEYLGLLEGA
LDELSGGMKPYFGGEKFGYMDIAFIPFASWFQAWEVMGNWKIPLETQFPRLHEWVNACMERESVKKV
LPHPEKVAEFAMQMRQRFVGSD

>Glycine soja(I3NNW1)

MGSEEVKLLSFFASPFGKRVEWALKLKGVEYEYIEQDIFNKTSLLLQLNPVHKKVPVLVHAHKPIAESFVIV
EYVDETWKQYPLLPQDPYQRALARFWANFAEQKLLDAAWIGMYSSGDEQQNAVKVAREAIEKIEEEIK
GKKYFGGENIGYLDIALGWISYWLPIWEEVGSIQIIDPLKFPAITAWITNFLSHPVIKDNLPPRDKMLVYFH
SRRTALSSTFQG

>Populus trichocarpa(D2X9R3)

MADVKLHGSWVSPFNYRVIWALKLKGVEFEHIVEDLTNKSELLLKYNPVYKKIPVLVHGGKPIAESLVILEY
IEETWPENPLLPTDPYERAMARFWIQYGATKTAAFGALFRASGEELEKAAKEVVEVLRVLEEQGLGDKKF
FGGDSINLVDISFGLFTCWLEAIEEAAGVKVLEPSTLPRLHAWAQNFIEVPLIKENIPDYDKLLLHMKGVRE
KMMNK

>Medicago truncatula(G7L6H8)

MAAIDEEVQLLGSVGSPFVIRVQIALKLKGIEHKYVEEKLGNLSETLLKYNPVYRMVPVLVHNGNPISESRV
IIEYIDEAWKQNPILPSDPYKRALDRFWSKFIDEKCTIAAWKSVFMPDEKEREKAREELFEALQFLENELKD

KFFGGEEIGFVDIAALFIPLFQEVAEKQLFPGDKFPKLQKWSLDFYNHPVVKEVMPSKEQQFGYFKARAA
SLAAPSK

>Pinus tabuliformis(L7S263)

MEGCGEDGQVKLLGVIHSPFVVRVRIALALKGIHYEFIEEEVLNNKSELLLQSNPVHKKIPVLIHNGKPVCE
SMIIVQYIEEAWGNKAFNFMPKDPYDRAIARFWAAFIDDKLTPSIWGVFNGVGEEQQKAVEESVANLLL
LEEALSGKAYFGGDEIGLVDIALGGLLVGVQTIERVTGSVLIDTLKMPLLSTWAHKFCKAEEVKEVLTDPAK
LFELLSEIRANLTSPPAGN

>Pinus tabuliformis(L7S255)

MATEGEKGQVKLLGATLSPFVVRVRIALALKGIDYEFIQESMHPKSELLVKSNPVHKKIPVLIHNGKPVCES
MIIVQYIEEGWGNKAPNLMPEDPYDRAIARFWAAFVDDKLFPCLRGVFTGQGEQLQKAVEDSVTNFLLI
EEALRTNHCFAGKAYFGGDQIGLIDIALGGLSAFIKGLEKATDSVLIDPEKMPLLSAWMDRFFKSDGVKEV
MPDVTKQVEFISTRRASMISPPTSN

Theta Class GSTs

>Alitta succinea(A6MN06)

MSRLKLYFDLMSQPSRAVWIFLKATGIPFEEKPVALRKGEHQTEEFAKINPFQLVPVIDDGGFVLYESLSIC
KYLAKSRNLADHWYPSELKHRARIESYLQWHCLMVRLFASQVFRIQVIEPRAMNKPVDRQRLAKYENILS
VVLDSFETVWLKDTPYICSNEISIADVACICELMQVYAVDYPLWEDRPKLEAWSKRVRERLNPHYDQANF
MVDKVRNHFLASKI

>Homo sapiens(P0CG29)

MGLELFLDLVSQPSRAVYIFAKKNGIPLELRTVDLVKGQHKSKEFLQINSLGKLPTLKDGDFILTESSAILIYLS
CKYQTPDHWYPSDLQARARVHEYLGWHADCIRGTFGIPLWVQVLGPLIGVQVPKEKVERNRTAMDQA
LQWLEDKFLGDRPFLAGQQVTLADLMALEELMQPVALGYELFEGRPRLAAWRGRVEAFLGAELCQEAH
SIILSILEQAAKKTLPTPSPEAYQAMLLRIARIP

>Homo sapiens(P30711)

MGLELYLDLLSQPCRAVYIFAKKNDIPFELRIVDLIKGQHLSDAFAQVNPLKKVPALKDGDFTLTESVAILLY
LTRKYKVPDYWYPQDLQARARVDEYLAWQHTTLRRSCLRALWHKVMFPVFLGEPVSPQTLAATLAELD
VTLQLLEDKFLQNKAFLTGPHISLADLVAITELMHPVGAGCQVFEGRPKLATWRQRVEAAVGEDLFQEA
HEVILKAKDFPPADPTIKQKLMPWVLAMIR

>Bos taurus(Q2NL00)

MGLELYLDLLSQPCRAIYIFAKKNRIPFELRTVDLRKGQHLSDAFAQVNPLQKVPILKDGDFILTESVAILLYL
ARKYKVPDHWYPQDLQACARVDEYLAWQHTALRRNCLRALWHKVMLPVFLGEPVSPEMLATTLAELD
MALQVLEGKFLQDKAFLTGSHISLADLVAITELMHPVGAGCQVFKGRPKLAAWRQRVEAAVGEVLFQE
AHEVILKAKDSQPADPTLKQKMLPKVLAMIQ

>Rattus norvegicus(Q4V8E6)

MGLELYMDLLSAPCRAVYIFARKNGIPFDFQFVDLLKGHHHSKEYIEINPLRKVPSLRDGKFILSESVAILCY
LCRKYSAPSHWYPPDLHMRARVDEFMAWQHTAIQVPMSKILWIKLIIPMITGEEVPTERLDKTLDEVNK

NIKQFEEKFLQDKLFITGDHISLADLVALVEMMQPMGTNHNVFISSKLAEWRMRVELAIGSGLFWEAHD
RLVKLPSWDCSTLDPSIKMKICEFLQKYK

>Mus musculus(Q64471)

MVLELYLDLLSQPCRAIYIFAKKNNIPFQMHTVELRKGEHLSDAFARVNPMKRVPAMMDGGFTLCESVA
ILLYLAHKYKVPDHWYPQDLQARARVDEYLAWQHTGLRRSCLRALWHKVMFPVFLGEQIPPETLAATLA
ELDVNLQVLEDKFLQDKDFLVGPHISLADLVAITELMHPVGGGCPVFEGHPRLAAWYQRVEAAVGKDLF
REAHEVILKVKDCPPADLIIKQKLMPRVLAMIQ

>Arabidopsis thaliana(Q9ZRT5)

MMKLKVYADRMSQPSRAVIIFCKVNGIQFDEVLISLAKRQQLSPEFKDINPLGKVPAIVDGRLKLFESHAIL
IYLSSAFPSVADHWYPNDLSKRAKIHSVLDWHHTNLRRGAAGYVLNSVLGPALGLPLNPKAAAEAEQLLT
KSLSTLETFWLKGNAKFLLGSNQPSIADLSLVCELMQLQVLDDKDRLRLLSTHKKVEQWIENTKKATMPH
FDETHEILFKVKEGFQKRREMGTLSKPGLQSKI

>Erythranthe guttata(A0A022Q7B5)

MELKVYADRMSQPSRAILIFCKANGLEFEEVPIQLAKKQHHSPEFAEINPMKQVPAIVHGDFKLFESHAILI
YLASAFPGVADHWYPADARKRAKIHSVLDWHHSNLRRGSVGYIFNNTIALAFGLPLNPKAEAEGEKLLLA
SLSTIESLWLEDGPFLLGNSKPSIADLSLVCEITQLEFADEKDRERILSPHKKVLKWMEDTKKASAPYFEEIH
SSMPPGIAQLKALKAQVISQLSASKH

>Macaca fascicularis(A0A023JCC0)

MGLELYLDLLSQPCRAVYIFAKKNGIPFELRIVDLIKGQHLSDAFAQVNPLKKVPALKDGDFTLTESVAILLY
LTRKYKVPDYWYPQDLQARARVDEYLAWQHTTLRRSCLRALWHKVMFPVFLDEPVSPQTLAATLAELD
VNLQLLEDKFLQNKAFLTGPHISLADLVAITELMHPVGAGCQVFESRPKLATWRQRVEAAVGEDLFREA
HEVILKAKDFPPADPTIKQKLMPWVLAMIR

>Mus musculus(Q61133)

MGLELYLDLLSQPSRAVYIFAKKNGIPFQTRTVDILKGQHMSEQFSQVNCLNKVPVLKDGSFVLTESTAILI
YLSSKYQVADHWYPADLQARAQVHEYLGWHADNIRGTFGVLLWTKVLGPLIGVQVPQEKVERNRDRM
VLVLQQLEDKFLRDRAFLVGQQVTLADLMSLEELMQPVALGYNLFEGRPQLTAWRERVEAFLGAELCQE
AHSTILSILGQAAKKMLPVPPPEVHASMQLRIARIP

>Rattus norvegicus(Q01579)

MVLELYLDLLSQPCRAIYIFAKKNNIPFQMHTVELRKGEHLSDAFAQVNPMKKVPAMKDGGFTLCESVAI
LLYLAHKYKVPDHWYPQDLQARARVDEYLAWQHTTLRRSCLRTLWHKVMFPVFLGEQIRPEMLAATLA
DLDVNVQVLEDQFLQDKDFLVGPHISLADVVAITELMHPVGGGCPVFEGRPRLAAWYRRVEAAVGKDL
FLEAHEVILKVRDCPPADPVIKQKLMPRVLTMIQ

>Rattus norvegicus(P30713)

MGLELYLDLLSQPSRAVYIFAKKNGIPFQLRTVDLLKGQHLSEQFSQVNCLKKVPVLKDGSFVLTESTAILIY
LSSKYQVADHWYPADLQARAQVHEYLGWHADNIRGTFGVLLWTKVLGPLIGVQVPEEKVERNRNSMV
LALQRLEDKFLRDRAFIAGQQVTLADLMSLEELIQPVALGCNLFEGRPQLTAWRERVEAFLGAELCQEAH
NPIMSVLGQAAKKTLPVPPPEAHASMMLRIARIP

>Homo sapiens(P0CG30)

MGLELFLDLVSQPSRAVYIFAKKNGIPLELRTVDLVKGQHKSKEFLQINSLGKLPTLKDGDFILTESSAILIYLS
CKYQTPDHWYPSDLQARARVHEYLGWHADCIRGTFGIPLWVQVLGPLIGVQVPEEKVERNRTAMDQA
LQWLEDKFLGDRPFLAGQQVTLADLMALEELMQPVALGYELFEGRPRLAAWRGRVEAFLGAELCQEAH
SIILSILEQAAKKTLPTPSPEAYQAMLLRIARIP

>Gallus gallus(P20135)

MGLELYLDLLSQPCRSIYIFARTNNIPFEFKHVELFKDSVLGKKPAAASGAERPRTGPSNSEGDGKISLLKKV
PVLKDGDFTLAECTAILLYLSRKYNTPDHWYPSDIKKRAQVDEYLSWHHANIRANAPKTMWIKVLIPLFT
GQPQPSEKLQEVMEGLSTSLKQFEERFLQDKAFIIGSEISLADLVAIVELMQPVGVGCDIFEDRPRLMEW
RRRVEEAVGKELFFQAHEMILNIKELSNIQIDPQLKEHLAPVLMKMLK

>Lucilia cuprina(P42860)

MDFYYLPGSAPCRSVLMTAKALGIELNKKLLNLQAGEHLKPEFLKINPQHTIPTLVDGDFALWESRAIMVY
LVEKYGKNDSLFPKCPKKRAVINQRLYFDMGTLYKSFADYYYPQIFAKAPADPELYKKMEAAFDFLNTFLE
GHQYVAGDSLTVADLALLASVSTFEVAGFDFSKYANVAKWYANAKTVAPGFDENWEGCLEFKKFFN

>Homo sapiens(P0CG30)

MGLELFLDLVSQPSRAVYIFAKKNGIPLELRTVDLVKGQHKSKEFLQINSLGKLPTLKDGDFILTESSAILIYLS
CKYQTPDHWYPSDLQARARVHEYLGWHADCIRGTFGIPLWVQVLGPLIGVQVPEEKVERNRTAMDQA
LQWLEDKFLGDRPFLAGQQVTLADLMALEELMQPVALGYELFEGRPRLAAWRGRVEAFLGAELCQEAH
SIILSILEQAAKKTLPTPSPEAYQAMLLRIARIP

>Nasonia vitripennis(K7ITL4)

MSVKFYMDLMSQPSRALYIFMKTTNIPFEKKVTSLKNGENYKDGFEKISPFNKLPVIQHNGFNLTESVAIV
RYLAREFNVEDHWYPKDSKAQAKVDEYLEWQHLNTRLHCASYFAVKFLWPIIKGQHIEPKTVVEHEARM
IECLDQIENIWLKDNKPFLVGDRITVADLFGACEIEQPRVGGFNPREGRPVLTAWLDRVAKETAPYYEEA
HSPMNKVTERNAKQSKL

>Danaus plexippus(A0A212EZ10)

MTLKLYYDLMSQPSRSLYILLKVSKCDFEPKFIDLRKGEHYSNEYAKINRFQKVPVIDHNGFVLSESVAILRY
LSGENLIPPSLYPKDNQIRARVDEFLEWHHIEFRLHLSMYFRVKHMDPIITGIPPNPKTLKGYERRLISALET
FESQWLKNGNEFITGNNITVADLFAASELEQPRMAGYNPAERFLRIGSWWKNVREHFSPYYDEGHVILN
KIVNKQKQQSSKL

>Aedes aegypti(Q17DF6)

MPRPVKFFYDLLSTHSRALYMFFEATKIPYDPIPVCATKGEHLTDEYRECVNRFQQVPSIIDDGFKLSNGVT
ILKYLIREKLIPEHWYPRDSQLRAKIDEYLEWQHDNSSKVCNAFVQEKWSLLNIEDERSSEQKVEEYRRQM
EQNLNQLEREWLVPGRFIIGDRITIADILAACEIEQPKIVGMDPFQGRPKLAAWLEKVRYTMTPYYQEAH
QDFYKFTEKASVKN

>Aedes aegypti(Q16X19)

MANGRSIRFYYDLISQPCRALYIFLEQNKIHYQKCPIALRKCEHTTPEYLQNVNRFGKVPAIVDGKNFKLAE
SIAILRYLAREFTVPDHWYPRDSRRRARVDEYLEWQHSNTRLHCAGYVRYVWRGPLRGETMDPRVAKR
LKAEMVGCLDFIETNVLQRDVHFIAGDEISIADLVAACEIEQPKLAGYDARVGRPKLTAWMQRVKETTQP
DYDEAHKVLNKFAPTAT

>Plutella xylostella(217AA)

MSQPSRAVYILLKKSNINFEPKYVDLRKGVHYTDEYSNNINRFKKVPVIDHNGFILTESVAIIRYLGRENVLP
EALFPRADKVLNTRLDEFLEWQHLGLRAPLAMYFRVVMFSPDSEKIPSYQKRMETALDEFSTLWLGRGN
QYILGDTATVADLLAACEVEQPRMTGYDCTANYPVIREWMDRVRSYFNPHYAEASSIVEKIAAKRIPMKK
PTAKL

>Drosophila melanogaster(E1JJS1)

MSVSFLASLLGLSNDEDQLQVAFDEVLKRRVPSRQPTNLRMSAPIRYYYDLMSQPSRALFIIFRLSNMPFE
DCVVALRNGEHLTEDFKKEINRFQRVPCIHDNGYKLAESVAILRYLSAKGKIPEHLYPKYFVDQSRVDEFLE
WQHMSLRLTCAMYFRTVWLEPLLTGRTPSEAKIETFRMQMERNLDVVEEVWLEGKDFLTGSSLTVADI
FAACEIEQTRMADYDVRIKYPKIRAWLKRVRQSCNPYYDVAHEFVYKISGTGPQAKL

>Anopheles gambiae(Q8MUQ2)

MSRSVKLYYDLMSQPSRALYIFLSTNKIPFDRCPIALRKMQHKTDEYRRQVNRYGKVPCIVDGSFRLAESV
AIYRYLCREFPTDGHWYPSDTVRQARVDEYLSWQHLNLRADVSLYFFHVWLNPLLGKEPDAGKTERLRR
RLDGVLNFFDQELLSAGSGQAFLAGDRISIADLSAACEIEQAKIAGYDPCEGRPALASWLTAVRERTNPYY
DEAHKYVYRLSPDHIVTPVVAEDE

>Aedes aegypti(Q5PY76)

SKLRYFYDLMSQPSRMLYIFLESTKIPYERCLVNLGKGEHLTDKFKAINRFQKVPCIVDKNDLHLAESVAIVR
YLAREYPFSDHWYPKDSQKRARIDEYLEWQQHNTRAVCATYFQYVWLRPKLMGTKVNPERAEEYKQK
MEDCLDFIESDYLGGGNPFLVGNEISVADLFAACEIEQPKMAGFDPCVGRPKMTAWMARVREATNPH
YDEAHKLVYRIAPDS

>Drosophila melanogaster(Q7K0B6)

MSKAIKYYYDFLSQPSRALWIAMKLGKTPFEDCPVALRKQEQLTDEYRSINRFQKVPAIVDGKFQLGESVS
IVRYLADKGVFSEQLYPKTLEERARVDEFLEWQHFNVRLVCSLFFRQVWLLPAKGLAPAPKPESVKKLIKD
VESNLGLLERLWLEKDFLVGDKLTVADIFGSSEINQMKLCQYNVNEKQFPKVAKWMERVRDATNPYYD
EAHSFVYKTSQQAVKAKN

>Drosophila melanogaster(A1Z7X7)

MSKPIRFYYDLLSPIARGLWIGLKFSNSPVEYCPIALRKFEQLTDEYKKINRFQKVPAIVGGDFHLSETIAIIRY
LADKGQFDEKLYPKTLENRARVDEFLEWQHLNIRLACSMYFRDAWLFPMNGIAPKPKPEQIQALIEGVE
NNLGLLERLWLENDFLVGKNLTMADILGSSEINQLRLCQYRVDEKKFPKVVKWLERVRVSANPYHDEGL
TFIDRKSKQSTAAKL

>Bombyx mori(B0LB14)

MVLKLYYDLMSQPSRVLYILLKTMKYDFEPKYVNLRKAEHYSEDFTKVNRMQRVPVIDHNGFILTESIAILK
YLSRENVIAESLYSKESKLQARIEEFLEWQHIGLRLHCAMYFRVVHMDPILTGRKSDEKTIQGYKRRMMM

ALDDFDTKWLGRGTAFIVGETPTVADLVAACELEQPRMAGFEPKDHFPNIAAWWPKVRDHFAPHYED
AHVILNKIINKMDRAANSKL

>Anopheles gambiae(Q8MUQ1)

KNLKYYYDLMSQPSRALWIFLEKTKLPYEKCLINLGKGEHLTEEFKAINRFQKVPCITDSQIKLAESVAIFRYL
CREYQVPDHWYPADSRRQALVDEYLEWQHHNTRATCAIYFQYVWLRPRMFGTKVDPKQAEKYRGQM
EGTLDFIEREYLGSGARFIAGDEITVADLLAACEIEQPRMAGYDPCEGRPNLTQWMARVRESTNPYYDQ
AHKLVNKFAQDTAS

Xi Class GSTs

>Natrialba magadii(D3SS28)

ACPWAHRTLVTRTLKGLEDAISVSVVDPYRAEDGWQFTPEKEGCTHDHVHDVDYLRELYVRAAPDVTC
RVTVPVLWDTEEDTIVNNESEEIMRMFDTEFDEFADHTVDLYPEGYQEKVDQIIDNIYEPINNGVYRAGF
ATEQEPYDEAVAELFGALAHWDDVLADQRYLAGDRLTEADIAMFTTLVRFDNVYHTHFMCNVQYIREF
DNLWPYLRDLYQTHGIAETVEMDHITEHYYTTHPDVNP

>Haloferax volcanii(D4GT00)

GRYHLYICRACPWAHRTAMTRALKGLEDAISLSLVEPVRIDDGWEFSEDLPDPLYGEEFLRDIYLRADDEF
TGRVTVPVLWDKQRETIVNNESREIMRMLDEAFDPLAERDVDLNPDGYEEEVDRLVDEIYEPVNNGVYR
AGFATTQEAYEEAVEELFDALDHWDEVLDDQRFLAGDVVTEADIAMFATLIRFDHVYHTHFKCNKKAIH
EYDNLWNYTKELYQLPGVAKTVNMDHIVRHY

>Halobacterium salinarum(Q9HN26)

GRYHLYVSYACPWAHRTLLVRALLGLEDAISVSVVDPVRYDQGWTFDPEKPGCTPDHVFGGTHLRDVYT
EADPEYTGRVTVPVLYDTDADTIVNNESEEIMRMLDVAFDDHAARDVDLYPEGYRDEVDRLIEDIYDPIN
NGVYRAGFADSQRAYDNAVDDLFEALAHYDDVLAEQRYLAGDVLTEADVAMFTTLYRFDEVYHTHFKC
NRKRISDYDNLWPYLRELCQLPGVADTLYMEHVKQHYYRS

>Halorhabdus utahensis(C7NPF2)

GRYHLYVCRACPWAHRTLVTRALKGLEDAITVDYVDPYRGEDGWQFTPEKDGCTPDTVNGSDYLREVY
VEADPDMTGRVTVPVLWDKQEETIVNNESAEIMRMLDTEFDDVAEHDVDLYPEGYQEDIDEIIEAIYEPI
NNGVYKAGFADSQAAYDEAVEELFDALDHWDSVLEDQRYLAGDRLTEADIAMFTTLVRFDEVYHTHFM
CNHKLIAEYDNLWPYLRDVYTTDGVAETVDIDHIKEHYYTTHPDVSP

Zeta Class GSTs

>Homo sapiens(O43708)

MQAGKPILYSYFRSSCSWRVRIALALKGIDYKTVPINLIKDRGQQFSKDFQALNPMKQVPTLKIDGITIHQS
LAIIEYLEEMRPTPRLLPQDPKKRASVRMISDLIAGGIQPLQNLSVLKQVGEEMQLTWAQNAITCGFNAL
EQILQSTAGIYCVGDEVTMADLCLVPQVANAERFKVDLTPYPTISSINKRLLVLEAFQVSHPCRQPDTPTEL
RA

>Mus musculus(Q9WVL0)

GKPILYSYFRSSCSWRVRIALALKGIDYEIVPINLIKDGGQQFTEEFQTLNPMKQVPALKIDGITIVQSLAIM
EYLEETRPIPRLLPQDPQKRAIVRMISDLIASGIQPLQNLSVLKQVGQENQMQWAQKVITSGFNALEKILQ
STAGKYCVGDEVSMADVCLVPQVANAERFKVDLSPYPTISHINKELLALEVFQVSHPRRQPDTPA

>Coccidioides immitis(D2YW48)

MTTPNFELYGYFRSSCSGRLRIAFHLKSIPYTRHPVNLLKGEQHSDTYKSLNPTNTVPLLVVSNINNTVSPS
SASFSIGQSLAALEYLEEALPTNARPLLPPISNPVARAHVRTICNIIACDVQPVTNLKIQKKVKALDGDPTV
WSRDLATQGFGAVEKLLELSAGRFCVGDEITLADVCLVPAVWAAERVGMDLARFPITKRVFEEMLKEEA
VQKAHWQKQEDTPEDLRA

>Dianthus caryophyllus(P28342)

MSSSETQKMQLYSFSLSSCAWRVRIALHLKGLDFEYKAVDLFKGEHLTPEFLKLNPLGYVPVLVHGDIVIA
DSLAIIMYLEEKFPENPLLPQDLQKRALNYQAANIVTSNIQPLQNLAVLNYIEEKLGSDEKLSWAKHHIKKG
FSALEKLLKGHAGKYATGDEVGLADLFLAPQIIASITGFGMDMAEFPLLKSLNDAYLKYQHFRMRCQRISP
MLDEAKS

>Euphorbia esula(P57108)

MASVEQPNKPKLKLYSYFRSSCSFRVRIALNLKGLDYEYVPVNLLKGEQFTPEFLKINPIGYVPALVDGEDVI
SDSFAILMYLEEKYPEHPILPADIHKKAINYQAANIVSSSIQPLQNLAVLNFIGEKVSPDEKVPWVQRHISK
GFAALEKLLQGHAGRFATGDEVYLADLFLEPQIHAAITRFNVDMTQFPLLLRLHEAYSQLPEFQNAMPDK
QPDSTSPTAS

>Dianthus caryophyllus(Q03425)

MSSSETQKMQLYSYSSSSCAWRVRIALHLKGLDFEYKAVDLLKGEHLTPEFLKLNPLGYVPALVHGDIVIA
DSLAIIMYLEEKFPENPLLPRDLQKRALNYQAANIVASNIQPFQNLAVLNYIEEKLGSDEKLSWANHHIKK
GFS

>Aedes albopictus(A0A023EJH2)

YWRSSCSWRVRIALNLKEIPYDIKPISLIKSGGEQHCNEYREVNPMEQVPALQIDGHTLVESLAIMHYLEET
RPQRPLLPQDVLKRAKVREICEVIASGVQPLQNLIVLIHVGEEKKKEWAQHWITRGFRAIEKLLSTSAGKFC
VGDEITLADCCLVPQVFNARRFHVDLRPYPIILRIDRELEGHPAFRAAHPSNQPDCPPEAAK

>Aedes albopictus(A0A023EMR1)

SACQTAIATSATHPPSSGLHLFLSSRWNDTAAKINYSKFHNAMSLSAMSKPILYSYWRSSCSWRVRIALNL
KEIPYDIKPISLIKSGGEQHCNEYREVNPMEQVPALQIDGHTLVESLAIMHYLEETRPQRPLLPQDVLKRAK
VREICEVIASGVQPLQNLIVLIHVGEEKKKEWAQHWITRGFRAIEKLLSTSAGKFCVGDEITLADCCLVPQV
FNARRFHVDLRPYPII

>Triatoma infestans(A0A023FAZ9)

MSLIGKPVLYSYWRSSCSWRVRIALNLKEIPYDIKPISLIKGGGEQHSNEFREINPMEHVPALQIDGHTLIES
LSIMYYLEETRPQRPLLPQDVYKRAKVREICDVIASGIQPLQNIGVLIYVGEEKKKEWAQHWITRGFRAVE
KLLSSSAGKYSVGDELTLADCCLIPQVFNARRFQVDLRPYPIILRIDRELENHPAFRAAHPSNQPDCPPEAT
K

>Amblyomma cajennense(A0A023FJC9)

MSKPVLLSNYLSSCAWRVRIVLEVKKIPYEYRTVNLKPVDGEQQTDKFKALNPMGQVPVLLVDGKSISQS
VAIMEYLEEKYPEPRLLPADAYLRAKCREVVELLVSGIQPLQSIGLIPLLGKAEWKKWADRTITRGFTALETI
FAETAGKYCFGDEVTFADACLVPQVCNAYRFGVDVTPXXTIRRIYDALQQHPLVKKADPSCQPDAHLH

>Amblyomma parvum(A0A023FXL6)

MSKPVLLSNYLSSCAWRVRIVLEAKKIPYEYRTVDLKPANGEQQTDKFKALNPMGQVPVLLVDGKPISQS
VAIMEYLEEKYPEPRLLPADSYLRAKCREVVELLVSGIQPLQSIGLIPLLGKAEWKKWADRTISRGFTALEAI
FAETAGKYCFGDELTFADACLVPQVCNAYRFGVDVTPXPTIRRIYDALQQHPLVKKADPSCQPDTPPTGI
PNT

>Amblyomma triste(A0A023GJ13)

MSKPVLLSNFLSSCAWRVRIVLEVKKISYEYRTVDLKPKDGEQQTDKFKALNPMGQVPVLLVDGKPISQS
VAIMEYLEEKYPQPSLLPTDLYFRAKCREVVEVLVSGIQPLQSIGLIPLLGKAEWKKWADRNITRGFTALEAI
LTETAGKYCFGDEVTFADACLVPQVCNAYRFGVDVTPFPTIRRIYEALQQHPLVKKADPSCQPDTPPTGIP
NTTDLFKGPGDQ

>Escherichia coli(A0A023Z6H5)

MKLYSFFNSSASYRVRIALALKGINYQTEGVNIRIGQQNELAYRRMNPVGLVPTLLTDEGQSLGQSLAIID
WLERHYPQVPLVPQEEPARNKVLEIVYAIACDIHPLNNLRVLRYLTEELNVSEEEKKRWYAHWIQQGLSA
VEQLLRQNQSGQFCVGETPTLADCCLVPQWANALRMNCDLSGYPRCKAVYDACTQLPAFIAAAPENQ
QDKISA

>Serpentinomonas raichei(A0A060NLQ4)

MKLYNYFRSSTSFRVRIALNLKGLDCEYASVHLARGEQREAAYRALSPDGLVPLLDLQGVPEPGLLSQSM
AIIEYLDEVYPQPPLLPPDPLGRSRVRALAQSVACEMHPINNLRVLKYLAQPLGLNEEQRAAWYNHWVV
QGLLAYEQRLRELQAERAARALPPARYSYGDAPSLADCCLVPQLINGRRFGLSYDALDIPLTLAVLQACLEL
DAFRRALPENCPDAPPV

>Theobroma cacao(A0A061DU28)

MAPKDGGEASSKLVLYSYWQSSCSWRVRFALNLKGLSYEYKAVNLAKGEQFTPEFEKLNPLHFVPVLVD
GDVVVSDSYAILMYLEEKYPQRTLLPADPQQKALNLQVASIISSSIQPLLMLSILKYLEEKVGLEERLLFVQT
NIEKGFLALEKLLKDFVGKYATGEEVYMADVFMAPQIAVATERFKIDMSKFPTLSRIYESQRALPEFLAASP
ERQPDAVH

>Theobroma cacao(A0A061DV17)

MAPKDGGEASSKLVLYSYWQSSCSWRVRFALNLKGLSYEYKAVNLAKGEQFTPEFEKLNPLHFVPVLVD
GDVVVSDSYAILMYLEEKYPQRTLLPADPQQKALNLQKYLEEKVGLEERLLFVQTNIEKGFLALEKLLKDFV
GKYATGEEVYMADVFMAPQIAVATERFKIDMSKFPTLSRIYESQRALPEFLAASPERQPDAVH

>Theobroma cacao(A0A061DVZ4)

MAPKDGGEASSKLVLYSYWQSSCSWRVRFALNLKGLSYEYKAVNLAKGEQFTPEFEKLNPLHFVPVLVD
GDVVVSDSYAILMYLEEKYPQRTLLPADPQQKALNLQVASIISSSIQPLLMLSILKYLEEKVGLEERLLFVQT
NIEKGFLALEKLLKDFVGKYATGEEVYMGFVICRLMYLWHLRLLWLQNGLRLTCPSSLL

>Panstrongylus megistus(A0A069DPN4)

MSLIGKPVLYSYWRSSCSWRVRIALNLKEIPYDIKPISLIKGGGEQHSNEFREINPMEHVPALQIDGHTLIES
LSIMYYLEETRPQRPLLPQDVYKRAKVREICDVIASGIQPLQNIGVLIYVGEEKKKEWAQHWITRGFRAVE
KLLSSSAGKYSVGDELTLADCCLIPQVFNARRFQVDLRPYPIILRIDRELENHPAFRAAHPSNQPDCPPEAT
K

>Erythrobacter litoralis(A0A074MD55)

MKLHGYYRSSTSYRLRIALELKGLAFENVPVNLLESAQKDAAFTSRNPFASVPMLEADGRDRAQSMALIE
WLDEAYPAKPLLPADIEARYTARELAYAIATELHAPLNLPVLKYLKEEYGKSPGEIGEWYRHWLARTLVPV
EQRLEQLGTGDFLFDAPGLFETVLVPQLYNARRFEYDLSASPRMTRIEAACLALEPFRRAHPDNQNDNP
QRETS

>Erythrobacter longus(A0A074N282)

MKLYGYFRSSTSYRLRIALELKGLAYENIGVNLLHSEQKDEGFTSRNPFGSVPLLEADGRDRAQSMAMIE
WLDEAYPQKPLLPSDIEDRYTARELTYAIATELHAPLNLPVLKYLKNEYGKSQDEIDIWYRHWLARTLVPLE
QRLAQLGTGDFLFDKPGIFEVVLMPQIYNARRFAYDFGDAPHMMRIEAACLALPEFQRAHPDNQNDNP
ERT

>Spodoptera litura(A0A075X262)

MGDDVETKPVLYSYWRSSCSWRVRIALNLKEIPYDIKAVSLIKGGGEQHCNEYREVNPMEQVPSLCIDGH
TLVESLNIMHYLEETRPNQRPLMPQDCFKRAKVREICEVISSGIQPLQNLIVLIYVGEEKKKEWAQHWITR
GFRAVEKLLSACAGKYCVGDEITLADCCLVPQVFNARRFHVDLRPFPIILRIDRELENHPAFRAAHPSSQPD
CPPEVAK

>Spodoptera litura(A0A075X2K1)

MAENRAILYAYWLSSCSWRVRAALHFKGIPFEERPIDIVKTNQQKTEQFRAINPAQKVPALVIDNVTLVES
MAIVQYLEDTHPEPTLTPKTPVLRARMRELCEVVVSGIQPLQNIGLRSQFDTTEQYTKFTKYWTDRGLMT
LEDLLQKSAGKYCVGDQLTLADLCLVPQLYNAVTRHALDISKYPTVSKLYESLLKENVFKETHPESVKSKM

>Triticum aestivum(O04437)

MATAKPILYGAWISSCSHRVRIALNLKGVDYEYKAVNPRTDPDYEKINPIKYIPALVDGDFVLSDSLAIMLY
LEDKYPQHPLVPKDIKTKGLDLQIANIVCSSIQPLQGYGVIGLHEGRLSPDESLEVVQRYIDKGFRAIEKLLD
GCDSKYCVGDEVHLGDVCLAPQIHAAINRFQIDMTKYPILSRLHDAYMKIPAFQAALPQNQPDAPSAK

>Nasonia vitripennis(K7JNH5)

MSVIGKPILYSYWRSSCSWRVRIALNLKEIPYDIKPISLVKNGGEQHSNEFREINPMEQVPALHIDNHTLIE
SLNILLYLEETRPHRPLMPVDPVKRARVREICEVIASGIQPLQNLIVLIYVGEERKKEWAQHWITRGLKAVE
KLLSASAGKYCVGDEITLADCCLVPQIFNARRFHVDLRPFPTILRVDRHLENHPAFTAAHPNNQPDCPPEA
TK

>Apis mellifera(A0A088AB84)

MSVMGKPILYSYWRSSCSWRVRIALNLKEIPYDIKPVSLIKGGGEQHSNEFREINPMEQVPALHIDNHTLI
ESLNILQYLEETRPHRPLMPADPVKRARVREICEVIASGIQPLQNLVVLIYVGEERKKEWAQHWITRGLTA
VEKLLSSSAGKYCVGDEITLADCCLIPQIFNARRFLVDLRPFPTILRVDRHLENHPAFTAAHPNNQPDCPPE
ATK

>Plutella xylostella(214AA)

MAKPVLYSYWRSSCSWRVRIALNLKEIPYDIKAVSLIKGGGEQHCNEYRECEVISSGIQPLQNLVVLIYVGE
EKKKEWAAHWMTRGFRAVERLLSGSAGKYCVGDEITLADCCLVPQVFNARRFHVDLRPFPIILRIDRELE
HHPAFRAAHPSTQPDCPPEAAK

>Plutella xylostella(213AA)

MASPAVLHGFFASSCTWRVRAALVLKSIPFEERHVDIVQLKTHLSDQYQAVHPAQKVPALEIDGTTLVES
MAILQYLEDTRPRPALAPAAPLPRARMREIVETIVSGIQPLQNVGVRGLLGSDEEYSAFSRGAARRALQTL
EALLARSAGQYCVGDQLSMADLCFVPQLFNAVGRLKLDISDLPTISKLYAKLSKEEIFMKTHPRTVKHLSET

>Drosophila melanogaster(Q9VHD2)

MSTNLCPNASSSDIQPILYSYWRSSCSWRVRIAMNLKEIPYDIKPISLIKSGGEQHCNEYREVNPMEQVPA
LQIDGHTLIESVAIMHYLEETRPQRPLLPQDVHKRAKVREIVEIICSGIQPLQNLIVLIHVGEEKKKEWAQH
WITRGFRAVEKALSTSAGKYCVGDEISMADCCLVPQVFNARRFHVDLRPYPIILRIDRELESNPAFRAAHP
SNQPDCPPELPNK

>Bombyx mori(Q2I0J4)

MGKQPVLYSYWRSSCSWRVRIALNLKEIPYDIKAVSLIKGGGEQHCNEYREVNPMEQVPSLCIDGHTLIES
LNIMHYLEETRPQRPLMPQDCFKRAKVREICEMIASGIQPLQNLIVLIYVGEEKKKEWSQHWITRGFRAIE
KLLSTTAGKYCVGDEITLADCCLVPQVFNARRFHVDLRPFPIILRIDRELENHPAFRAAHPSSQPDCPPEVA
K

>Bombyx mori(A5HSJ9)

MVENRVILHAYWLSSCSWRVRAMLHAKSIPFEERPVDIVKTGKQLTEEYRAINPAQKVPALEIDGVTLVE
STAIIQYIEDTRPEPKLMPDTALQRARMREICETIVSGIQPLQNFGLKKHLGTEEKFLSFTKYWTERGLQTL
NDLLAKTSGAYCIGDQITLADICLVPQIYNGVSRHKLDLKTYPIVSKVYENLLKEELYQATHPKATKEKLKINL

>Danaus plexippus(A0A212FF03)

MAETRAILYGFWASSCTWRVRAALHFKGIAFEEKSVDIVTEQKQLTDEYRYINPSQKVPALVMNGETIVE
SMAIIQYVEEIKPEASLVPTTPILRARMREICETIVSGIQPLQNIGLKRRFDSEHKFKEFAEYFTSRGLESVEEL
LKKTAGRFCIGDQITVADLCLVPQVYNGIVRYKMNMEKFPIVSSVYEHLLKEKTFIETHPKNIK

>Arabidopsis thaliana(Q9ZVQ4)

AKLKLYSYWRSSCAHRVRIALTLKGLDYEYIPVNLLKGDQSDSDFKKINPMGTVPALVDGDVVINDSFAII
MYLDDKYPEPPLLPSDYHKRAVNYQATSIVMSGIQPHQNMALFRYLEDKINAEEKTAWITNAITKGFTAL
EKLLVSCAGKYATGDEVYLADLFLAPQIHAAFNRFHINMEPFPTLARFYESYNELPAFQNAVPEKQPDTPS

>Arabidopsis thaliana(Q9ZVQ3)

EKLKLYSYWRSSCAHRVRIALALKGLDYEYIPVNLLKGDQFDSDFKKINPMGTVPALVDGDVVINDSFAII
MYLDEKYPEPPLLPRDLHKRAVNYQAMSIVLSGIQPHQNLAVIRYIEEKINVEEKTAWVNNAITKGFTALE
KLLVNCAGKHATGDEIYLADLFLAPQIHGAINRFQINMEPYPTLAKCYESYNELPAFQNALPEKQPDAPS

Out Group

>Oryctolagus cuniculus(P08628)

MVKQIESKSAFQEVLDSAGDKLVVVDFSATWCGPCKMIKPFFHALSEKFNNVVFIEVDVDDCKDIAAECE
VKCMPTFQFFKKGQKVGEFSGANKEKLEATINELL

APPENDIX C

CHAPTER 3 SUPPLEMENTARY DATASET

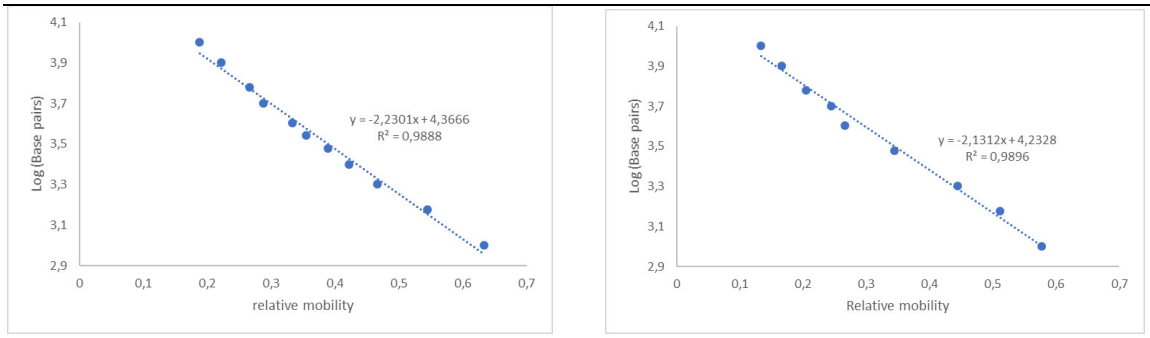


Figure S2: Standard curves relating the relative mobility of the DNA ladder size on 0,75 % agarose gels respective to their log base pair sizes to determine the restriction digested fragment sizes. A: The standard curve for the pET11a-C*p*GST and pCOLD1-C*p*GST restriction digest samples with the GeneRuler 1 Kb DNA Ladder (ThermoFisher Scientific , USA). B: The standard curve for the pCOLDTF-C*p*GST restriction digest samples with the 1 Kb DNA Ladder (New England Biolabs, USA).

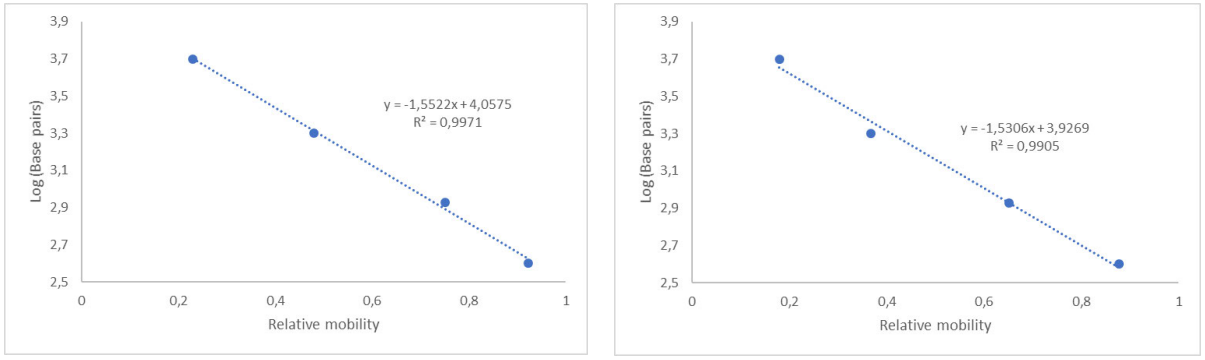


Figure S3: Standard curves relating the relative mobility of the DNA ladder size on 1 % agarose gels respective to their log base pair sizes to determine the PCR product sizes. A: The standard curve for the pCOLD1-C*p*GST PCR product samples. B: The standard curve for the pCOLDTF-C*p*GST PCR product samples. Both samples had the FastRuler middle range DNA Ladder (ThermoFisher Scientific , USA).

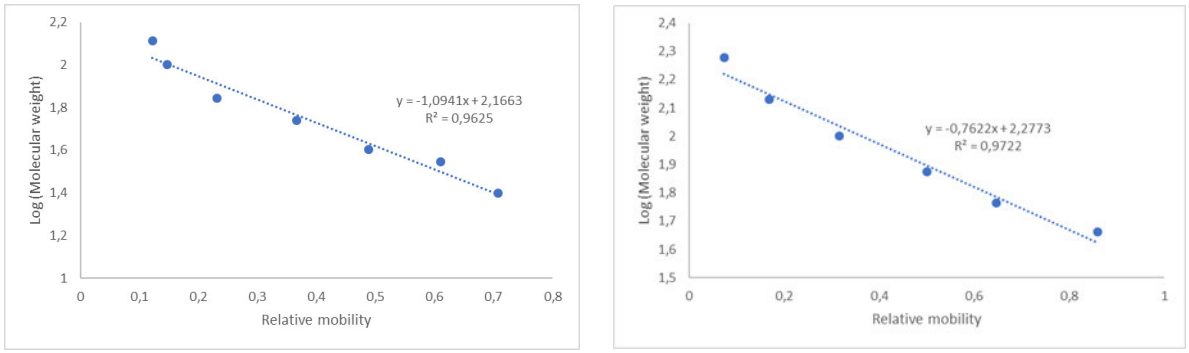


Figure S4: Standard curve relating the relative mobility of the molecular weight marker proteins on 12.5 % SDS-PAGE laemmli system respective to their log molecular weight. A: The standard curve for the pCOLD1-C*p*GST expression samples with the blue prestained protein standard (New England Biolabs, USA). B: The standard curve for the pCOLDTF-C*p*GST expression samples with the BLUeye prestained ladder (Sigma-Aldrich, USA).

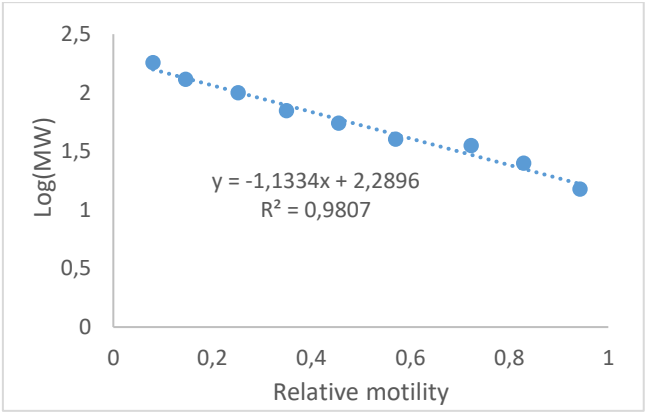


Figure S5: Standard curve relating the relative mobility of the molecular weight marker proteins on 12.5 % SDS-PAGE Laemmli system respective to their log molecular weight. The calibration curve was constructed by plotting the Log (Mr) of the protein ladder against the relative mobility of that the proteins travelled. A: The standard curve for the pCOLDTF-C*p*GST expression and eluents after purification. The IMAC and SEC gels were run against the PageRuler™ prestained protein ladder (ThermoFisher Scientific, USA)

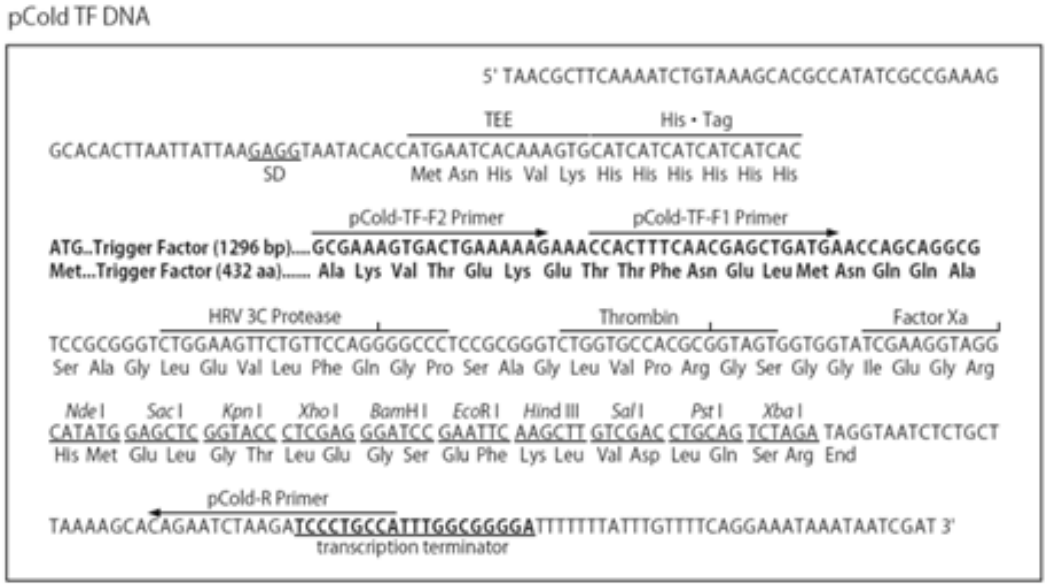


Figure S6: pCOLDTF vector map showing the primer recognition sequences.