Identification of potential inhibitors of the folate biosynthesis enzymes HPPK of *Salmonella enterica* and *Escherichia coli* and pteridine reductase of *Trypanosoma brucei* through molecular docking and enzyme assays

Thesis submitted in the fulfilment of the requirements for the degree of

MASTER OF SCIENCE IN PHARMACY

Faculty of Pharmacy

Rhodes University

Grahamstown/Makhanda

Eastern Cape

South Africa

By

Tiaan Marc Gerwel

February 2021

Principle supervisor: Professor H. C. Hoppe

Co-supervisor: Doctor C. G. L. Veale

Abstract

Antimicrobial resistance has become a serious threat to the survival of the human species especially those living in rural areas where access to medicine and the knowledge for proper use is scarce. It has been estimated that the number of extreme untreatable resistant infections in Africa will increase to as much as 10 million by the year 2050. Thus the need for novel drugs to act as therapeutic agents is becoming more compelling each year. The subspecies of *Trypanosoma brucei* (*T. brucei*) is responsible for the Human African Trypanosomiasis (HAT) also known as sleeping sickness and results in large numbers of deaths and loss of income for many homes. Resistance to current therapeutic agents has been observed and is on the rise increasing the difficulty to treat the infection. *Salmonella enterica* (*S. enterica*) serotypes are responsible for acute diarrhoeal disease in humans ranging from invasive typhoid to non-invasive non-typhoid disease, resulting in a large number of deaths, with an estimated 180 million people falling ill annually. The pathogen is spread via the faecal-oral route or through food contaminated with bacterium and prepared in an unsanitary environment. The chance of recovery in rural populations is low. *Escherichia coli* (*E. coli*), forming part of commensal gut flora, spread via the faecal-oral route or unhygienic practices, causes a diarrhoeal disease which can progress to a haemorrhagic phase. More than 241 million annual infections are caused by enterotoxigenic *E. coli*. A common strategy to develop antimicrobial agents is to target the biosynthesis of essential biological molecules, thereby rendering the microbes less viable. One such group of molecules are folates, which are generally synthesised *de novo* by bacteria. Higher organisms have a scavenger mechanism to obtain their folates from the extracellular environment and in some cases, organisms have both mechanisms. In this study, two enzymes falling into the folate scavenger and *de novo* synthesis groups were examined to identify potential agents to act as inhibitors. Pteridine reductase 1 (PTR1) is a scavenger enzyme used by a variety of trypanosomes and 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK) is an enzyme that overlaps the kingdoms of protozoa and bacteria forming part of the *de novo* biosynthetic pathway. Homology modelling was performed on the HPPK enzyme of *S. enterica* using *Yersinia pestis* (*Y. pestis*) as a template providing a model to use for docking studies. The *E. coli* HPPK enzyme structure was retrieved from the Protein Data Bank (PDB) and bound molecules were removed to render the enzyme in an apo-state. Docking studies using the generated *S. enterica* homology model and apo *E. coli* HPPK was performed using the ZINC 15 database and resulted in 9 hit compounds which showed high binding affinities and binding energy to the enzyme. The HPPK and PTR1 enzyme coding sequences were cloned into pET-28a(+) plasmids and supplied by GenScript, to enable the expression of histidine-tagged proteins in T7 Express *lys*Y competent *E. coli* cells. Analytical scale expression studies showed the recombinant proteins to be in a soluble form

and purification was achieved using nickel-NTA affinity chromatography. The purified PTR1 recombinant protein was used to establish and optimise an NADPH absorbance microplate assay to screen compounds previously identified in docking studies by Kimuda, Laming, Hoppe, & Bishop, (2019). The assay yielded a Z´-factor of above 0.8 indicating an excellent assay to use for screening. An unsuccessful attempt was made to use resazurin reduction as an alternative method to demonstrate PTR1 enzyme activity. HPPK purified recombinant proteins were used to establish and optimise a luminescence microplate assay for the screening of compounds identified in *in silico* docking studies against the HPPK enzymes of *S. enterica* and *E. coli*. The Z´-factor of the luminescence assay was above 0.5, indicative of a functional assay with good separation between enzyme activity signal and negative control of reaction without enzyme. The target-based enzyme screening resulted in the confirmation of compound 3 (3-chloro-N-[(4-oxo-3,4-dihydrophthalazin-1-yl)methyl]benzamide) as an inhibitor of *S. enterica* HPPK with an IC_{50} of 10.4 µM. At 50 µM none of the compounds decreased *E. coli* HPPK enzyme activity by 50%. Further bacterial studies would provide more compelling data to motivate the optimisation of compound 3 as an *S. enterica* inhibitor. This study demonstrated the effectiveness of using computational methods in the drug discovery process, correlating *in silico* results with those obtained from target-based assays producing a hit compound that can be used for future drug optimisation.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

%	Percentage
<	Less than
>	Greater than
≤	Less than or equal to
≥	Greater than or equal to
°C	Degrees Celsius
µ	Mean
µg	Micrograms
µg/ml	Micrograms per milliliter
µl	Microliters
µm	Micrometer
µM	Micromolar
1D	One dimensional
2D	Two dimensional
3D	Three dimensional
8MG	8-mercaptoguanine
8TG	8-thioguanine
A	Adenine
Å	Angstrom
AAT6	Amino acid transporter 6
ABC	ATP-binding cassette
Abs_{620}	Absorbance at 620 nm
ADME	Absorption, distribution, metabolism and elimination
ADP	Adenosine diphosphate
Ala	Alanine

AMP	Adenosine monophosphate
AmpCs	Cephalosporinases
AQP2	Aquaglyceroporin 2
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
ATP	Adenosine triphosphate
ATP1	Aminopurine permease
AU	Absorbance units
BBB	Blood-brain barrier
blob	Binary Large OBject
BSA	Bovine serum albumin
BT1	Biopterin transporter 1
C	Cytosine
CBP	Chitin-binding protein
cDNA	Complementary DNA
CNS	Central nervous system
CO_2	Carbon dioxide
conc	Concentration
Cys	Cysteine
DAEC	Diffusely adherent *E. coli*
DALYs	Disability-adjusted life years
DBU	1,8-Diazabicyclo[5.4.0]undec-7-ene
DEAE	Diethylaminoethyl
DHF	Dihydrofolate
DHFR	Dihydrofolate reductase

DHFS	Dihydrofolate synthetase
DHNA	Dihydroneopterin aldolase
DHNE	7,8-Dihydroneopterin triphosphate epimerase
DHNP	Dihydroneopterin 3'-triphosphate
DHPS	Dihydropteroate synthase
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
e.g.	For example
EAEC	Enteroaggregative *E. coli*
EAST1	Enteroaggregative heat-stable toxin 1
EC_{50}	Half-maximal effective concentration
EDTA	Ethylenediaminetetraacetic acid
EIEC	Enteroinvasive *E. coli*
EPEC	Enteropathogenic *E. coli*
ESBLs	Extended-spectrum β-lactamases
et al.	And others
ETEC	Enterotoxigenic *E. coli*
F. tularensis	*Franscisella tularensis*
FASTA	Fast-all
FPGS	Folylpolyglutamate synthetase
G	Guanine
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
GST	Glutathione-S-transferase

GTP	Guanosine triphosphate
GTPCH1	GTP cyclohydrolase
H. influenzae	*Haemophilus influenzae*
H_2B	Dihydrobiopterin
H_4B	Tetrahydrobiopterin
HAT	Human African Trypanosomiasis
HCl	Hydrogen chloride
HeLa	Henrietta Lacks (Human cervix adenocarcinoma)
Hepes	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-tag	Polyhistidine tag
His	Histidine
HIV	Human immunodeficiency virus
HMDP	6-Hydroxymethyl-7,8-dihydropterin
HP-18	5'-*S*-[1-(2-{[(2-Amino-7,7-dimethyl-4-oxo-3,4,7,8-tetrahydropteridin-6-yl)carbonyl]amino}ethyl)piperidin-4-yl]-5'-thioadenosine
HP-19	2-Amino-7,7-dimethyl-4-oxo-3,4,7,8-tetrahydro-pteridine-6-carboxylic acid(2-{4-[5-(6-amino-purin-9-yl)-3 4-dihydroxy-tetrahydro-furan-2-ylmethylsulfanyl]-piperidin-1-yl}-ethyl)-amide
HP-26	2-Amino-7,7-dimethyl-4-oxo-3,4,7,8-tetrahydro-pteridine-6-carboxylic acid(2-{2-[5-(6-amino-purin-9-yl)-3,4-dihydroxy-tetrahydro-furan-2-ylmethanesulfonyl]-ethylcarbamoyl}-ethyl)-amide
HP_2A	P^1-(6-hydroxymethylpterin)-P^2-(5´-adenosyl)diphosphate
HP_3A	P^1-(6-hydroxymethylpterin)-P^3-(5´-adenosyl)triphosphate
HP_4A	P^1-(6-hydroxymethylpterin)-P^4-(5´-adenosyl)tetraphosphate
HPLC	High-performance liquid chromatography

HPPK	6-Hydroxymethyl-7,8-dihydropterin pyrophosphokinase
HPPP	6-Hydroxymethyl-7,8-dihydropterin pyrophosphate
HTS	High-throughput screening
HUS	Haemolytic uremic syndrome
Hz	Hertz
IC_{50}	Half-maximal inhibitory concentration
Ile	Isoleucine
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
ISG75	Invariant surface glycoprotein 75
IUPAC	International Union of Pure and Applied Chemistry
kcal/mol	Kilocalories per mole
K_D	Dissociation constant
kDa	Kilodalton
lacI	*lac* repressor
LB	Luria broth
LDL	Low-density lipoprotein
Leu	Leucine
LIB_66	4-(Benzyloxy)pyrimidine-2,6-diamine
logP	Partition coefficient
Lys	Lysine
M	Molar
M. tuberculosis	*Mycobacterium tuberculosis*
M1	Muscarinic receptor type 1
MBP	Maltose-binding protein
Met	Methionine

mg/L	Milligram per liter
mg/ml	Milligram per milliliter
Mg^{2+}	Magnesium
$MgCl_2$	Magnesium Chloride
min	Minutes
ml	Milliliters
mM	Millimolar
Mn^{2+}	Manganese
mRNA	Messenger RNA
MRPA	Multidrug-resistant protein A
MTX	Methotrexate
N_2	Nitrogen
NaCl	Sodium chloride
NADH	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
NBT/PES	Nitroblue tetrazolium / phenazine ethosulfate
NCBI	National Center for Biotechnology Information
Ni^{2+}	Nickel
$NiSO_4$	Nickel (II) Sulfate
nm	Nanometer
NMR	Nuclear magnetic resonance
NTA	Nitriloacetic acid
NTR-1	Type I nitroreductase
NTS	Non-typhoid *Salmonella*
O_2	Oxygen
OD	Optical density

OD_{600}	Absorbance at 600 nm
ODC	Ornithine decarboxylase
P. falciparum	*Plasmodium falciparum*
P. jiroveci	*Pneumocystis jiroveci*
P2	Adenine nucleotide
PAINS	Pan-assay interference compounds
PBPs	Penicillin-binding proteins
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDB	Protein database
pH	Potential of hydrogen
Phe	Phenylalanine
pLDH	Parasite lactate dehydrogenase
pmol	Picomoles
PMQR	Plasmid-mediated quinolone resistance
PRIMO	Protein interactive modelling
Pro	Proline
ProSA	Protein structure analysis
PTR1	Pteridine reductase
RFU	Relative fluorescence units
RLU	Relative luminescence units
RMSD	Root mean square deviation
RNA	Ribonucleic acid
RPMI	Roswell Park Memorial Institute
rRNA	Ribosomal RNA
S. aureus	*Staphyllococcus aureus*

S. bongori	*Salmonella bongori*
S. cerevisiae	*Saccharomyces cerevisiae*
S. enterica	*Salmonella enterica*
S. enterica enterica	*Salmonella enterica enterica*
S. enteritidis	*Salmonella enterica enterica enteritidis*
S. paratyphi	*Salmonella enterica enterica paratyphi*
S. typhi	*Salmonella enterica enterica typhi*
S. typhimurium	*Salmonella enterica enterica typhimurium*
SAR	Structure-activity relationship
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
sec	Seconds
Ser	Serine
SMILES	Simplified Molecular Input Line Entry System
SPI	*Salmonella* pathogenicity islands
STEC	Shiga toxin-producing *E. coli*
Strep-tag	Streptavidin tag
T	Thymine
T. b. gambiense	*Trypanosoma brucei gambiense*
T. b. rhodiense	*Trypanosoma brucei rhodiense*
T. brucei	*Trypanosoma brucei*
T. cruzi	*Trypanosoma cruzi*
TEMED	N,N,N´,N´-Tetramethyl-ethylenediamine
THF	Tetrahydrofolate
Thr	Threonine
TPSB	Tris protein storage buffer

tRNA	Transfer RNA
Trp	Tryptophan
TS	Typhoid *Salmonella*
Tyr	Tyrosine
UBP1	Upstream binding protein
US$	United States dollars
UV-Vis	Ultraviolet-visible
V	Volts
v/v	Volume per volume
Val	Valine
vdW	van der Waal's forces
VSG	Variant surface glycoprotein
VTEC	Verocytotoxigenic *E. coli*
w/v	Weight per volume
WHO	World Health Organisation
X	Times
xg	Times gravity
XlogP	Octanol/water partition coefficients
Y. pestis	*Yersinia pestis*
α	Alpha
β	Beta
γ	Gamma
Δ	Change
ΔAbs_{340}	Change in absorbance at 340nm
ΔLum	Change in luminescence
π	Pi

PABA	Para-aminobenzoic acid
σ	Standard deviation

Acknowledgements

I would like to express my deepest gratitude to my supervisor Professor Heinrich Hoppe and my co-supervisor Dr Clinton Veale for their unwavering support, patience, understanding and guidance throughout this project. Professor Hoppe, I thank you for taking time out of your day to teach me hands on the experimental procedures used in this project, and for listening to me and my ideas with regards to the project. During times when experiments were unsuccessful or results difficult to interpret, your input and advice provided the much-needed insight in developing a workable strategy towards success. Your support and patience when life got in the way, is sincerely appreciated. Magambo Phillip Kimuda, thank you for the assistance you provided me in performing the bioinformatics section of the work, your experience and guidance has taught me a great deal and Professor Özlem Tastan Bishop for allowing me into the group and assistance you provided.

To my parents Thomas and Moira, I would like to express my gratitude for your continuous support and motivation during my studies, I truly could not have gotten this far if it wasn't for you as you have been present in every facet of my academic life. My siblings Zann and Dayne, you have guided, motivated and been the ears needed whenever your presence was requested, thank you. To Zama and Arlo, you have truly walked this journey with me in every sense of the word, from experiments to lecturing, you have been there for me and for this I cannot thank you enough. The support family has shown me has been unimaginable and incredible and therefore a very big thank you goes to them.

The members of Professor Hoppe's lab, thank you. Leigh-anne Derry, Tarryn Swart and Dustin Laming, you have assisted in teaching me experimental procedures and provided insight into biochemistry I previously did not have.

My studies were funded by the Jakes Gerwel Family Trust, whose support is greatly appreciated.

Chapter 1

Introduction and Literature Review

1.1 Burden of Disease

1.1.1 Trypanosomiasis

Human African Trypanosomiasis (HAT), commonly known as sleeping sickness, forms part of the neglected tropical diseases occurring in sub-Saharan Africa. HAT is a zoonotic infection with the causative organism of the protozoan species *Trypanosoma brucei* causing infections in humans, with subspecies *T. b. gambiense* responsible for a slow-progressing disease endemic to regions in central and western Africa, while the faster-progressing *T. b. rhodesiense* is endemic to regions in southern and eastern Africa (Visser, 2020). *T. b. brucei* causes disease in animals and forms a third subspecies of *T. brucei*. The disease is not only limited to Africa but also has a similar disease prevalent in South America, called Chagas disease which does not have neurological symptoms, and is caused by *Trypanosoma cruzi* (Lenk *et al.*, 2018). In Africa, which will be the focus of this dissertation, HAT is spread across 36 African countries and has been responsible for the loss of millions of lives throughout recorded African history, with epidemic flare-ups happening regularly with the latest occurring in the 1990s, where more than 50 000 new cases were reported yearly (Mehlitz and Molyneux, 2019). Between 1999 and 2019 the number of reported cases due to *T. b. gambiense* fell from 27 862 to 864, a decrease of 97%. *T. b. gambiense* accounted for roughly 88% of reported HAT infections. The number of reported cases due to *T. b. rhodesiense* dropped by 81%, from 619 to 116, and accounted for 12% of HAT infections. Between 2000 and 2010, more than 100 cases were reported in non-endemic countries, indicating the spread of the disease due to human movement either for employment or tourism (Büscher *et al.*, 2017). The disease has a 100% fatality if left untreated. It is estimated that 39% of HAT cases and 92% of deaths due to HAT remain unreported making the true incidence a lot higher than reported (Lutumba *et al.*, 2007).

HAT is transmitted via the bite of a tsetse fly from the genus *Glossina.* The insect vectors haemolymph being incapable of providing sufficient amounts of nutrients for parasite survival are not mortally infected. The tsetse fly feeds on the blood of mammals, with a greater preference for the blood of animals and only feeding accidentally on humans (Ponte-Sucre, 2016). Tsetse flies are attracted to bright colours, very dark colours, metallic fabric, strongly to the colour blue, large fast-moving objects and carbon dioxide. These attraction factors are all applicable to humans, especially carbon dioxide which is released during the burning of wood commonly used for outside cooking in rural areas. The flies are commonly found in forest areas where they rest in holes in trees and roots

and areas close to food sources thus bringing them close to farming communities (Brun *et al.*, 2010). The female flies produce larvae and do not lay eggs. The larvae develop in the uterus over 10 days after which they are deposited into moist soil in shaded areas. The larvae will continue to develop and flies will emerge within 22-60 days dependent on favourable temperature conditions. Female tsetse flies are seen to typically mate only once during their lives and will give rise to a single larva, in utero, approximately once in a 10-day cycle, depending on nutrient availability and suitable habitat (Büscher *et al.*, 2017).

The *T. brucei* species are parasitic haemoflagellate protists. The species can change the constitution of their glycoprotein coat allowing them to escape the immune reaction of the host. The parasite follows a cyclical path between human hosts and insect vectors (Barrett, 2018). Figure 1, shows the complete life cycle of the trypanosome. Within the insect vector, the parasite will enter 3 major phases of development before being transmitted to the human host. When the tsetse fly feeds on the blood of an infected human, two major morphological forms exist, namely a slender (normal) form and a stumpy (non-replicating form) form (Kennedy, 2004). The non-replicating form of the parasite can survive the protease activity in the saliva and midgut of the flies. Once in the midgut of the fly, they transform into the procyclic form of the parasite. The mechanisms causing transformation are not well understood and include theories such as midgut protease causing a natural trigger in the parasite or the change in temperature from host to midgut, exposure of the parasite to high levels of citrate or inhibition of the effects of tyrosine phosphate (Kennedy, 2019). In the procyclic form, the parasite attaches to the cells of the ectoperitrophic space and begin to transform into motile forms called trypomastigotes which travel along the proventriculus of the midgut to the salivary glands of the fly. The kinetoplast organelle which powers trypomastigote flagellar motility is absorbed intracellularly causing the change to the long epimastigote form which settles on the basolateral membrane of the salivary glands where asymmetrical division results in a short epimastigote form (Büscher *et al.*, 2017). This developmental stage produces the metacyclic trypanosomes which are then able to infect humans. Subsequent transformations of the parasite will result in modifications of the glycoprotein coat to ensure the survival of the parasite in different environments. The procyclic and the epimastigote forms will multiply by binary fission (Kennedy, 2004). With the end of the fly portion of the life cycle, the short epimastigotes start to multiply, producing large numbers of the metacyclic trypanosomes which are then passed through the saliva into a new human host continuing its life cycle in humans. The life cycle in flies lasts approximately 3 weeks (Kennedy, 2019).

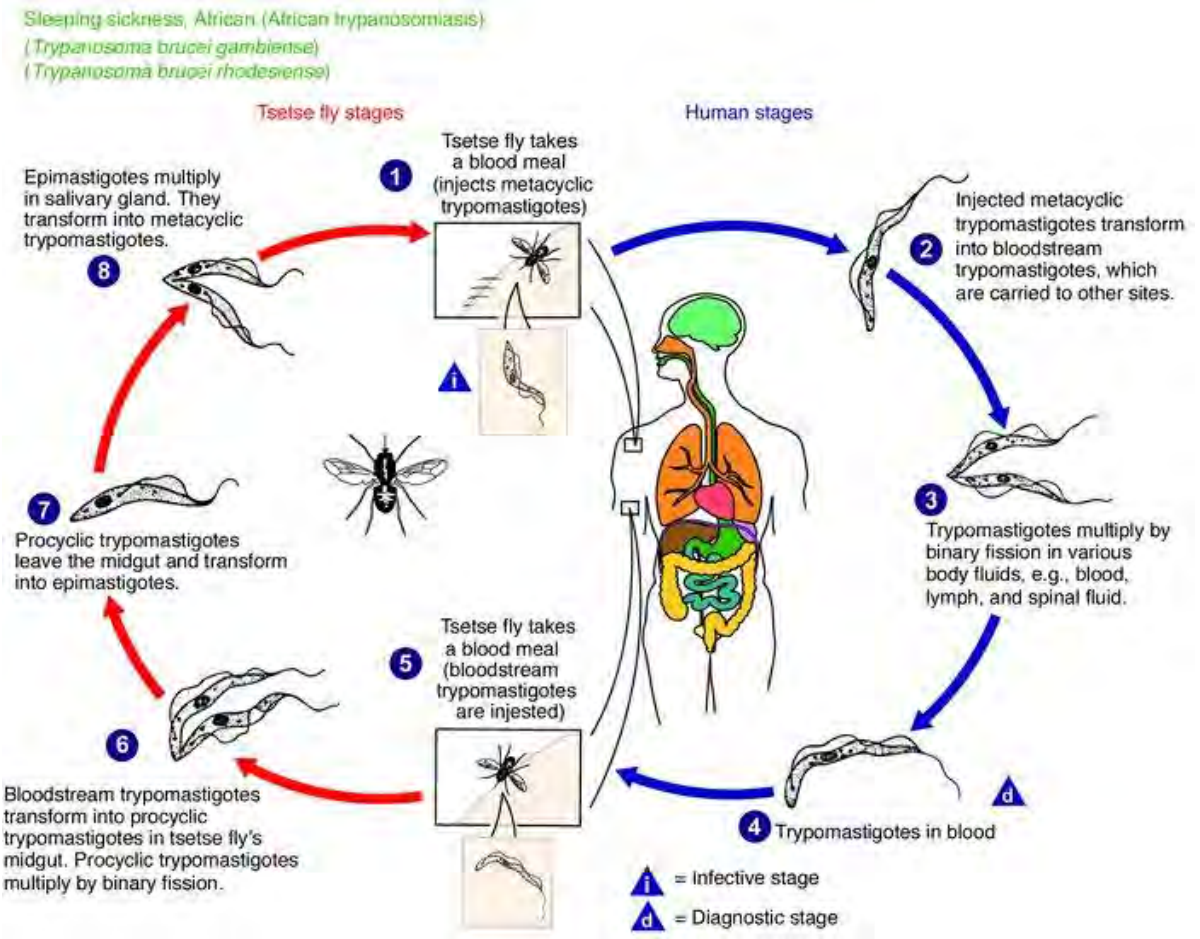


Figure 1: Insect and human life cycle of *T. brucei*. Image obtained from Kennedy (2004) with credit to Alexander J. da Silva and Melanie Moser, Centers for Disease Control Public Health Image Library.

Once the tsetse fly feeds on a human, the metacyclic form of the parasite enters the bloodstream, which results in a regulated cascade of biochemical signals to transform the parasite into the typical slender form. Lesions appear as the parasite tears through the capillary bed of the skin and the severity of the lesions depend on the subspecies, with *T. b. rhodosiense* causing a larger lesion as it is more fast-acting than *T. b. gambiense*. As the body attempts to clear the source of infection from the area through lymph, the proximal lymph nodes will become infected and spread to the entire lymphatic system (Mehlitz and Molyneux, 2019). The trypanosome parasite is digesting the cells of the lymphatic system leading to symptoms such as severe headaches, fever and fatigue. With the fast-progressing *T. b. rhodosiense*, the acute attack the body launches to clear the infection results in oedema, pancreatitis and congestive heart failure. *T. b. gambisiense* does not show these harsh bodily reactions and generally appears with lymphadenopathy and general malaise (Ponte-Sucre, 2016). One diagnostic clue of *T. b. gambisiense* infection is the presence of the Winterbottom's sign, which is lymphadenopathy occurring in the region of the trapezius of the thorax. This phase of the disease is known as the haemolytic phase. After a couple of weeks in the case of *T. b. rhodosiense*

infection and as much as a couple of years in the case of *T. b. gamsiense*, the parasites invade organ systems, as indicated by an increase in immune response (Fèvre *et al.*, 2008). The most deadly and devastating effect is on the central nervous system (CNS). The parasites enter the CNS, by tricking the blood-brain barrier (BBB) by expressing a parasitic version of cysteine protease which increases the number of changes in the oscillatory pattern of calcium ions in the cell. The attack on the CNS is known as the meningoencephalitic phase of the disease. This phase is characterised by an unusual sleeping pattern, during which the patient sleeps most of the day and suffers from intermittent insomnia at night (Kennedy, 2004). As the disease progresses, personality changes are noted in the form of anger and depression soon after which basic reasoning or mental processes are impaired. In the final stages of the disease, severe ataxic dyskinesia sets in and the loss of coordination of action occurs (Kennedy, 2019). Throughout the entire pathogenic cycle in humans, some of the parasites are converted to their stumpy form through a quorum sensing apparatus that promotes the rearrangement of the organelles and causes an arrest in mitotic activity. These stumpy parasites migrate to the surface of the skin and lie dormant in the capillary bed waiting to infect the insect vector. The particular protein involved in quorum sensing is unknown and has simply been referred to as a "stumpy inducing factor" (Hassan, Castanha and Wolfram, 2020).

The economic burden of HAT is enormous, as, in a 2008 study, Fèvre *et al.* (2008) estimated that approximately 60 million individuals are at risk of developing the disease. These at-risk individuals tend to live in remote areas, in agricultural communities and have a subsistence lifestyle. These communities tend to have no doctors available, no medication, health information is weak while the need is greatest, food security is low whilst hunger is high and the income of these individuals falls well below the poverty line (Bukachi, Wandibba and Nyamongo, 2017). Average estimates of the total time from infection to returning to work includes a 10 day hospitalisation period and a 90 day rest period after hospitalisation. A caregiver would on average spend 10 days taking care of the individual once at home. In communities where the average daily income is only US$ 1.2, the costs to treat HAT is estimated to be US$ 163.98, estimated to represent 43% of the annual revenue of a household. The disability-adjusted life years (DALYs) in rural communities in the Democratic Republic of Congo could total as much as 2145, indicating the number of years lost due to ill health, disability or death. As political tensions destabilise democracies, the incidence of HAT dramatically increases (Lutumba *et al.*, 2007).

Not only is the disease expensive to treat and has a devastating effect on households, but resistance to current treatment is also fast becoming widespread. The current treatment options approved and available are pentamidine, suramin, melarsoprol, eflornithine and nifurtimox, shown in Figure 2 A-E,

which are used in combination and only under special authorisation (Franco, Scarone and Comini, 2018). Unfortunately, these drugs are not anodyne as they possess a certain level of toxicity. Pentamidine and suramin are used during the early stages of the disease of both *T. b. gambiense* and *T. b. rhodiense*. Eflornithine is used as monotherapy for the second stage of *T. b. gambiense* infection as it is ineffective against *T. b. rhodiense*. It is also a challenge to administer the drug, requiring skilled staff and cumbersome equipment, making its use a logistical problem (de Koning, 2020). A combination of eflornithine and nifurtimox has improved the ease of delivery and improved safety and efficacy whilst decreasing administration time compared to that of eflornithine monotherapy. The combination is used to treat the late stage of *T. b. gambiense* infection. Melarsoprol is the only drug available for the treatment of late-stage *T. b. rhodesiense* infection and is effective in treating late-stage *T. b. gambiense* infection but is reserved as a second-line drug to lower drug resistance(Barrett, 2018). As of late 2019, the WHO has implemented the use of fexinidazole for the treatment of stage 1 or stage 2 *T. b. gambiense*. Fexinidazole (Figure 2F) needs to be administered under the supervision of skilled personnel (Lindner *et al.*, 2020).

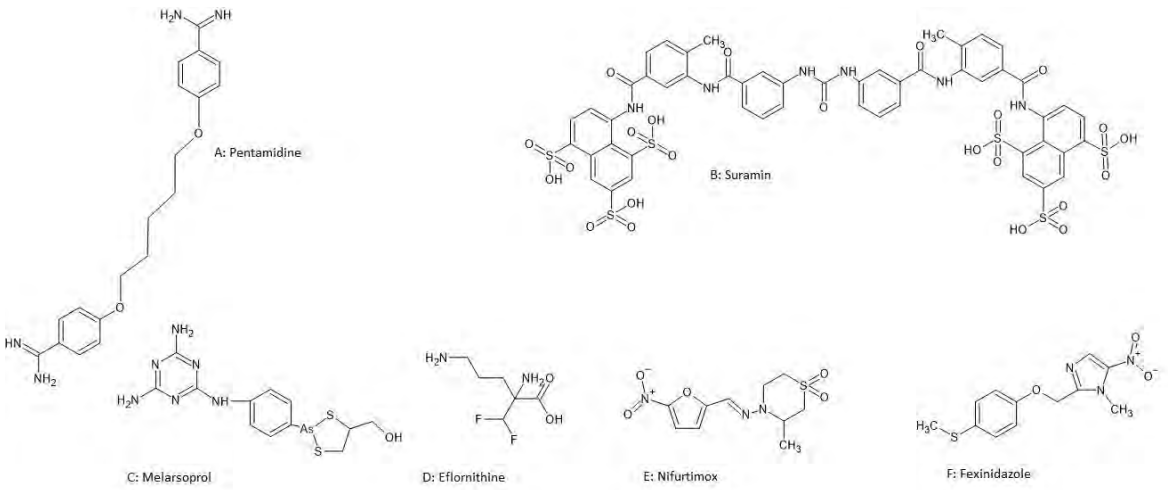


Figure 2: Drugs used in the treatment of *T. brucei* infection. A: Pentamidine, B: Suramin, C: Melarsoprol, D: Eflornithine, E: Nifurtimox, F: Fexinidazole. Images redrawn using ACD/ChemSketch from Franco, Scarone and Comini (2018) and Lindner *et al.* (2020)**.**

Pentamidine, Figure 2A, an aromatic water-soluble diamidine, enters the parasite through membrane transporters. For uptake into the parasite, pentamidine utilises the adenine nucleotide (P2) or aminopurine permease (ATP1) along with high-affinity pentamidine transporter, utilising aquaglyceroporin 2 (AQP2) as a receptor, and low-affinity pentamidine transporters contributing to cellular uptake (Franco, Scarone and Comini, 2018). These membrane transporters are also responsible for cellular nutrient uptake and are dispensable for the parasite, thus a loss-of-function mutation in corresponding genes have rendered the parasite resistant to pentamidine (de Koning,

2020). It has been noted that melarsoprol, Figure 2C, uses the same membrane transporters for cellular uptake indicating that the mutation will result in multi-drug resistance (Franco, Scarone and Comini, 2018). Melarsoprol is a trivalent arsenical compound, composed of a melaminophenylarsine moiety and a dimercaptopropanol moiety, making the drug highly toxic to patients with roughly 5% of patients receiving treating having a fatal encephalitic reaction. Genetic modifications underlying drug resistance shows large deletions of the AQP2 and AT1 sequences and/or point mutations in a single allele of the AT1 gene (Fairlamb and Horn, 2018). Single codon mutations of the RNA-upstream binding protein (UBP1), believed to regulate the function of mRNA levels, has been observed, but their precise role in drug resistance has not been determined. The multidrug-resistance protein A (MRPA), involved in the energy-dependent extrusion of compounds or active metabolites by ATP-binding cassette (ABC) transporters, has been demonstrated to confer melarsoprol resistance in *T. brucei* by active extrusion of melarsoprol-trypsnothione conjugates (Barrett *et al.*, 2011).

Eflornithine, Figure 2D, an α-difluorinated amino acid analogue, arose through rational drug design to specifically target the ornithine decarboxylase (ODC) enzyme as a mechanism-based suicide inhibitor, that upon decarboxylation binds covalently to a cysteine residue adjacent to the active site, irreversibly inactivating the enzyme (Franco, Scarone and Comini, 2018). ODC is the first enzyme in the polyamine biosynthetic pathway to yield spermidine, thus making it essential as polyamines modulate essential cellular processes such as proliferation, survival, differentiation, autophagy, stress tolerance and apoptosis. *T. bruei* relies mainly on de novo polyamine biosynthesis (Cullen and Mocerino, 2017). Eflornithine was initially developed as cancer chemotherapy, but due to the large adverse effect profile was not pursued for clinical application. Resistance to the drug is linked to amino acid transporter 6 (AAT6) which participates in the cellular uptake of eflornithine as indicated by gene deletion or silencing, generated resistant cell lines of *T. brucei* in laboratory experiments (Barrett *et al.*, 2011).

Nifurtimox, Figure 2E, is a nitroaromatic drug, which enters parasitic cells by passive diffusion. The drug is a prodrug requiring a one-electron reduction to a nitro anion radical by type II NADPH-dependent nitroreductase enzyme to be active. It has also been noted that type I nitroreductase (NTR-1) catalyses a two-electron reduction of the nitro group yielding cytotoxic reactive species (Dickie *et al.*, 2020). Furthermore, the activation of nifurtimox produces a nitroso and hydroxylamine group which then fragments and produces open-chain nitriles. These act as a Michael acceptor which form adducts with DNA and thiol groups in low molecular mass metabolites and proteins, resulting in DNA and macromolecule damage and thiol depletion within the parasite (Spaulding,

Gallerstein and Ferrins, 2019). Resistance is noted with the production of defective NTR-1 enzyme by rearrangement of the corresponding genomic locus. Several mutations in the components involved in DNA repair are observed proving the susceptibility to alkylating and interstrand cross-linking of the nifurtimox metabolites. It should be noted that although a relatively new drug, fexinidazole, Figure 2F, resistance will most likely occur along the same lines as for nifurtimox that being the alteration in the genes responsible for the NTR-1 enzyme (de Koning, 2020). Fexinidazole is a 5-nitroimidazole analogue, which undergoes the same intracellular activation as nifurtimox, forming reactive sulfoxide and sulfone metabolites, which have the same anti-trypanosomal action as nifurtimox (Franco, Scarone and Comini, 2018).

Suramin, Figure 2B, is one of the first drugs developed in a medicinal chemistry program in 1916 by Bayer. It is a sulphated naphthylamine compound presenting as a highly charged molecule at physiological pH, thus being unable to cross cell membranes by passive diffusion. Suramin possesses an extremely long half-life of approximately 40 days and this is due to it being highly plasma protein-bound. The exact mechanism of action of suramin is unknown, with theories speculating that suramin, being approximately 15% low-density lipoproteins (LDL) bound, alters cholesterol homeostasis of the parasite, as *T. brucei* exclusively obtains cholesterol from mammalian hosts (Zoltner *et al.*, 2020). The flagellum and flagellar pockets of trypanosomes are densely populated by low- and high-affinity LDL receptors to ensure efficient cholesterol uptake. An alternative to the disruption of cholesterol homeostasis is the inhibition of ATP production by glycolysis, due to electrostatic binding to certain glycolytic enzymes. Transport into the parasite is mediated by the parasite-specific invariant surface glycoprotein 75 (ISG75) and movement from the endosomes to the lysosome is facilitated by the adaptin complex 1 EMP70 (Spaulding, Gallerstein and Ferrins, 2019). Suramin resistance is conveyed in this regard, by the adaptation and alteration of the parasites variant surface glycoprotein (VSG). These alterations will result in the downregulation of the endocytic pathway preventing suramin from entering the parasite (de Koning, 2020). The parasite contains more than 1000 genes responsible for the formation, maintenance and alteration of the VSG during different life cycles, thus making it an ideal mutation target. It has been noted that suramin does not affect the parasite when it is in the insect stage of its life cycle, again this is due to alterations in the VSG (Cullen and Mocerino, 2017).

1.1.2 Salmonellosis

Salmonella species are Gram-negative bacilli associated with animal and human diarrhoeal infections. The species infection results in high morbidity and mortality rates in both developed and developing countries, having the worst impact on developing nations (Eng *et al.*, 2015). The standardised taxonomic classification has assigned two species to the *Salmonella* genus, namely *S.*

bongori and *S. enterica*. *S. entrica* further subdivides into six subspecies named: *enterica*, *salamae*, *arizonae*, *diarizonae*, *outenae* and *indica*; while *S. bongori* has no subspecies. *S. enterica enterica* has several thousand serotypes (estimated to be more than 2500) of which serotype *Typhi* (*S. typhi*) and *Paratyphi* (*S. paratyphi*) are clinically significant for Typhoid Salmonella (TS) while the remaining serotypes are responsible for Non-Typhoid Salmonella (NTS) (e.g. *S. enteritidis* and *S. typhimurium* to name two) (Tack *et al.*, 2020). The other subspecies of *S. enterica* are also responsible for the disease but to a lesser extent than *enterica* subspecies. Subspecies and their serotypes have developed to be host-specific with few having the ability to infect numerous species of animal such as the NTS strain *S. typhimurium* and its serotypes (*S. enteriditis* has a vastly decreased ability to infect multiple hosts) (Langridge, Wain and Nair, 2012).

It is estimated that *Salmonella* is responsible for 180 million infections, which is 9% of all diarrhoeal illnesses occurring globally each year. The number of deaths due to *Salmonella* is more than 298 thousand, with some estimating deaths due to *Salmonella* infection may be as high as 1 million, and accounts for approximately 41% of all diarrhoeal disease-associated deaths (Sánchez-Vargas, Abu-El-Haija and Gómez-Duarte, 2011)(Park *et al.*, 2019). The causative serotypes are distributed across the world with NTS serotypes being more common in Africa and TS serotypes more common in Southeast Asia. The majority of *Salmonella* cases (52% NTS and 37% TS) are believed to be foodborne, which is higher than other enteric pathogens. It is estimated that 79% of the global infections occur in sub-Saharan Africa and as many as 85% of the global mortalities (Smith, Seriki and Ajayi, 2016).

S. typhi and *S. paratyphi* have only humans as a reservoir and thus the mode of transmission is purely through ingestion of food or water contaminated with a carrier's faeces. Carriers are defined as having Salmonella positive stools for more than 12 months after initial acute infection. It is estimated that carriers of TS serotypes may continue to shed the bacterium for many decades, as they are responsible for all human infections in endemic areas as well as during outbreaks in non-endemic areas (Tack *et al.*, 2020). As many as 1-4% of patients infected with TS serotypes may become chronic carriers, as well as individuals who are infected but remain asymptomatic. It has been suggested that the liver houses the bacteria and intermittently excretes them into the gallbladder from where they are released together with gall into the intestines and passed out with faeces. A link has tentatively been made between chronic Salmonella carriers and carcinoma of the gallbladder (De Cesare, 2018).

NTS serotypes (*S. enteritidis* and *S. typhimurium*) have been isolated from the gastrointestinal tract of birds and mammals, including poultry and livestock. A faecal-oral route of transmission is the

main method of spread although a shared environment has also been linked to spreading through mechanisms such as touching an infected animal or coming in contact with animal blood or animal products. Even cheese has been implicated in the pathogen spread (De Lucia *et al.*, 2018). NTS serotypes have a broad spectrum of hosts and show a large degree of virulence amongst all host types. Slaughtering of animals in unhygienic facilities drives infections from animals to humans (Jaja *et al.*, 2019). It is worth noting that the bacterium can infect eggs from poultry. The bacterium infects the egg before the formation of the shell and continues to grow inside the egg once the shell has formed (MacDonald *et al.*, 2019).

Infection is initiated by the ingestion of Salmonella, where it travels through the stomach and attaches itself to specialised epithelial cells (M cells) of the Peyer patches lining the lumen of the small intestine by fimbriae or pili (Park *et al.*, 2019). Figure 3 shows *S. enterica* host infection cycle. The internalisation of the bacterium is achieved by receptor-mediated endocytosis, where they are transported within phagosomes to the lamina propria where they are released. The internalisation of the bacterium will induce an influx of macrophages (TS serotypes) or neutrophils (NTS serotypes) resulting in the further influx of inflammatory mediators such as cytokines resulting in the body increasing fluid secretion into the intestines in an attempt to flush out the pathogen (Tack *et al.*, 2020). Furthermore, Salmonella is capable of invading the intracellular environment of non-phagocytic cells. Virulence genes are involved in and required for intracellular survival and are clustered in large chromosomal DNA regions called Salmonella pathogenicity islands (SPIs). SPI-1 and -2 encode type III secretion systems which are multi-protein complexes that are used to develop a contiguous channel between the bacterial and epithelial cell membranes, resulting in translocation of bacterial effectors directly into the epithelial cell cytoplasm (Branchu, Bawn and Kingsley, 2018). Bacterial effectors will interact with host cell proteins to activate signal transduction pathways, rearranging the actin cytoskeleton leading to membrane ruffling and engulfment of the bacterium. Once inside cells, Salmonella resides within a membrane compartment known as the Salmonella-containing vacuole and actively remodels this vacuole for optimal survival and replication (Langridge, Wain and Nair, 2012). The bacterium can survive the intracellular environment of macrophages which it uses to travel to the spleen, lymph nodes and the entire reticuloendothelial system (Eng *et al.*, 2015).

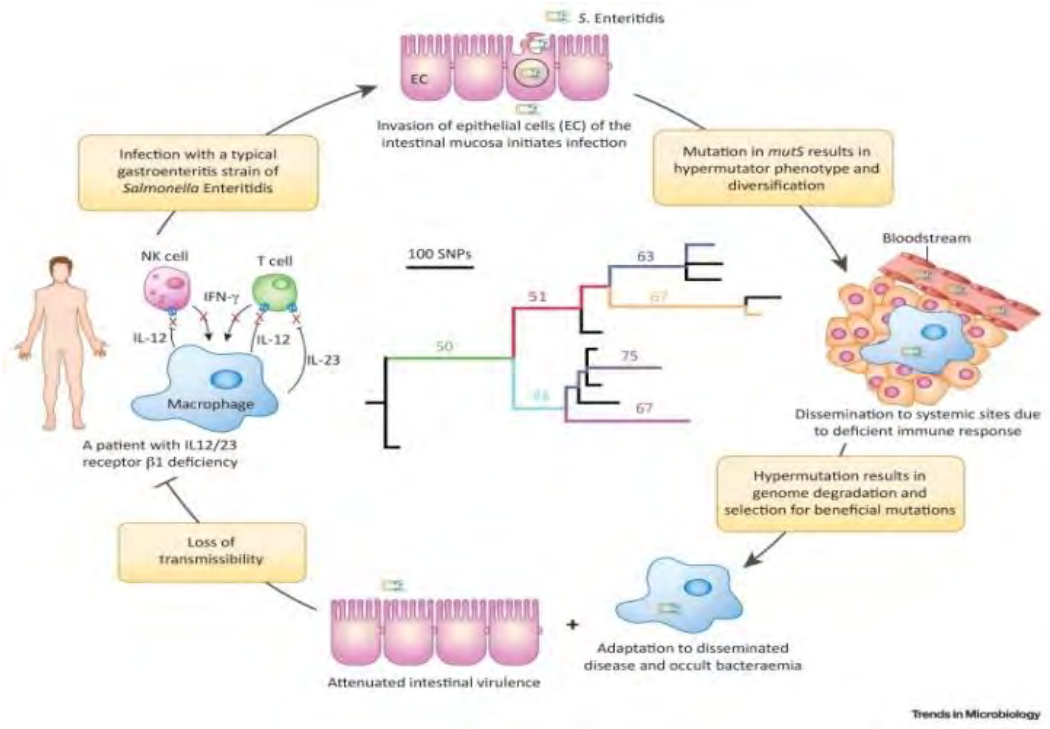


Figure 3: *S. enterica* infection within the host. Image obtained from Tanner and Kingsley (2018).

Enteric fever is caused by *S. typhi* or *S. paratyphi*, with a general incubation period of 7-14 days, but can range from 3-60 days. Before the onset of fever, the patient's symptoms may include headache, diarrhoea or constipation and abdominal pain, with other non-specific complaints including chills, loss of appetite, cough or myalgias. The disease initially has a low-grade fever followed by a sustained high fever during the second week of infection lastly for up to 4 weeks. The disease is more severe in children under the age of 1 year and patients with a compromised immune system (Olobatoke, 2017). Studies conducted in Tanzania have reported that the HIV status of a patient could provide an underdetermined protective measure against TS infection, as HIV-positive patients were less likely to develop bacteraemia than HIV-negative controls (Keddy *et al.*, 2017). NTS infections are associated with a different clinical syndrome, including but not limited to gastroenteritis and bacteraemia, with a short incubation period of 6-12 hours. Causative organisms include the serotypes of *S. enteritidis* and *S. typhimurium*. Common initial symptoms include nausea, vomiting and no bloody diarrhoea, with fever, chills and abdominal pain. The symptoms and the disease itself is usually self-limiting and represents a non-invasive NTS infection (Langridge, Wain and Nair, 2012). Invasion into the circulatory system, occurring via the mechanism described previously, commonly presents with bacteraemia of which the lungs are frequently compromised. Risk factors for the development of invasive NTS infection include immune system status, age extremes, decreased gastric acidity and recent use of antibiotics (Olobatoke, 2017).

Historically chloramphenicol, Figure 4A, has been the antibiotic of choice in the treatment of enteric fever (TS) infection, but resistance occurred and the treatment agent became ciprofloxacin, Figure 4B. With a decrease in susceptibility to ciprofloxacin and development of β-lactamase resistance, quinolones (fluoroquinolone), Figure 4C, azithromycin, Figure 4D, third-generation cephalosporins, Figure 4E, or trimethoprim-sulfamethoxazole, Figure 4F and G respectively, have become the treatment options for enteric fever (Tack *et al.*, 2020). The treatment of NTS does not recommend antibiotic therapy as the disease is generally self-limiting and antibiotic therapy tends to prolong or increase the patient's carrier state. However, in patients who are at risk for the development of invasive NTS, first-line therapy is the same as for enteric fever (Sánchez-Vargas, Abu-El-Haija and Gómez-Duarte, 2011).

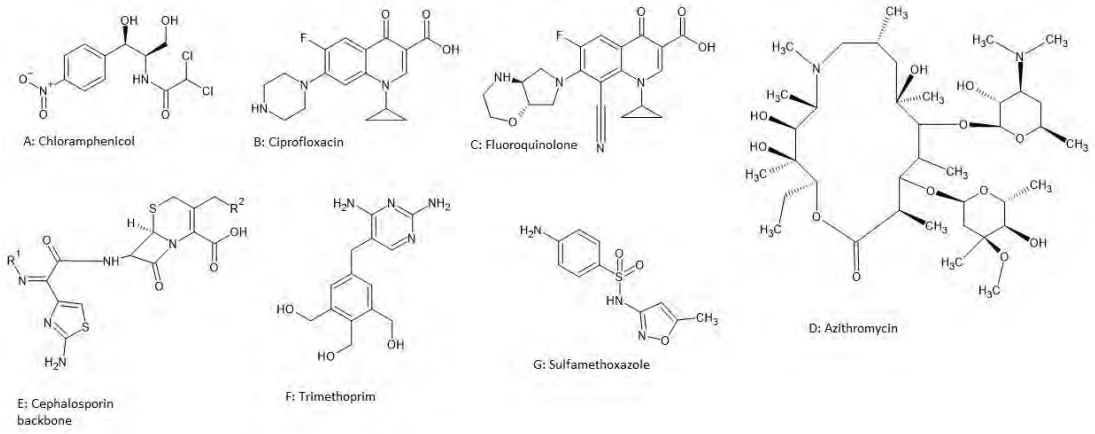


Figure 4: Previously used drugs in the treatment of *S. enterica*. A: Chloramphenicol, B: Ciprofloxacin, C: Fluoroquinolone, D: Azithromycin, E: Cephalosporin backbone, F: Trimethoprim, G: Sulfamethoxazole. Images redrawn using ACD/ChemSketch from Tack *et al.* (2020).

A serotype of NTS, known as definitive page type (DT104), shows resistance to ampicillin, Figure 5A, (interferes with cell wall synthesis by attaching to penicillin-binding proteins [PBPs] resulting in rupture of the cell wall and organism death), chloramphenicol, Figure 4A, (inhibits protein synthesis by preventing protein chain elongation via the inhibition of peptidyl transferase), streptomycin, Figure 5B, (irreversibly binds to the 16S rRNA and S12 protein within the bacterial 30S ribosomal subunit, thus interfering with the assembly initiation complex between mRNA and the bacterial ribosome), sulphonamides, Figure 5C, (structurally resemble ρ-aminobenzoic acid [ρABA], thus acting as a false substrate creating a non-functional folate) and tetracyclines, Figure 5D, (inhibits the 30S ribosomal subunit, hindering binding of the aminoacyl-tRNA to the acceptor site on the mRNA-ribosome complex, thus inhibiting protein synthesis), as well as nalidixic acid, Figure 5E, (inhibits a subunit of DNA gyrase and topoisomerase IV inducing formation of cleavage complexes) (Lynn *et al.*,

1998). *XbaI* and *BlnI* have been identified genes responsible for carrying resistance in Salmonella species. Additionally, multidrug-resistant serotypes were identified to have DNA substitution in the *gyrA* gene and these substitutions occurred in isolation from each other, e.g. a New Zealand serotype changed from GAG to AAC occurred at codon 87, whilst a Danish change occurred at codon 83 from TCC to TTC and a German serotype at the same position as the Danish serotype changed from TCC to TAC (Parkhill *et al.*, 2001). The current treatment for multidrug-resistant *Salmonella* infections is to use high doses of azithromycin (binds to the 50S subunit of the bacterial ribosome inhibiting translation of mRNA). Some evidence indicates that the use of β-lactam antibiotics, such as third-generation cephalosporins, Figure 4E, and carbapenems, Figure 5F, having a similar mechanism of action to penicillin, have success in treating resistant Salmonella strains (Braoudaki and Hilton, 2004).

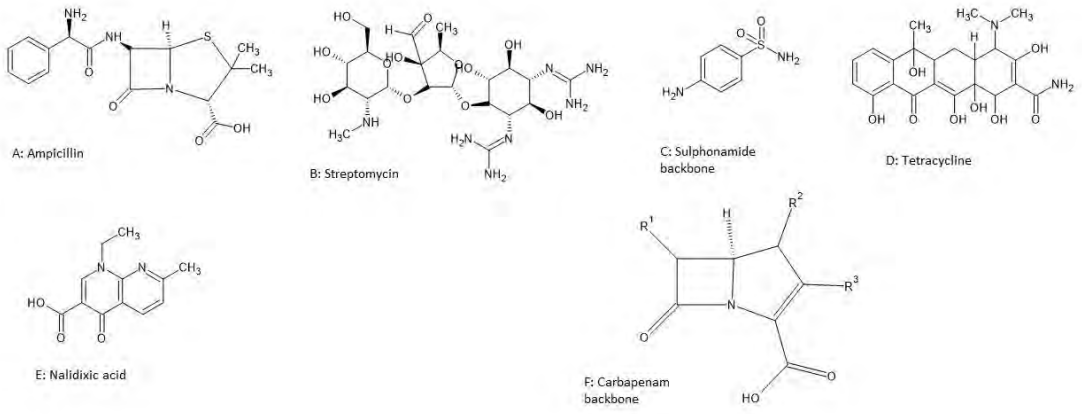


Figure 5: Currently used drugs in the treatment of *S. enterica*. A: Ampicillin, B: Streptomycin, C: Sulphonamide backbone, D: Tetracycline, E: Nalidixic acid and F: Carbapenem backbone. Images redrawn using ACD/ChemSketch from Lynn *et al.* (1998).

Antibiotic resistance is spread between serotypes as well as between bacterial species in several mechanisms. Conjugation between bacteria can result in the transference of genes found on plasmids or transporons. Transformation is the mechanism by which a bacterium takes up extracellular donor DNA, incorporating it into its own. Transduction occurs when a bacteriophage infects the recipient bacterium with foreign genes which are then incorporated into the bacterium's genome (Braoudaki and Hilton, 2004).

With the increase in resistance growing the need for alternative and safe therapeutic agents is increasing. The need for more stringent hygiene practices and education regarding such practices is shown by the high levels of infections due to Salmonella which can largely be prevented by hygiene

campaigns. Responsible farming practices and abattoir safety can decrease the number of infections, along with food safety measures being put in place specifically amongst street food vendors. Therapeutic agents targeting novel structures of pathways in the bacterium can greatly assist in the treatment of infections and possibly lower the alarmingly high mortality due to *Salmonella* infections.

1.1.3 Escherichiosis

E. coli is a Gram-negative facultative anaerobic bacillus commonly found as commensal flora (0.1%) in warm-blooded organisms. *E. coli* is the most extensively studied bacterium, with serotypes having been developed solely for research. Numerous diseases are caused by *E. coli*, including cholecystitis, bacteraemia, cholangitis, urinary tract infections and diarrhoea, with neonatal meningitis and pneumonia being a serious concern. Amongst the different serotypes, some are more suited to causing a specific disease, e.g. enterotoxigenic *E. coli* causing as many as 241 million cases of diarrhoea globally (Irek *et al.*, 2018). Many of the non-infectious serotypes play an important role amongst the autochthonous microbiome of hosts, providing a beneficial role to host organisms in that they can synthesise vitamin K and some of the B complex vitamins. The capacity of the organism to survive outside a host body is limited in the majority of serotypes, but a growing number of serotypes have developed the ability to survive in the environment for several days (Aijuka *et al.*, 2018). *E. coli* forms a large portion of foodborne diseases as it can infect numerous animals consumed by humans. Unsanitary food working conditions are a major cause of spread, as the pathogen spreads easily and is linked to faecal contamination of food and water and has been identified as causing a large portion of diarrhoeal diseases in livestock around the world. Resistant *E. coli* serotypes have also been identified as being one of the biggest sources of nosocomial infections, resulting in high patient deaths, especially post-operatively (Ingle *et al.*, 2018)(Malande *et al.*, 2019).

Diarrhoea causing *E. coli* strains in humans are classified as enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC), enteropathogenic *E. coli* (EPEC), enteroinvasive *E. coli* (EIEC), diffusely adherent *E. coli* (DAEC) or verocytotoxigenic *E. coli* (VTEC). One of the VTEC strains associated with diarrhoea, bloody diarrhoea, haemorrhagic colitis and haemolytic uremic syndrome (HUS) is the Shiga toxin-producing Escherichia coli (STEC) O157:H7. STEC O157:H7 has quickly spread to become a serious source of infection in food products and the environment (Lupindu, 2018). This serotype was initially rarely isolated from stool samples but has in the last decade increased in prevalence, with the ability to pass on its genetic material to other serotypes as well as other bacterial species increasing the number of high virulence bacteria. Factors such as low infective dose, the ability to express different virulence factors and long survival time in the environment

increase its difficulty to treat. This coupled with the widespread resistance genes amongst *E. coli* serotypes, could result in the development of a super, pan-resistant bacterium (Mbelle *et al.*, 2019).

Figure 6 shows how enterotoxigenic *E. coli* is spread. Serotypes possess the ability to infect numerous hosts whereby virulence and resistance genes can pass between infected animals and animals to humans via direct contact or food products, indicating the possibility that *E. coli* can serve as a reservoir of resistance genes, which spreads via horizontal gene transfer. A high number of *E. coli* serotypes have been directly linked to treatment failure through antibiotic resistance (Roussel *et al.*, 2017).

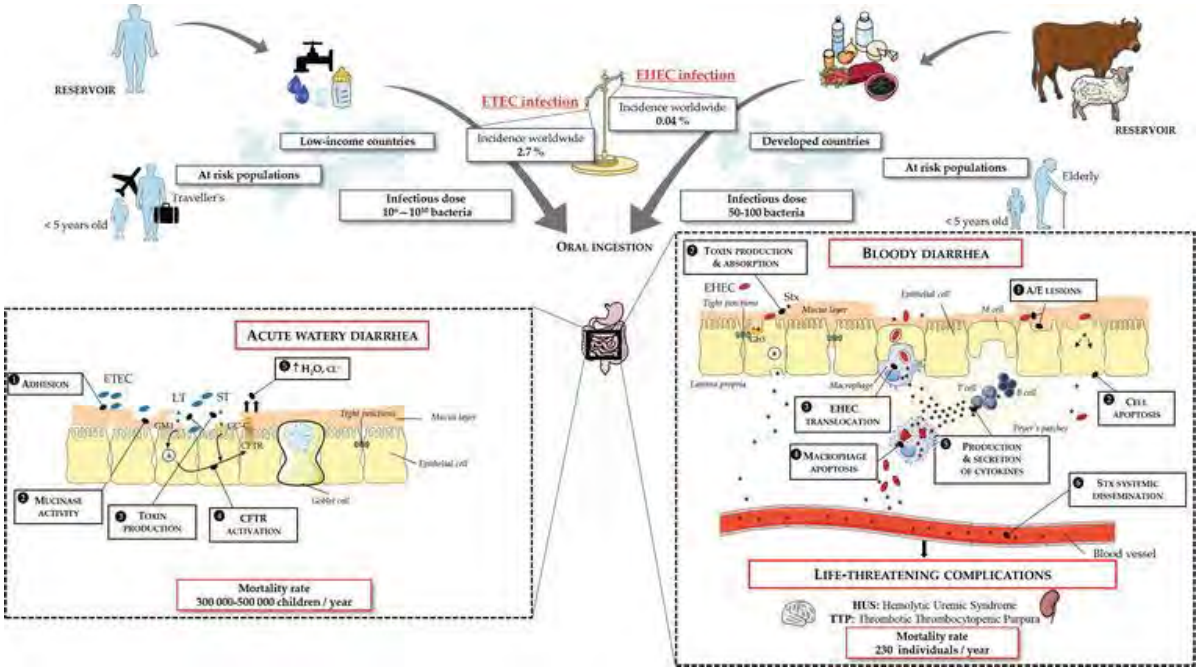


Figure 6: Spread of enterotoxigenic *E. coli*. Imaged obtained from Roussel *et al.* (2017).

The intestinal pathogenic *E. coli* serotypes are responsible for disorders in the gut ranging from mild diarrhoea to severe colitis, while the extra-intestinal pathogenic *E. coli* serotypes are mostly asymptomatic inhabitants of the intestinal tract that cause extra-intestinal diseases after migrating to other parts of the body, such as the urinary tract or the bloodstream. The number of genes present in the *E. coli* genome varies from 4,000 to 5,000 genes, with approximately 3,000 genes shared by the different serotypes, whereas the others mostly correspond to colonization or virulence determinants. These ETEC serotypes bind and colonize the intestinal epithelium through adhesins such as the F4, F5, F6, F17, and F18 fimbriae. ETEC also produces various enterotoxins, of which enteroaggregative heat-stable toxin 1 (EAST1) results in diarrhoea (Aijuka *et al.*, 2018).

Human and animal genes are present in *E. coli* which confer resistance to β-lactams such as $bla_{\text{TEM-1}}$, coming from animals, it codes for the inactivation of penicillins and aminopenicillins, $bla_{\text{CTX-M-1}}$, $bla_{\text{CTX-}}$

$_{M-14}$, *bla*$_{TEM-\ 52}$, and *bla*$_{SHV-12}$ genes coding for extended-spectrum β-lactamases (ESBLs) or overexpressed cephalosporinases (AmpCs) and recently carbapenem-resistant genes have been identified, all having been identified as being introduced to the bacterium from mammalian genetic material (Mbelle *et al.*, 2019).

Quinolones and fluoroquinolones, refer to Figure 4C, which have been known to be bactericidal against effectively all bacteria, treating a large number of infections in humans, have been identified as ineffective against *E. coli* serotypes, which resulted from mutations in the drug target (Tuem *et al.*, 2018). The genes coding for DNA gyrase and topoisomerase IV have been altered as well as the bacterium having reduced permeability of the cell membrane and upregulation of efflux pumps. Topoisomerase IV consists of four subunits, two ParC and two ParE subunits (Mbelle *et al.*, 2019). The majority of the mutations appear in the quinolone resistance determining region in GyrA. Initially, the plasmid-mediated quinolone resistance (PMQR) determinant, qnrA1, was discovered in 1997 and has since become a grave concern in the spread of PMQR genes. Mechanisms conveying resistance via plasmids have since been identified which include, Qnr-like proteins (QnrA, QnrB, QnrC, QnrD and QnrS) which prevent quinolone from binding to DNA, the AAC(6')-Ib-cr acetyltransferase that alters the structures of certain fluoroquinolones, ciprofloxacin and enrolfloxacin, and efflux pumps which actively excrete quinolones from cells (QepA and OqxAB). These alterations do not confer overall resistance, but rather reduces the susceptibility of the bacteria to quinolones and when multiple resistance genes are encoded for then absolute quinolone resistance is conferred (Poirel *et al.*, 2018).

Aminoglycosides, Figure 7, which inhibit protein synthesis by binding to ribosomal subunits, have been noted to be less susceptible to resistance than other antibiotics. Resistance develops through point mutations in the genes encoding for the 16S RNA, either the S5 or S12 ribosomal proteins (Mbelle *et al.*, 2019). The effectiveness of this strategy would be high only if the bacterium has a limited number of 16S RNA encoding operons. Fortunately, *E. coli* has several copies of the operon, which largely reduces its chances to develop aminoglycoside resistance through this mechanism. Aminoglycosides are rather inactivated by enzymes, which fall under a group called aminoglycoside-modifying enzymes (acetyltransferases, nucleotidyltransferases, and phosphotransferases), resulting in aminoglycosides being unable to reach their target (Tuem *et al.*, 2018).

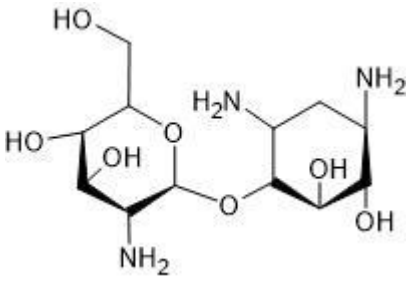


Figure 7: Structure of aminoglycoside. Image redrawn using ACD/ChemSketch from Mbelle *et al.* (2019).

Resistance to tetracycline, refer to Figure 5D, results from two mechanisms, namely active efflux out the bacterial cell and ribosomal protection. Nine tetracycline efflux genes, and two tetracycline resistance genes encoding ribosome protective proteins and one gene coding for an oxidoreductase that inactivates tetracyclines have been identified in *E. coli* and all belonging to the *tet* family (Mbelle *et al.*, 2019). Resistance to anti-folates, sulphonamides, refer to Figure 5C, and trimethoprim, refer to Figure 4F, have been identified in which the bacteria modify the target enzymes (DHFR or dihydropteroate synthase[DHPS]) resulting in an enzyme capable of performing its function in an alternative manner. The genes responsible for these modifications are the *sul* (*sul* 1-3) genes encoding for DHPS and the *dfr* (A and B) genes encoding DHFR (Montso *et al.*, 2019).

The resistance profiles discussed although the spread between the various *E. coli* serotypes can be spread between serotypes and other bacteria. From this, it is noted that *E. coli* serves as a reservoir for many resistance genes and coupled with their wide presence in the animal kingdom, can spread antimicrobial resistance not only to human pathogens but also to animal pathogens. Appropriate use of antimicrobial agents, patient education and hygienic practices with food products will greatly assist to curb the spread of resistance genes, but the large resistance profile exhibited by *E. coli* indicates the need for novel therapeutic agents by an alternative mechanism, targeting essential pathways in bacteria. Resistance to antimicrobials targeting specific enzymes within the folate biosynthetic pathway has been discussed, thus making the need for alternative novel agents targeting different enzymes a high priority in the drug discovery process, as the folate biosynthetic pathway is a validated pathway.

1.2 Folate Biosynthetic pathway

The enzymes catalysing the folate biosynthetic pathway have been described as a potent source of drug targets, especially in prokaryotes and lower eukaryotes, as they are unable to source folate from their environment and thus synthesise it *de novo* (Bermingham and Derrick, 2002). Tetrahydrofolate (THF), the ultimate product in the pathway is used for many biological functions, serving as a critical co-factor by shuttling methyl and formyl groups which are utilised in one-carbon

transfer reactions. These reactions synthesise purines, amino acids, S-adenosylmethionine and formyl-methionine. Inhibition of the folate biosynthetic pathway will result in the cessation of cellular replication by the starvation of the cell of essential components, thus resulting in effective anti-cancer, anti-plasmodial and anti-bacterial treatments (Bertacine Dias *et al.*, 2018). Prokaryotes synthesise THF *de novo,* from the starting products, guanosine triphosphate (GTP), ρ-aminobenzoic acid (ρABA) and glutamate. These will form the pterin portion of THF and ultimately THF through the action of several enzymes. In bacteria, archaea, fungi, algae, some protozoans and plants, the shikimate pathway is utilised for the production of GTP. It is worth noting that some organisms can both synthesise folate *de novo* and salvage folate from their environment, as is the case with the malaria parasite which obtains folate from the serum of an infected host. Observations indicate that multiple enzymes within the folate biosynthetic pathway can be inhibited with the use of a single compound as has been noted to occur with 8-mercaptoguanine (discussed in a subsequent section) (Bourne, 2014). Figure 8 shows an overview of the *de novo* folate biosynthetic pathway.

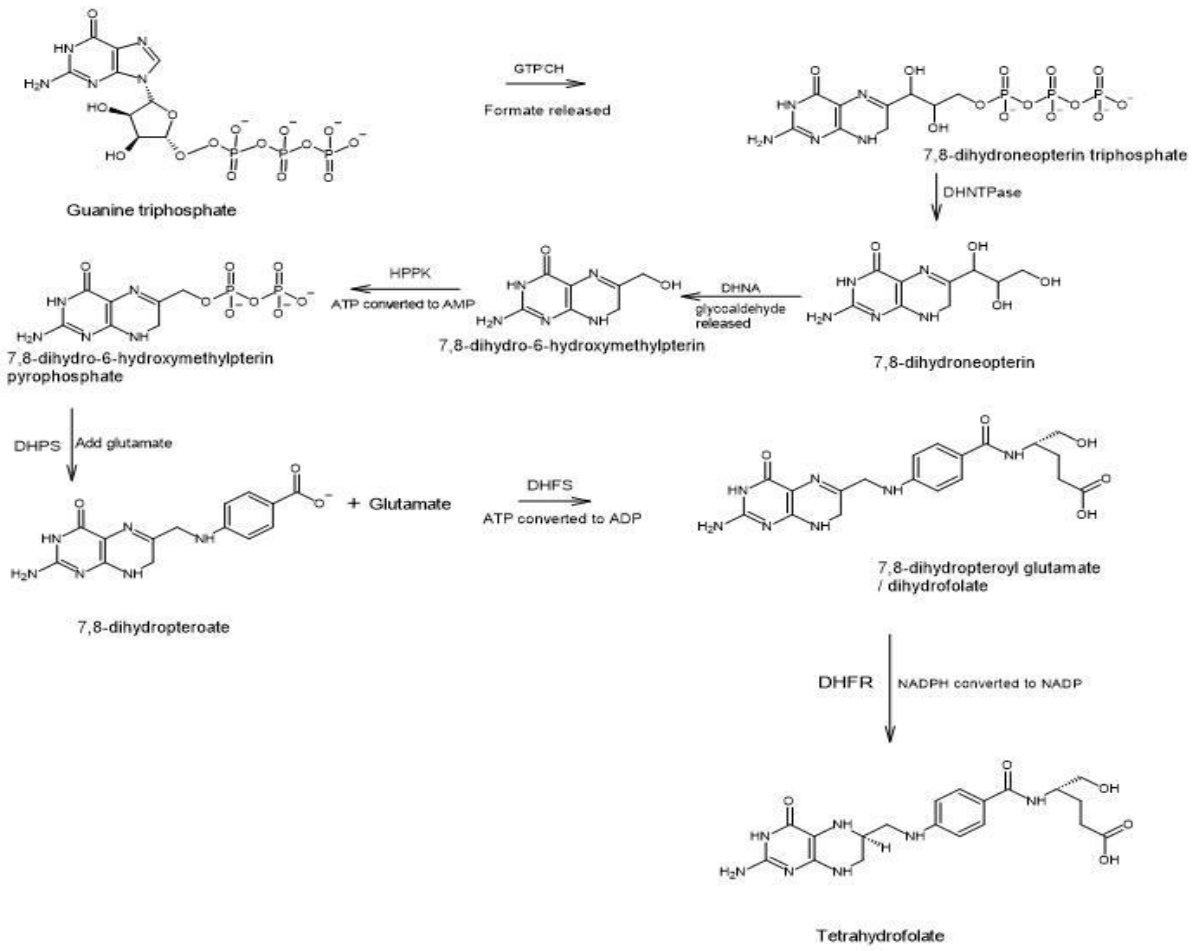


Figure 8: Overview of *de novo* folate biosynthetic pathway of *E. coli*. Image redrawn and edited using ACD/ChemSketch from Swarbrick *et al.* (2009) and Bourne (2014).

1.2.1 GTP Cyclohydrolase (GTPCH 1)

Initiation of pterin synthesis starts with GTP which undergoes ring-opening, rearranging and reclosing forming dihydroneopterin 3'-triphosphate (DHNP) and is catalysed by GTPCH 1 which is encoded by the *folE* gene. The reaction is considered the most complicated and is the rate-limiting step. Initially, the guanine imidazole ring is broken, after which cleavage at the N9-C8 bonds occurs producing the N-formyl pyrimidine intermediate and additional release of formate from C8, while the ribose moiety undergoes an Amadori rearrangement producing the dihydropyrazine ring. The dihydropyrazine ring will recyclize by a condensation reaction to provide the pteridine ring moiety (Bourne, 2014). GTPCH 1 is conserved in bacteria, protozoa, fungi, plants and vertebrates, although its product is a substrate in more than one pathway among different organisms. In bacteria, fungi and plants the enzyme catalyses the first step in the de novo biosynthetic folate pathway, while in vertebrates it catalyses the biosynthesis of tetrahydrobiopterin, a key co-factor for nitric oxide-producing enzymes, melanin and neurotransmitters (Bermingham and Derrick, 2002). GTPCH 1, composed of a pentameric beta-barrel surrounded by alpha-helices with an active site contribution from three subunits, requires a zinc cation which is coordinated by two cysteines and one histidine residue and ordered water molecules to function. Competitive inhibition of the bacterial enzymes, in in vitro studies, has been confirmed using derivatives of 8-oxoguanine, by mimicking the transition state configuration. These derivatives were found to interact with a cavity of the enzyme by using the same interactions as those observed with pterin-binding proteins. The by-product of this reaction is formic acid (Bertacine Dias *et al.*, 2018). The reaction catalysed by GTPCH 1 is shown in Figure 9.

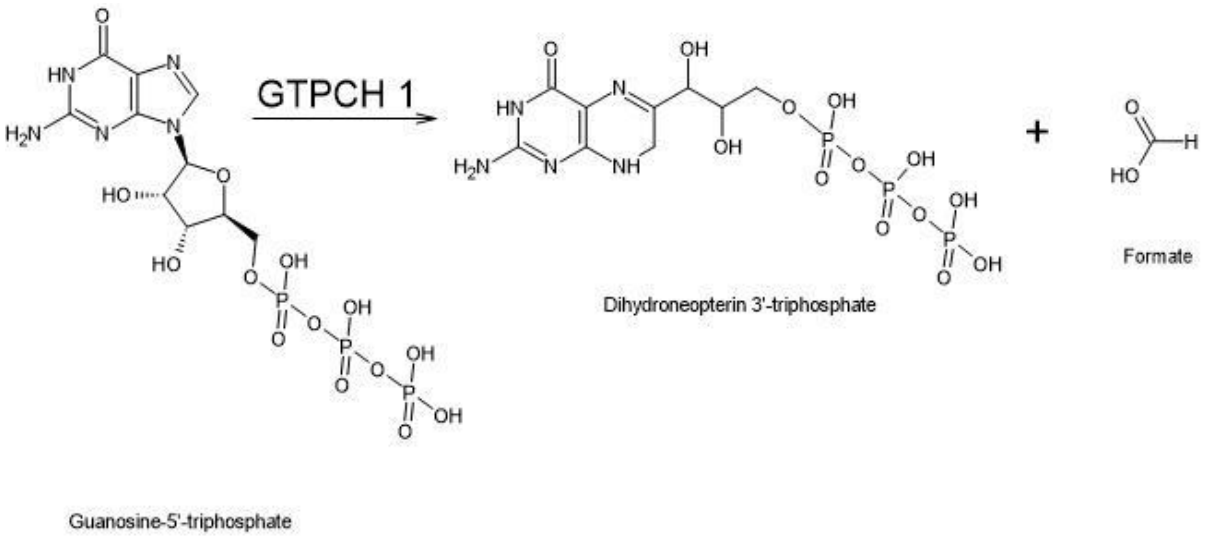


Figure 9: GTPCH 1 enzyme activity. Image redrawn using ACD/ChemSketch from Bourne (2014).

1.2.2 7,8-Dihydroneopterin triphosphate epimerase (DHNE)

This poorly characterised enzyme performs the essential epimerase activity in several bacterial species such as *E. coli* and *P. aeruginosa*, which is encoded by the *folX* gene. DHNE catalyses the inversion of the 2' carbon of the 7,8-dihyfroneopterin triphosphate, which is the first intermediate in the folate biosynthetic pathway (Bermingham and Derrick, 2002). Recent studies have indicated that DHNE can catalyse the reaction performed by dihydroneopterin aldolase (DHNA), but at a much slower rate and requires a divalent metal cation such as magnesium (Mg^{2+}) or manganese (Mn^{2+}). DHNE consists of four monomers packed together, producing a cylindrical core consisting of a 16-stranded antiparallel β-barrel. An octamer is formed by stacking two tetramers head-to-head thus producing potentially eight active sites. The α/β folds contained in DHNE are very similar to other pterin-binding enzymes, such as GTPCH 1 and DHNA, although they do not share significant sequence homology, except having a conserved glutamate residue, seeming to have a common pterin ring recognition function (Bertacine Dias *et al.*, 2018). The exact function and mechanism of DHNE are unclear but a possible catalytic mechanism is suggested in which epimerisation begins with the protonation of the N5 of the pterin ring. This results in an imin-enamine tautomerisation step which increases the acidity of the C2' hydroxyl group proton. An abstraction of the hydroxyl group from the C2' atom then occurs which leads to rearrangements and ultimately cleavage of the C1'-C2' bonds. From gene knockout studies performed in *E. coli*, lack of DHNE proved to be non-lethal to the organism and thus proves to be unlikely to be an effective drug target. The non-lethality in *E. coli* does not guarantee non-lethality in other organisms and further research into DHNE needs to be performed (Bourne, 2014). The reaction catalysed by DHNE is shown in Figure 10.

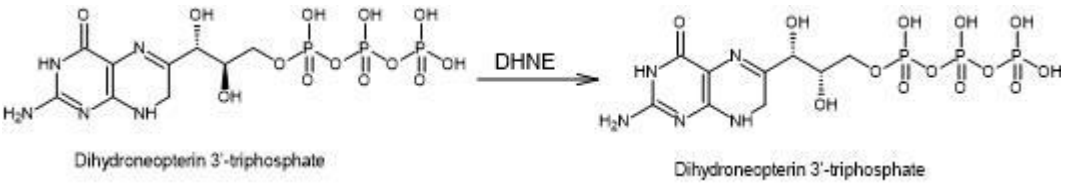


Figure 10: DHNE enzyme activity. Image redrawn using ACD/ChemSketch from Bourne (2014).

1.2.3 Dihydroneopterin aldolase (DHNA)

The reversible conversion of DHNP to 6-hydroxymethyl-7,8-dihydropterin (HMDP) is catalysed by DHNA, with glycolaldehyde as a by-product. DHNA is reported to catalyse the epimerization at the 2'-carbon converting 7,8-dihydromonapterin to its substrate 7,8-dihydroneopterin and recently the DHNA enzyme from *Mycobacterium tuberculosis* was confirmed to have oxygenase activity and thus produce 7,8-dihydroxyxanthopterin (Bourne, 2014). The enzyme is encoded by the gene *folB* and is composed of a four-stranded antiparallel beta-sheet with two long helices along one side of each monomer. Oligomerization results in a tetramer with a 16-stranded beta-barrel with longer alpha-

helices surrounding the outside. The active sites, composed of residues from two monomers and located externally to the beta-barrel, bind the pterin substrate by specific hydrogen bonds from the main polar atoms of the enzyme and acidic glutamic acid and/or aspartic acid residues. It has been found that *Pneumocystis jiroveci (P. jiroveci)* has a bifunctional enzyme system, in which DHNA is fused to 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK) allowing for increased turnover of the product (Bermingham and Derrick, 2002). Three specific hydrogen bonds contributed by the main chain polar atoms are responsible for pterin binding, in addition to acidic glutamic acid and/or aspartic acid. The catalytic site is distal to the pterin-docking cavity, in which the hydrophobic areas of the heterocycle are sandwiched between hydrophobic residues increasing binding affinities (Bourne, 2014). Drug discovery has identified potential inhibitors based on pterin or pyrimidine analogues to anchor into the pterin and pyrimidine binding pockets. In an extensive campaign, Sanders *et al.* (2004)identified 4 lead compounds of which further optimisations lead to the improvement in IC_{50} from 28 µM to 1.5 µM and finally 68 nM. Unfortunately the most potent *in vitro* inhibitor failed to elicit inhibition of bacterial growth. Apicomplexan parasites lack both *folB* and *folX* genes and thus it has been proposed that the enzyme 6-pyruvoyltetrahydropterin synthase bypasses the activity of DHNE and DHNA (Bertacine Dias *et al.*, 2018). The reaction catalysed by DHNA is shown in Figure 11.

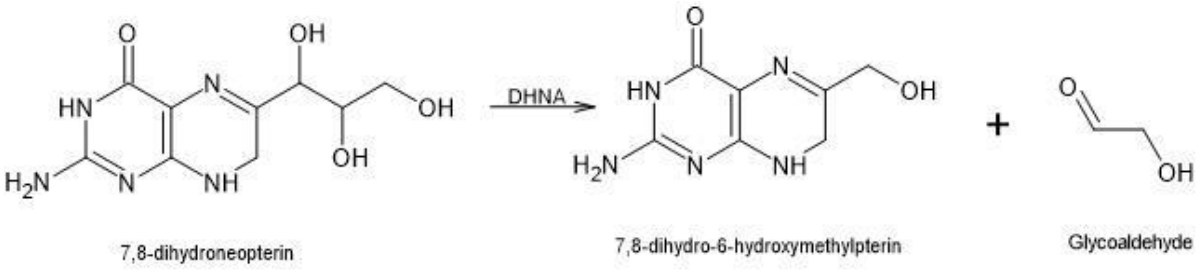


Figure 11: DHNA enzyme activity. Image redrawn using ACD/ChemSketch from Bourne (2014).

1.2.4 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase (HPPK)

HPPK, an ATP-binding enzyme, transfers a pyrophosphoryl group from ATP to HMDP, the product of DHNA, to produce hydroxymethyl-7,8-dihydropterin pyrophosphate (HPPP) and AMP. It is a metalo-coordination enzyme requiring Mg^{2+} for the transfer of the pyrophosphate moiety. The pyrophosphate is transferred at the beta phosphate rather than the alpha phosphate, making it rather unique amongst pyrophosphate kinases (Bourne, 2014). The monomeric protein contains a thioredoxin-like fold composed of a four-stranded antiparallel beta-sheet with helices added to each face. HPPK requires an ordered binding series, starting with an Mg^{2+}-ATP unit binding and triggering the movement of a loop (L3), exposing the catalytic binding site. L3 movement results in the ordering of critical arginine residues which completes the HMDP binding site. Bound pterin moieties

will stabilise the binding site and subsequent closure of L1 and L3 around the site. For further catalysis to occur the product HPPP needs to be completely released from the catalytic site. It has been estimated that the cycle has six steps, including transitional states (Bertacine Dias *et al.*, 2018). Interaction between HPPK and the pterin moiety include hydrogen bonding and aromatic residues stacking on the pterin heterocycle. A notable inhibitor, 8-mercaptoguanine (8MG), contains the pterin moiety and has an IC_{50} of 41 µM. 8MG was also shown to competitively inhibit dihydropteroate synthase (DHPS), and investigations using *Saccharomyces cerevisiae* HPPK-DHPS bifunctional enzyme indicated that both enzyme's active sites were being inhibited by the pterin monophosphate moiety of 8MG, lending support to the multiple enzyme inhibition by a single compound (Bermingham and Derrick, 2002). The reaction catalysed by HPPK is shown in Figure 12

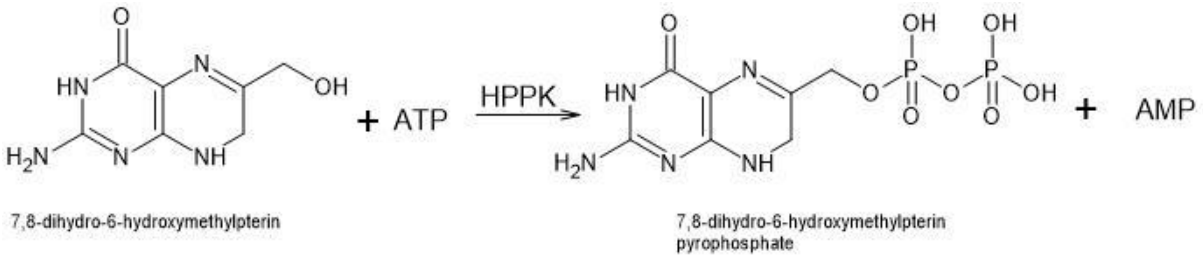


Figure 12: HPPK enzyme activity. Image redrawn using ACD/ChemSketch from Bourne (2014).

1.2.5 Dihydropteroate synthase (DHPS)

DHPS catalyses the condensation reaction of HPPP with *para*-aminobenzoic acid (ρABA) yielding 7,8-dihydropteroate, through the formation of a carbon-nitrogen bond. The enzyme is a dimeric triosephosphate isomerase-type barrel protein consisting of eight parallel beta-strands which are surrounded by eight alpha-helices. The active site of the enzyme is situated at the C-terminal of the central β-sheet, where the loops influence substrate access. This area exhibits a high degree of sequence diversity (Bourne, 2014). The HPPP binding site lies deep within a cleft which causes the plane ring system of HPPP to sit parallel to the axis plane of the eight-stranded β-sheet. The dihydropterin moiety binds in a narrow hydrophilic pocket. It is noted that binding occurs in an ordered manner, in which the pterin substrate is required to bind before ρABA binding. DHPS forms the junction between the folate biosynthetic pathway and the chorismate pathway, which produces ρABA. The reaction is initiated by the removal of the pyrophosphate moiety independently from the condensation reaction before the addition of ρABA. Magnesium is required for the stabilization and release of pyrophosphate (Bermingham and Derrick, 2002). A well-known antimicrobial drug class that interacts with the DHPS enzyme is sulphonamides. The sulpha drugs act as an alternative substrate to ρABA and results in a dead-end dihydropterin-sulphonamide product. Sulpha antimicrobials inhibit the action of mammalian sepiapterin reductase, which is critical for nerve cells

to produce tetrahydrobiopterin which serves as a co-factor for neurotransmitter synthesis, thus increasing the need for antimicrobials with lower adverse effects. Numerous mutations in the DHPS sequence have resulted in resistance to sulphonamide therapy and these relate to the ρABA binding site (Bertacine Dias *et al.*, 2018). The reaction catalysed by DHPS is shown in Figure 13.

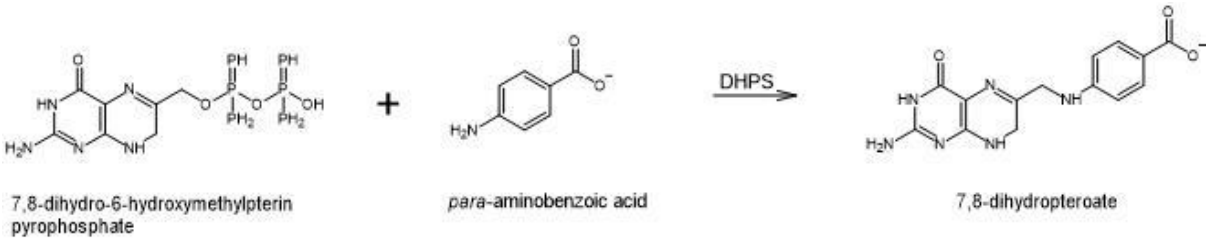


Figure 13: DHPS enzyme activity. Image redrawn using ACD/ChemSketch from Bourne (2014).

1.2.6 Dihydrofolate synthetase (DHFS)/Folylpolyglutamate synthetase (FPGS)

DHFS attaches an L-glutamate tail to dihydropteroate, the product of DHPS, in an ATP-dependent reaction. The negatively charged glutamate moiety bound to the folate prevents diffusion of the folate compound through cell membranes and additionally facilitates the one-carbon transfer in downstream enzymes. Further polyglutamylation can occur once reduction of the DHFR substrate to tetrahydrofolate has occurred and is carried out by a ubiquitous FPGS activity (Bourne, 2014). The initial glutamination is performed by DHPS, producing dihydrofolate, whereupon further glutamate moieties are added by FPGS with the number of added glutamates differing between organisms. The pterin binding site in DHFS lies between the ATPase domain, which conforms to the Mur synthetase superfamily and a Rossmann-fold domain, alongside the ATP binding site which is stabilised by the interaction with water molecules. The phosphate groups are coordinated with conserved residues and two magnesium ions which neutralise the negative charges. The heterocycle of the pterin is secluded within a cavity, stacked between hydrophobic residues, and the binding site which is poorly ordered only becomes ordered when it is occupied (Bermingham and Derrick, 2002). Shared activity between DHFS and FPGS within a single protein will results in separate inhibitory profiles for each of the separate activity portions of the protein. Since FGPS is active in mammalian cells it is an important target for cancer chemotherapy using agents such as methotrexate and pemetrexed, blocking folate biosynthesis through inhibition of polyglutamylation. A feedback mechanism has been identified in which the buildup of DHFR product, dihydrofolate, will result in inhibition of DHFS, from which novel folate analogues have been generated (Bertacine Dias *et al.*, 2018). The reaction catalysed by DHFS is shown in Figure 14

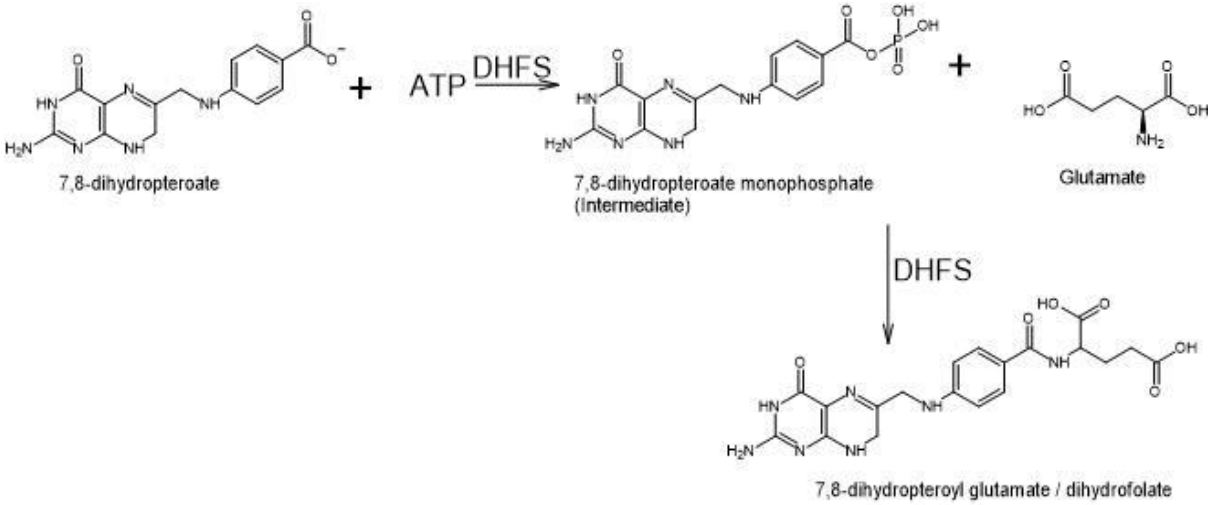


Figure 14: DHFS enzyme activity. Image redrawn using ACD/ChemSketch from Bourne (2014).

1.2.7 Dihydrofolate reductase (DHFR)

DHFR reduces dihydrofolate to tetrahydrofolate in an NADPH-dependent reaction in which the NADPH wraps around the outside of the enzyme whilst the dihydrofolate and nicotinamide moiety of NADPH are held adjacent in their respective binding pockets. A proton is donated from the NADPH to the N5 position of dihydrofolate with concomitant hybrid transference from NADPH to the C6 pterin ring. DHFR consists of an eight-stranded β-sheet flanked by four α-helices, containing two structural subdomains, namely the adenosine binding and loop subdomains, involved in conformational changes during the enzymatic catalytic cycle (Bourne, 2014). Studies have revealed that three key loops are responsible for the conformational changes which occur at the above-mentioned subdomains and have been named Met20, F-G and G-H, with Met20 having a further four different conformations, namely occluded, closed, open and disordered occurring at different times during catalysis. DHFR is the most targeted enzyme in the folate pathway. The best-known inhibitor is methotrexate, a substrate mimetic, which inhibits all known DHFR enzymes (Bermingham and Derrick, 2002). Currently, eight inhibitors fall into two classes, namely classical and non-classical based on the presence of a polyglutamate moiety. Classical inhibitors have the polyglutamate moiety and are typically seen in anti-cancer agents, allowing uptake into cells. Non-classical inhibitors are most often noted in antimicrobials and rely on cell entry via diffusion, which has made them susceptible to the action of efflux pumps mainly in Gram-negative bacteria. Agents such as trimethoprim and pyrimethamine are used to target bacterial, protozoal and fungal infections via inhibition of the DHFR enzyme (Bertacine Dias *et al.*, 2018). The reaction catalysed by DHFR is shown in Figure 15

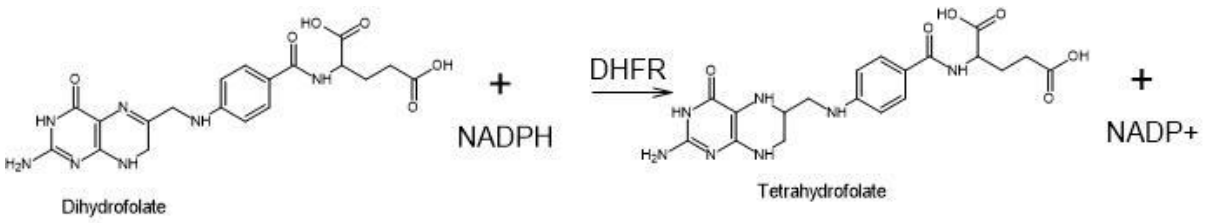


Figure 15: DHFR enzyme activity. Image redrawn using ACD/ChemSketch from Bourne (2014).

1.3 Evaluation of the HPPK enzyme

HPPK is responsible for the conversion of HMDP (Figure 16) to HPPP by the transfer of a pyrophosphate moiety from ATP, refer to Figure 12. Mg^{2+} is required for the successful catalysis of this reaction. A large degree of conservation is observed in the amino acid sequence of isozymes of HPPK as can be seen by the similarity between *E. coli* and other organisms such as *Haemophilus influenzae* (*H. influenzae*) (56%), *Pneumocystis carinii* (*P. carinii*) (40%) and *Mycobacterium tuberculosis* (*M. tuberculosis*) (33%) (Gao *et al.*, 2015). This high degree of similarity can thus be used to provide structural information derived from *E. coli* HPPK in the development of novel broad-spectrum antimicrobials. The HPPK crystal structure of *E. coli* reveals 158 amino acid residues formed in a three-layered αβα fold formed by six β strands and four α helices in the sequence β1-α1-β2-β3-α-2β4-β5-β6-α3-α4. In the center, the six β strands are organised into antiparallel β sheets and a β hairpin (Marimuthu, Singaravelu and Namasivayam, 2017). A valley is created in an unbound state measuring approximately 10 Å wide, 10 Å deep and 26 Å long, containing the catalytic binding centers of the enzyme. This valley is partially sealed when bound with an HMDP-MgAMP complex leaving an open end. This valley is formed by functional flexible loops, which assume differing conformational states depending on the substrate or ATP binding. Loops 2 and 3 fold into the valley covering the catalytic binding center in the presence of two Mg^{2+}. ATP and HMDP when bound in the catalytic centers, results in loop conformational changes which cover the catalytic centers preventing further interaction whilst catalysis is underway (Bermingham *et al.*, 2000). To date, it appears as though 13 amino acid residues are conserved throughout several organisms and these residues are located within the three flexible loops. These loops are stabilised by hydrostatic forces and also stabilise the HPPK-HMDP-MgATP complex. A reaction pathway based on current thermodynamic and kinetic data consists of the following steps: 1) apo-enzyme, 2) binding of MgATP, 3) binding of HMDP, 4) formation of the reaction intermediate HPPK-MgATP-HMDP, 5) transfer of the pyrophosphate to form HPPK-MgAMP-HPPP, 6) release of MgAMP, 7) release of HPPP (Li *et al.*, 2006).

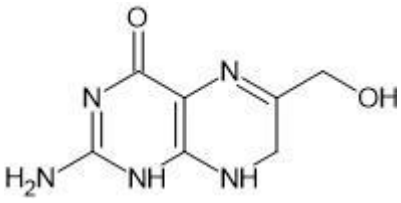


Figure 16: Structure of HMDP. Structure redrawn using ACD/ChemSketch from Bourne (2014).

An inhibitor, 8MG, Figure 17, was found to be effective against *E. coli* HPPK, identified by *in silico* docking studies, and was explored as a potential inhibitor of other HPPK enzymes such as the HPPK enzyme from *Staphyllococcus aureus* (*S. aureus*). It was noted that 8MG and HMDP have similar chemical moieties and are therefore worth exploring in *in vitro* plate-based assays. 8MG and HMDP share the same pyrimidine ring which forms hydrogen bonds with conserved residues Threonine 43, Alanine 44, Valine 46 and Aspartic acid 56 (Chhabra *et al.*, 2012). The pyrimidine ring of 8MG is stacked between Phe54 and Phe123 whilst a small cavity near N7 of 8MG forms a hydrogen bond with a water molecule and the side chain of Asp95. Two interactions that form between HMDP and HPPK are lacking in the case of 8MG, namely the imidazole ring of 8MG is too far away to interact with Asp95 which forms hydrogen bonds with the hydroxymethyl group of HMDP and van der Waals forces interacting with Trp89 is also missing. 8MG was determined to have an approximate dissociation constant (K_D) of 160 µM for *E. coli* HPPK and a K_D of 12.6 µM and an IC_{50} of 41 µM for *S. aureus* HPPK (Chhabra *et al.*, 2013). It is noted that *in vivo* growth inhibition did not occur with 8MG when performed on *E. coli* cell-based assays most likely due to poor membrane permeability, which may be disadvantageous for pterin-like molecules. Chhabra et al. found that 8MG does not require Mg^{2+} or ATP as a cofactor for competitive binding to the enzyme and that the catalytic binding sites on Loop2 and 3 become more rigid when 8MG binds to the enzyme (Yun *et al.*, 2014).

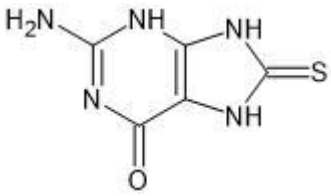


Figure 17: Structure of 8-mercaptoguanine (8MG). Structure redrawn using ACD/ChemSketch from Chhabra *et al.* (2013).

Dennis et al. (2016) synthesised a series of S-functionalised 8MG analogues as substrate site inhibitors of *E. coli* and *S. aureus* HPPK and quantified their binding affinities. They found that acetophenone-based substituents were similar for the enzyme of both organisms, while benzyl-substituted 8MG derivatives proved more selective for *S. aureus* HPPK due to its binding to a unique enzyme pocket with the enzyme. In particular, one benzyl-substituted 8MG derivative containing a

methyl group at the para position of the benzyl moiety proved to have particularly strong binding affinities to both *E. coli* (1.76 µM) and *S. aureus* (120 nM), providing a strong lead compound for further development into a potentially broad-spectrum therapeutic agent. The success of this agent will allow an antimicrobial capable of inhibiting both Gram-negative and Gram-positive organisms (Dennis *et al.*, 2016).

8-thioguanine (8TG), Figure 18, was shown to possess the inhibitory activity and further investigations into 8TG derivatives resulted in positive results from aryl-substituted 8TG compounds, in which structural analysis showed the compounds engaged with the pterin-binding pocket of the HPPK enzyme and again with the unique binding pocket. Although binding affinities and IC_{50} were substantially weaker than those observed for 8MG analogues, K_D 9.7 µM and IC_{50} of 139 µM for *S. aureus* HPPK, they prove promising in the light of potential novel agents worth further exploration (Yun *et al.*, 2014).

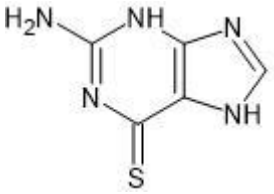


Figure 18: Structure of 8-thioguanine (8TG). Structure redrawn using ACD/ChemSketch from Yun *et al.* (2014).

Three bisubstrate analogues were synthesised by Shi et al. (2012) which have provided improved inhibitory activity against the *E. coli* HPPK enzyme. All three consist of a pterin moiety, an adenosine moiety and two to four phosphoryl groups abbreviated to HP, Figure 19. The first P^1-(6-hydroxymethylpterin)-P^2-(5′-adenosyl)diphosphate (HP_2A) showed little binding affinity and inhibitory effects. The second, P^1-(6-hydroxymethylpterin)-P^3-(5′-adenosyl)triphosphate (HP_3A), had a moderate binding affinity and inhibitory activity with a K_D of 4.25 µM and an IC_{50} of 1.27 µM. the third analogue, P^1-(6-hydroxymethylpterin)-P^4-(5′-adenosyl)tetraphosphate (HP_4A), had a K_D of 0.47 µM and an IC_{50} of 0.44 µM. These analogues were shown to occupy both the binding sites of HMDP as well as ATP, providing significant conformational changes which result in an altered phosphate-binding pattern (Shi, Shaw, Liang, *et al.*, 2012). Unlike most kinase enzymes, in the pyrophosphoryl transfer catalysed by HPPK, the substitution reaction takes place at the β-phosphate. This β substitution is altered to a γ substitution by the binding of the bisubstrate analogues indicating an alteration in enzyme activity. Later work by Shi, Shaw, Li, et al. (2012) in which improved protein binding was observed, was a synthesis in which the pterin moiety was replaced with a 7,7-dimethyl-7,8-dihydropterin and the phosphate bridge was replaced with piperidine linked thioether. Of these

compounds one, 2-amino-7,7-dimethyl-4-oxo-3,4,7,8-tetrahydro-pteridine-6-carboxylic acid(2-{4-[5-(6-amino-purin-9-yl)-3 4-dihydroxy-tetrahydro-furan-2-ylmethylsulfanyl]-piperidin-1-yl}-ethyl)-amide (HP-19), had vastly improved activity and consisted of the above mentioned structural changes joined to an adenosyl moiety, giving the compound improved binding specifically to Loop 3 of the catalytic site. HP-19 had a K_D 2.55 µM and IC50 of 3.16 µM (Shi, Shaw, Li, *et al.*, 2012).

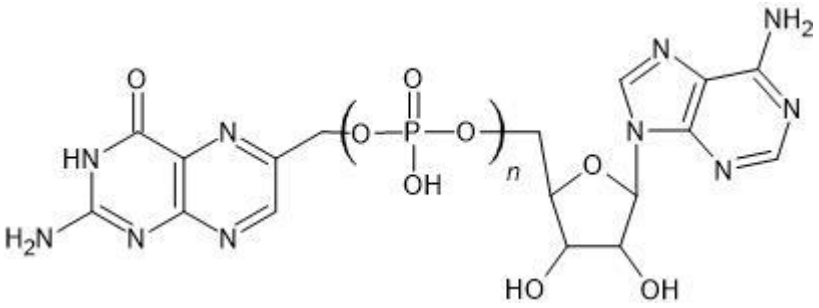


Figure 19: Basic structural backbone of the HP analogues. Structure redrawn using ACD/ChemSketch from Shi, Shaw, Li, *et al.* (2012).

As mentioned earlier, in some organisms folate enzymes are bifunctional and their function will depend on whether they are bi- or tri-functional. Shaw et al. (2014) set out to understand the bifunctional HPPK-DHPS enzyme from the biological warfare agent *Franscisella tularensis* (*F. tularensis*). The binding and inhibitory effects from HMDP derivatives were examined and were found to have significant isozyme selectivity to the *F. tularensis* HPPK-DHPS enzyme. The work was based on the compounds designed by Shi, Shaw, Liang, *et al.* (2012) and found that 5'-*S*-[1-(2-{[(2-amino-7,7-dimethyl-4-oxo-3,4,7,8-tetrahydropteridin-6-yl)carbonyl]amino}ethyl)piperidin-4-yl]-5'-thioadenosine (HP-18) inhibited *E. coli* HPPK with a K_D of 2.6 µM and an IC_{50} of 3.2 µM and the *F. tularensis* K_D was not determined but an IC_{50} of 180 µM was found. Compound 2-amino-7,7-dimethyl-4-oxo-3,4,7,8-tetrahydro-pteridine-6-carboxylic acid(2-{2-[5-(6-amino-purin-9-yl)-3,4-dihydroxy-tetrahydro-furan-2-ylmethanesulfonyl]-ethylcarbamoyl}-ethyl)-amide (HP-26) inhibited *E .coli* with K_D 4.2 µM and IC_{50} of 9.5 µM and *F. tularensis* with a K_D 2.0 µM and IC_{50} of 5.1 µM. Thus HP-26 appears to be a promising agent for further development of inhibitors of bifunctional HPPK enzymes (Shaw *et al.*, 2014).

The majority of studies performed on HPPK have relied on the use of bioinformatics for improved kinetic, molecular and dynamic parameters, and use as a screening tool, and have used several commercially available or self-synthesised compounds. To date, the focus has been on the exploitation of the catalytic centers and findings have indicated the best inhibitors to be fairly large molecules capable of spanning both the substrate and ATP binding pockets of the enzyme. Drug design strategies have incorporated the knowledge of amino acid residues to improve the binding of

compounds. As mentioned the high degree of similarity of HPPK enzymes amongst microbial species has made it possible for the results of studies performed on one species to be carried over and used in subsequent studies on the different organisms as can be seen by the use of results from *E. coli* studies in *S. aureus* and *F. tularensis* studies. In 2019, compounds were patented by Xinhua Ji, Genbin Shi and X. Shaw, based on the work of Shi, Shaw, Liang, et al. (2012), an example of which being HP-19 and the derivatives thereof and the other compounds mentioned in the study, for use as antimicrobial agents (Ji, Shi and Shaw, 2019).

1.4 Evaluation of the PTR1 enzyme

Organisms such as *T. brucei* are auxotrophic for pterins and folates and rely on salvage mechanisms to obtain pterins and folates from their environment. PTR1 is an enzyme involved in the salvage mechanism and is a functional short-chain homo-tetramer, falling into the dehydrogenase reductase family. PTR1 exhibits dual functionality in that it is responsible for the catalysis of biopterin to dihydrobiopterin (H_2B) and H_2B tetrahydrobiopterin (H_4B), as well as reducing folate to dihydrofolate (H_2F) and H_2F to tetrahydrofolate (H_4F), a function usually performed by DHFR, shown in Figure 20. These reactions are performed using NADPH as a cofactor (Kimuda *et al.*, 2019). Trypanosomes have both DHFR and PTR1 and thus the inhibition of DHFR, which has been the classical anti-folate inhibitory target, can be circumvented and will result in the overproduction of PTR1 by the parasite. PTR1 is also less susceptible to traditional anti-folate inhibitors. Studies performed using live parasite, in which gene knockout or downregulation has occurred in *T. brucei* resulted in parasite death and indicated the essential role played by PTR1. Furthermore, a lack of PTR1 has also resulted in phenotypic defects and reduced in vivo virulence of the parasites. The enzyme exists as an asymmetric tetramer of which each monomer consists of two α-/β-domains in which seven β-sheets are between two sets of α-helices (Panecka-Hofman *et al.*, 2017). An L-shaped depression houses the active site, mainly formed by a single subunit connected by the C-terminus of the partner subunit. The cofactor and a phenylalanine 97 contribute to the formation of the catalytic center. NADPH binds to each subunit of the homo-tetramer, stabilised by a tight network of conserved H-bonds. The substrate pterin moiety binds in a peculiar π-sandwich between the nicotinamide ring NADPH and the aromatic side chain of phenylalanine 97 of the PTR1 catalytic center, suggesting an ordered binding reaction mechanism in which substrate binding is preceded by cofactor binding. Within *Leishmania* species, which also contain PTR1, salvage of extracellular biopterin is mediated primarily through biopterin transporter 1 (BT1) and the fully oxidised petrin is then reduced sequentially to H_2B and H_4B, by PTR1 (Cullia *et al.*, 2018).

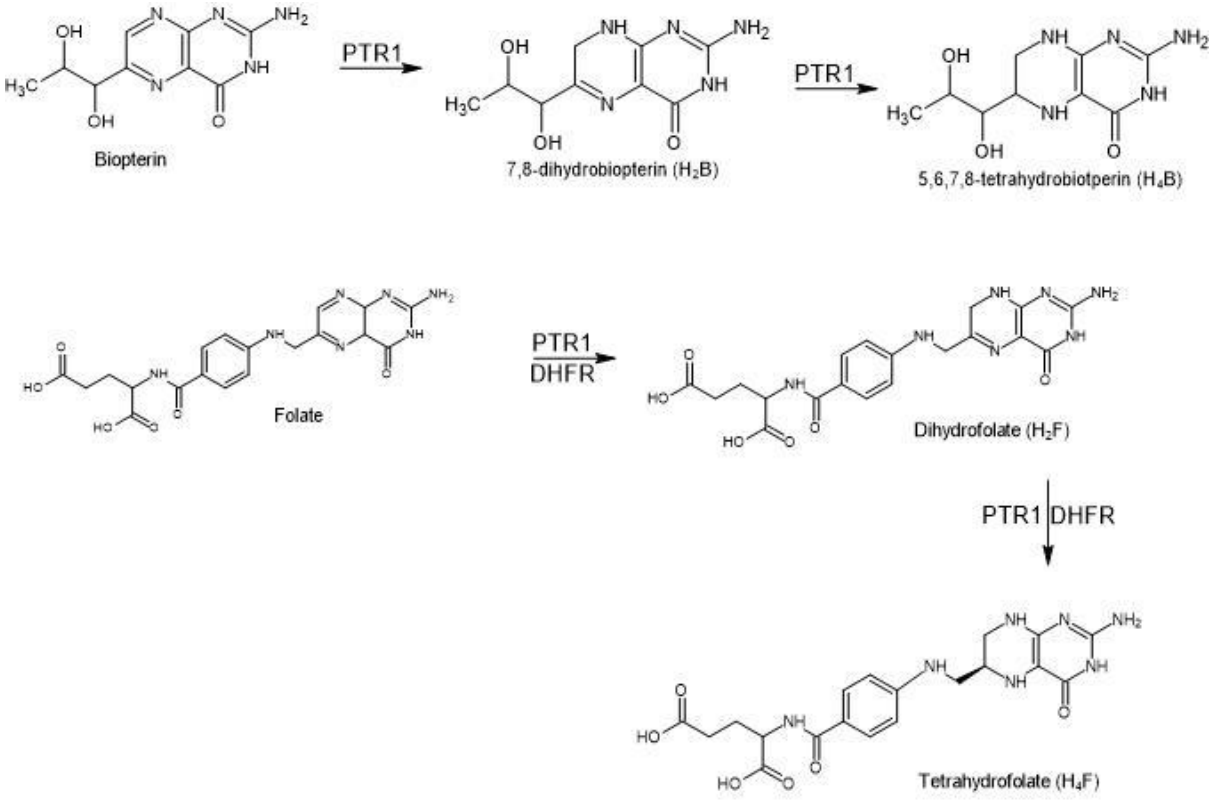


Figure 20: Enzyme activity of PTR1. Showing the dual enzyme activity of PTR1, namely the conversion of biopterin to H_2B and H_2B to H_4B, as well as the conversion of folate to H_2F and H_2F to H_4F. The latter reaction is also performed by the enzyme DHFR. Reactions redrawn using ACD/ChemSketch from Kimuda *et al.* (2019).

To date, the majority of PTR1 inhibitors were designed as competitive inhibitors of biopterin or folic acid and although few non-competitive inhibitors have been identified, they possess a substrate-like structure ascribed to the pyrimidine class or its bioisosteres (Pöhner *et al.*, 2020). It is noted that methotrexate (MTX), Figure 21, binds into the biopterin binding site preventing substrate binding thus acting as a competitive inhibitor. MTX also has activity against DHFR and has been used synergistically with PTR1 inhibitors to inhibit all folate and pterin uptake by the parasite (Linciano *et al.*, 2019).

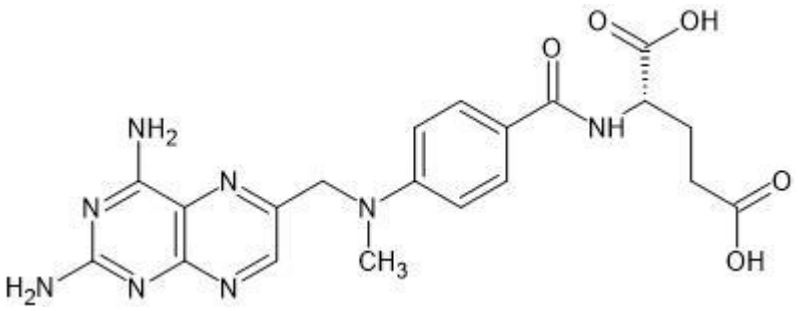


Figure 21: Structure of methotrexate. Structure redrawn using ACD/ChemSketch from Linciano *et al.* (2019).

Linciano et al. (2020) conducted an extensive *in silico* screening study of 522 compounds in the LIBRA compound library against the PTR1 enzymes of *T. brucei* and *Leishmania major* (*L. major*) and found nine compounds to be active and were selected for cell-based parasite screening as single agents and in combinations with MTX. The results indicated that the most effective inhibitor was 4-(benzyloxy)pyrimidine-2,6-diamine (name LIB_66) shown in Figure 22, from which it was found that the pyrimidine ring stacks in the active site in a similar manner as described for the pterin moiety of biopterin, while the phenyl ring establishes a face-to-edge stacking interaction with the aromatic side chains of the tryptophan 221 of the catalytic center. LIB_66 had an IC_{50} of 0.6 µM against *T. brucei* PTR1 and was used as a lead compound to synthesise further compounds with improved IC_{50} profiles with the lowest having an IC_{50} of 0.2 µM. The LIB_66 derivatives had half-maximal effective concentration (EC_{50}) in cell-based parasite screening ranging from 0.27-6.1 µM in combination with MTX (Linciano *et al.*, 2020).

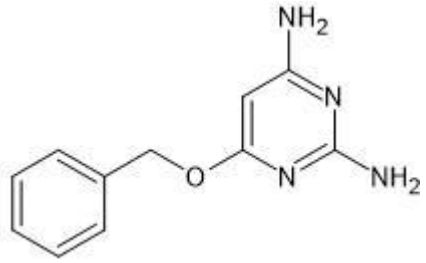


Figure 22: Structure of 4-(benzyloxy)pyrimidine-2,6-diamine. Structure redrawn using ACD/ChemSketch from Linciano *et al.* (2020).

Flavonoids have been identified as possessing antiparasitic effects, in particular, PTR1 inhibition and in a study to further investigate and develop flavonoid compounds, the inhibitory effects of chroman-4-one and chromen-4-one, shown in Figure 23 A and B respectively, derivatives were investigated and found to show inhibitory effects against the target PTR1 enzyme and *T. brucei* parasite (Di Pisa *et al.*, 2017). These derivatives contained phenoxy, acetophenone and benzaldehyde substitution to increase binding affinities between the compounds and the binding center of PTR1, like the binding mode of the pterin substrate. The phenoxy substitution proved to be the most effective in inhibition with an IC_{50} of 31 µM against PTR1. This inhibition provides a basis for further exploration of flavonoid compounds and further structural improvements in chemical structure. These results were correlated in a study performed by Omolabi, Iwuchukwu, Odeniran, & Soliman (2020).

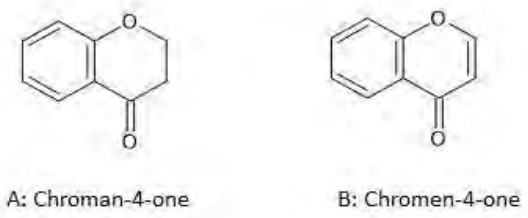


Figure 23: Structure of chroman-4-one and chromen-one. Structure redrawn using ACD/ChemSketch from Di Pisa *et al.* (2017).

Inhibitors of PTR1 to date have been largely monocyclic and bicyclic aromatic systems, yet the catalytic center allows for the accommodation of expanded molecular cores. In an attempt to investigate tricyclic-based aromatic compounds Landi et al. (2020) synthesised and evaluated 2,4-Diaminopyrimido[4,5-b]indol-6-ol, Figure 24, as a potential PTR1 inhibitor. It was found that the binding mechanism adopted a substrate-like orientation within the catalytic center which maximises hydrophobic and hydrogen-bond interactions. Inhibition studies further indicated that an IC_{50} result of 83 µM was obtaining providing a positive indication for the use and optimisation of tricyclic aromatic compounds (Landi *et al.*, 2020).

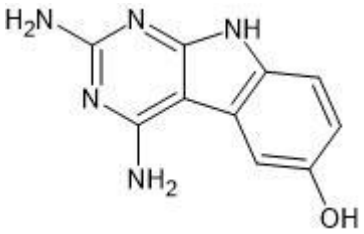


Figure 24: Structure of 2,4-diaminopyrimidol[4,5-b]indol-6-ol. Structure redrawn using ACD/ChemSketch from Landi *et al.* (2020).

The use of computational methods for the development and functional activity of PTR1 inhibitors is heavily relied upon. The crystal structure of *T. brucei* PTR1 and *L. major* PTR1 has been resolved and used as a template to investigate substrate and cofactor binding with the enzyme, thus providing pivotal information in the development of PTR1 inhibitors. It is noted that the majority of inhibitors act as competitive inhibitors of PTR1 substrate, biopterin, and interact with the catalytic center in a substrate-like manner. Inhibition of PTR1 in combination with DHFR inhibition by MTX proves to be the most effective manner of parasite death, although, as indicated, PTR1 is an essential parasitic enzyme and when lacking results in phenotypic abnormalities, decreased virulence and parasite death. Effective inhibitors have been designed with low micromolar and nanomolar inhibitory effects, providing a strong foundation from which further therapeutic agents can be designed.

1.5 Problem Statement

Antimicrobial resistance is fast becoming a serious threat as the treatment of many resistant organisms is left to highly toxic last resort agents. This is especially of concern in Sub-Saharan Africa where proper knowledge and expertise is sorely lacking (Spellberg *et al.*, 2008). The identification of the folate *de novo* synthesis and salvage pathways as a source of potential drug targets has been around for some time with great success, yet the need for novel therapeutic agents remains high and provides a sliver of hope for the ongoing resistance problem (Bermingham and Derrick, 2002)(Spellberg *et al.*, 2008).

As mentioned previously, HPPK is a potential drug target that has been extensively studied and in 2020 a patent was registered for compounds to use as therapeutic agents (Marimuthu, Singaravelu and Namasivayam, 2017). PTR1 in *T. brucei* also studied extensively, has not yielded any clinically applicable drug to date (Panecka-Hofman *et al.*, 2017). Previous studies on HPPK has resulted in a compound, 8MG, which in and of itself does not have a very good IC_{50} and thus derivatives of 8MG are being studied for improvements in the IC_{50} (Shi, Shaw, Liang, *et al.*, 2012). As improvements in bioinformatics techniques are being made with their applicability to drug discovery, its utility should be made full use of. It has shown its usefulness in the identification of hit compounds with success rates similar to those of high-throughput screening, requiring fewer resources and ease of applicability (Gershell and Atkins, 2003). Many HPPK and PTR1 enzymes have been crystalized and knowledge of the enzyme catalytic centers allow these enzymes to be used in bioinformatics (Kok, Steegenga and McKay, 2018). Currently, a crystal structure of *S. enterica* HPPK does not exist neither has any docking studies been performed on a modelled enzyme to yield potential hit compounds. A crystal structure of *E. coli* HPPK does exist and few hit compounds have been produced from *in silico* docking studies. Identification of *T. brucei* PTR1 has been performed by Kimuda et al. (2019) and hit compounds from *in silico* docking studies have been tested against the parasite and shown to inhibit their growth. Confirmation that PTR1 inhibition is the mechanism by which parasite death occurs has not been demonstrated.

1.6 Aims and Objectives

The aim of this research project was firstly, to confirm that compounds predicted by Kimuda et al. (2019) to be PTR1 inhibitors and found to inhibit cultured *T. brucei* parasites were indeed able to inhibit PTR1 in a colorimetric PTR1 enzyme assay, thus confirming the mechanism of action of the compounds. Secondly, the project aimed to model the HPPK enzyme of *S. enterica* and use the model for *in silico* docking studies along with the HPPK enzyme of *E. coli* to identify novel hit compounds, which were tested against the enzyme to confirm activity using a luminescence assay.

To achieve these aims, the following objectives were pursued:

1. Develop a homology model of the *S. enterica* HPPK enzyme and using this model and the crystal structure of *E. coli* HPPK, perform *in silico* docking studies for hit compound identification.
2. Express and purify the His-tagged *S. enterica* HPPK, *E. coli* HPPK and *T. brucei* PTR1 proteins.
3. Establish and optimise a plate-based NADPH absorbance assay for *T. brucei* PTR1 activity and inhibition studies.
4. Establish and optimise a plate-based luminescence assay for *S. enterica* HPPK and *E. coli* HPPK activity and inhibition studies.
5. Identify any compounds which show sufficient inhibition of *S. enterica* HPPK and *E. coli* HPPK which could potentially be explored further in subsequent research projects.

CHAPTER 2

Homology Modelling of *S. enterica* HPPK and in silico screening of *S. enterica* HPPK and *E. coli* HPPK

2.1 Introduction

Improvements in technology and our understanding of how computational methods could improve and increase the speed at which novel drugs are discovered, or the repurposing of drugs achieved, is being produced and developed continuously. The field of bioinformatics has become a very useful tool in this process, not only in delivering potential drugs but also in its application to produce drug target models through the use of homology modelling (Cavasotto and Phatak, 2009). The need for the rapid identification of small molecules which bind to drug targets is one of the greatest driving forces behind the continuous improvements made in this discipline. It is a fast solution to an expensive problem. It is estimated that the success rate of computational drug development strategies is on par with non-computational methods, such as high-throughput screening, in identifying active compounds and can cut costs by as much as 50% (Hillisch, Pineda and Hilgenfeld, 2004). Two strategies are adopted within the field of computational drug development: 1) structure-based methods, in which information regarding the structure of the target and the ligand are required; and 2) ligand-based methods, in which a more in-depth understanding of the chemical properties of the ligand is required and it is screened non-specifically (Hillisch, Pineda and Hilgenfeld, 2004).

Homology modelling is a computational method used to derive the structure of an unknown sequenced protein target from that of a closely related protein, whose structure has been resolved (Szilagyi and Zhang, 2014). The resolved structure, referred to as the template, is usually obtained from the protein database (PDB), and the desired structure is referred to as the target. Static protein structures are typically determined using X-ray crystallography, or solution structures, which are solved using two dimensional (2D) nuclear magnetic resonance (NMR) of isotopically labelled proteins (Morris and Lim-Wilby, 2006). X-ray crystallography is a technique that uses a crystal of a specific protein to diffract incident x-rays in a specific direction (Joe, 2005). The electron density of the crystal is measured using the angle of x-ray diffraction and the intensity of the diffracted beam, from which atomic positions can be determined, as well as information regarding the chemical bonds of the crystal. The image produced is given a resolution. Resolution describes the ability to distinguish between neighbouring features in an electron density map (Tramontano, 2017). By convention, it is defined as the minimum plane spacing given by Bragg's law for a particular set of x-ray diffraction intensities. Resolution is the distance corresponding to the smallest observable

feature; if two objects are closer than this distance they will appear as one combined binary large object (blob) rather than two separate objects (Szilagyi and Zhang, 2014). The higher the resolution, the higher the quality of the crystal structure, conversely the lower the numeric value of the resolution the higher the quality of the image.

Protein NMR spectroscopy is used to obtain structural and dynamic properties of proteins, nucleic acids and their complexes. Structural determination consisting of several phases uses a highly specialised technique involving the quantum properties of the nucleus of an atom (Cavalli *et al.*, 2007). The local molecular environment and measurements provide a map of the chemical linkage of the atoms, the distance in space and the speed with which they move with respect to each other (Herrmann, Güntert and Wüthrich, 2002). Protein samples should be of high purity in an aqueous solution in the range of 300-600 microlitres (µl), with a concentration of 0.1-3 millimolar (nM). Each of the distinct nuclei within the molecule will experience a distinct electronic environment, resulting in a distinct chemical shift used for recognition (Yamazaki *et al.*, 1998). Due to the large size of proteins, the number of resonances can typically be several thousand causing incidental overlaps in a one-dimensional (1D) spectrum. To overcome the overlaps, multidimensional experiments correlate the frequencies of distinct nuclei to decrease the overlap and has larger information content (Montelione *et al.*, 2000). It is important to obtain information regarding which chemical shift corresponds to which atoms, which is typically achieved by sequential walking - the use of information derived from several different types of NMR experiments.

Homology modelling constructs an atomic-resolution model of a target protein, that has an unknown structure, from its amino acid sequence, using an experimentally determined three-dimensional (3D) structure of a related homologous protein, that has a known structure (Bates *et al.*, 2001). Protein structures have proven to be more conserved than protein sequences, hence related proteins would not only have similar sequences but also structures. Thus if a given protein has vastly diverged from a related protein in sequence but still has partial similarity, it will still share common structural properties, particularly the overall fold (Pozzan, 2006). The sequential steps in model production are as follows: 1) template selection, 2) target-template alignment, 3) model construction and 4) model assessment. Steps one and two are generally performed as a single step in which the best template structure is identified by a serial pairwise sequence alignment aided by the use of sequence databases (Davies *et al.*, 2002). Some databases use more sensitive algorithms such as a position-specific scoring matrix to identify more distantly related homologs, thus producing a larger number of potential templates and improved templates for the target sequence. Protein threading, which is the recognition of protein folds or 3D-1D alignment, is an alternative option for

template identification and has proven to be more sensitive than purely sequence-based methods, especially when only distantly related templates are available(Bates *et al.*, 2001). Model construction uses three major classes to plot the 3D structure as a set of Cartesian coordinates. These are 1) fragment assembly, which assembles a complete model from conserved structural fragments from solved structures; 2) segment matching, which splits the sequence into several smaller segments which are matched to the chosen template; and 3) satisfaction of spatial restraints, using calculations which construct a 3D structure from NMR spectroscopy, in which geometrical criteria is converted to probability density functions (Tramontano, 2017). Model assessment for a target is performed either by statistical potentials or energy calculations. Unfortunately, neither correlates well with true structural accuracy.

Another very important factor in homology modelling is the sequence of the two proteins used for modelling. Homology modelling is done by using a known template crystal structure to predict the tertiary structure of a related protein based on similarities in their amino acid sequences, using *in silico* algorithms (Pozzan, 2006). The higher the similarities between protein sequences the better the chances of producing a workable model. This does not mean that the template and the target necessarily need to be functional homologues of each other; it only depends on the similarity of their amino acid protein sequences. However, the two should have the same biological role (e.g. from different species), as the function of a biological molecule is highly dependent on its structure and vice versa (Szilagyi and Zhang, 2014).

From this we can see that for an optimal model of a protein with an unknown structure to be produced, a template with high crystal structure resolution, high protein sequence homology, and of a similar biological function should be used (Hillisch, Pineda and Hilgenfeld, 2004). The software used will give the user several options to choose which template to use for modelling with resolution and chain identity as options. The resultant model needs to be verified for accuracy and several evaluation methods exist. ProSa provides verifications for models which uses the thermodynamic hypothesis of protein folding, stating that native proteins would fold into their 3D structure within equilibrium conditions which corresponds to the global minimum of the free energy under physiological conditions (Luthy, Bowei and Einsenberg, 1997). These plots provide information regarding the nature of the molecular structure of the generated model. Furthermore, the root mean square deviation (RMSD) is used to measure the structural similarity between a model and a template (Pozzan, 2006). It is the square root of the average squared distances between equivalent atoms after an optimal superposition. An RMSD value should ideally be ≤ 2Å from the used template if the template has a ≥ 60% chain similarity (Tramontano, 2017).

In silico docking studies are used to determine structural parameters for an interaction between a protein structure and a ligand. For drug discovery purposes, the ligands are usually small drug-like molecules that conform to Lipinski's Rule of Five (<5 hydrogen bond donors, <10 hydrogen bond acceptors, molecular mass <500, logP<5) (Lipinski, 2000). It assists in the exploration of a protein structure for potential active sites that can be exploited to inhibit protein function with drug-like ligands and predicts molecules that may bind strongly to this site and allows for compounds to be prioritised for bioassay (Hughey and Krogh, 1996). These compounds are typically ranked according to their binding affinities and binding distance. This allows for the inhibitory potency of the predicted ligands against these targets to be either tested *in vitro* or *in vivo* (Davies *et al.*, 2002). *In silico* docking results, in combination with *in vitro* studies, can be used for the optimisation of a potential drug, thus decreasing the turnover time of hit compounds to lead compounds to marketable drugs. Autodock Vina is a software program that allows for a set of ligands to be superimposed in different conformations onto the target protein at the selected amino acid residues, allowing for the best conformational fit to the protein (Nguyen *et al.*, 2020). The ligand-protein interactions are measured in terms of their binding affinities that is the energy needed to break the bond. These predictions are based on interactions observed from known ligand-protein interactions and state the most likely type of bond to form between the amino acids of the protein and the ligand (Green, 2003). Binding distance is another important parameter to use, as the further from the protein: the ligand is the less likely it is to interact strong enough and long enough to be worth exploring further.

The experiments performed in this chapter aimed to:

i. Model the tertiary structure of the HPPK enzyme of *S. enterica* using homology modelling techniques.
ii. Determine how successful the modelling was using validation methods.
iii. Perform in-silico screening against the HPPK enzyme of *S. enterica* and *E. coli* using docking methods for identifying potential inhibitory ligands.

2.2 Methods and Materials

2.2.1 Modelling of *S. enterica* HPPK enzyme.

The National Center for Biotechnology Information (NCBI) website was accessed. HPPK Salmonella was entered as a text search option and the results were shown. The RefSeq source database option was selected for further refinement of the results, giving a total of 38 results. The sequence which was selected was option 9, selected due to correspondence to validated, peer-reviewed journal article structure, showing a full protein sequence for *Salmonella enterica* (2-amino-4-hydroxy-6-hydroxymethyldihydropteridine diphosphokinase [Salmonella enterica], Accession: WP_023178310.1, GI: 554677281). The FASTA tab was selected and the sequence was copied onto a Word document.

On a new tab, the PRotein Interactive MOdeling (PRIMO), was opened. The job name was entered with the modeller key. The selected sequence was entered into the space labelled target sequence. The optional input dropdown menu was selected and the job description was completed. Under the same optional input dropdown menu, the following parameters were set: 1) template identification was selected to automatically identify templates with the HHSearch from the HHSuite database version:pdb70_from_mmcif_190904.tar.gz selected to manually identify the templates to use. 2) the alignment was set to manually edit the generated alignment using the Pseudo-Expresso option under the T-COFFEE set. 3) the model prefix was set to model, the number of models set to 4, and the refinement method was set to Very slow. The Start option was selected, which began the identification process.

Once Template Identification was complete, the top 10 best templates were shown with their Protein Database crystal structure number, which chain of the crystal structure it best matched to, the % identity of the template structure and that of the sequence that needed to be modelled, the number of amino acid residues covered by the templates and the resolution of the crystal structure. The template structure was viewed, as well as how the two sequences aligned with each other. Template 2qx0 was selected.

The target and the template alignment are generated automatically on the PRIMO software using T-COFFEE in 3D-COFFEE mode.

Once modelling was complete, the software automatically listed the four models it was programmed to provide, from most favourable to least favourable. Each model was given a number, a Dope Z-Score, and an RMSD. The model was viewed using the software. The model was evaluated using the

Procheck option and links to other servers were provided namely ProSA and Verify 3D, which are both local and global servers.

The ProSA webpage allows one to upload the model to the website and state the template number used, 2qx0. After these two were completed, the analyse option was selected and the software analysed the model quality. It provided the overall model quality with a Z-Score, local model quality, and energy distribution on the model.

The model was also uploaded onto the Verify 3D webpage and the analysis run. This webpage scored the model on a pass or fail basis and gave the number of residues that have averaged a 3D-1D score of greater than or equal to 0.2.

2.2.2 Preparation of *S. enterica* HPPK and *E. coli* HPPK structures for docking studies.

Using Google Chrome the PDB was accessed from where the *E. coli* HPPK crystal structure was obtained, reference number 3hcx. The structure was downloaded and saved in a .pdb format.

Discovery Studio was used to open the downloaded PDB crystal structure. From here the protein was stripped of all bound ligands rending it in its apo state. The *E. coli* HPPK Apo file was saved to be used for docking.

The modelled *S. enterica* HPPK was modelled without any bound ligands and was thus in its apo state.

The enzymes, in their apo states, were measured to determine their dimensions, which were used to set a finite area in which docking could take place by using these dimensions to draw a box around the enzymes if docking outside of the desired area were to occur.

Using one enzyme at a time they were uploaded onto the Autodock Vina software along with their dimension. Input into the desired docking area, namely the catalytic and ATP binding sites was performed.

2.2.3 Preparation of the ligand library

The ligand library was downloaded from the ZINC Database website, which provides the compounds in their SMILES format which is ready for docking. The subset of the database used was ZINC15, which contain compounds that already conform to the Lipinski rule of 5 and are commercially available drug-like compounds.

Further preparation was performed by filtering the 10 639 555 compounds for compounds with an XlogP<=3, have fewer than 4 rotatable bonds, having at least 2 hydrogen donors, a net charge of 0 and a molecular weight of <=490, that is lead-like molecules.

This reduced the number of compounds to be screened to only 5107.

2.2.4 Docking the selected ligand set to *S. enterica* HPPK and *E. coli* HPPK

The enzymes were uploaded into the AutoDock Vina software and the dimensions of the enzymes were set. The catalytic and ATP binding centers amino acid residues, which was obtained from NCBI, were input to provide a preferential area of binding to occur.

The filtered ZINC15 compounds were uploaded and the software was set to perform the docking.

Once docking was completed, output files were automatically generated in a Microsoft Excel sheet to account for binding energy and binding distance, ranking the compounds from most favourable to least favourable. The two files were combined into one, from where the top molecules for each enzyme were selected. Further analysis for overlapping molecules was performed and a selection of 9 compounds were made; 3 compounds solely for *S. enterica* HPPK, 3 compounds solely for *E. coli* HPPK and 3 compounds that bound favourably to both.

2.3 Results

2.3.1 Homology Modelling

Homology modelling has proven to be a verified method of structural determination, which can be used for *in silico* docking studies. As the structure of the HPPK enzyme of *S. enterica* has not been experimentally determined, the purpose of this experiment was to use homology modelling as a method for structural determination for use in further experiments. The predicted *S. enterica* HPPK protein sequence was entered into the PriMo software to find sequences of similar proteins and perform an alignment to determine which sequence proved to be the best match. Once the top sequence was determined, the software provided the desired model of the *S. enterica* HPPK enzyme.

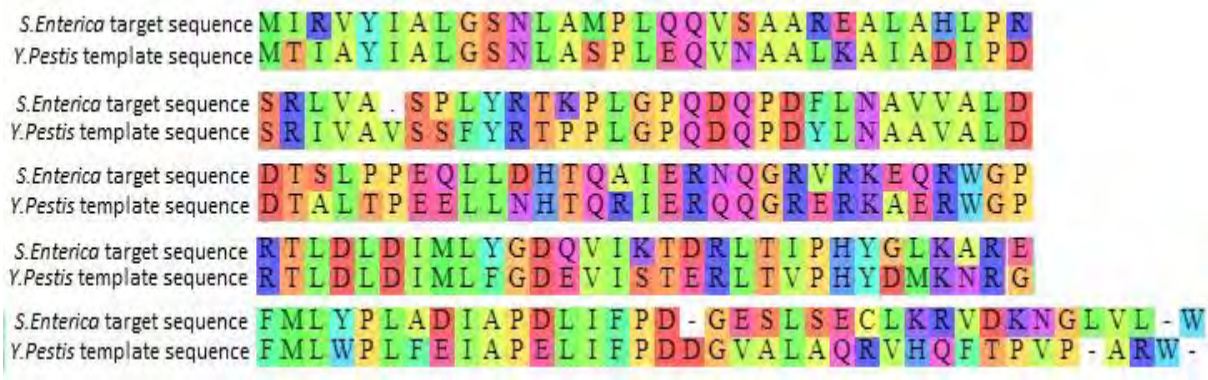


Figure 25: *S. enterica* HPPK sequence (target) aligned against that of *Yersinia pestis* (*Y. pestis*) HPPK (template). Gaps are represented by a (–) and modified residues are represented by a (.). Amino acids are represented by colours corresponding to Orange for small polar, green for hydrophobic, magenta for polar, red for negatively charged, and blue for positively charged amino acids.

Figure 25 shows the template and target sequences aligned. The known crystal structure which PRIMO identified as the ideal template to use for modelling was that of *Y. pestis* HPPK (PBD reference number 2qx0). The template has been crystallised as a dimer and chain A was used for modelling. The two proteins have a 66% amino acid identity and the template covered 96% of the amino acid residues of *S. enterica* HPPK. The resolution of the crystal structure was 1.80

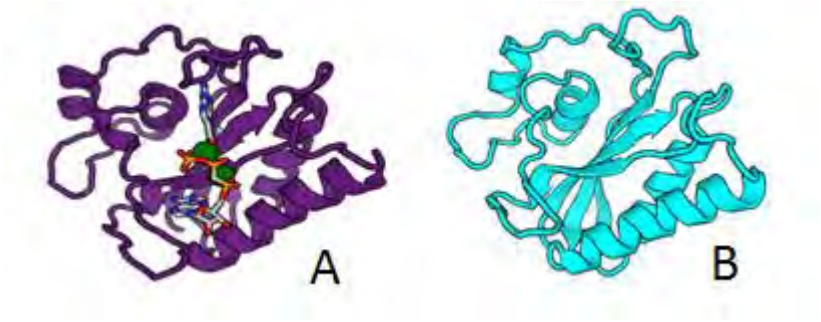


Figure 26: Crystal structure of *Y. pestis* HPPK (2qx0) (A) and modelled *S. enterica* HPPK (B). The template crystal structure *Y. pestis* HPPK,2qx0, shown here in purple with bound ligand and

coordinating metal ions (Mg^{2+}, shown in green), and the modelled *S. enterica* HPPK shown here in blue without a bound ligand or coordinating metal ions.

From Figure 26, the model which most closely resembles that of the template was selected and had a Dope Z-Score of -1.785 and an RMSD of 0.35 showing that the model and the template are significantly similar after superimposition.

2.3.2 Model Verification

Verification of the structure determined using homology modelling is performed to judge the accuracy and validity of the structure. The structure of *S. enterica* HPPK enzyme was modelled, using homology modelling, and as such the purpose of this experiment was to ascertain the accuracy and validity of the structure, for use in further experimentation. The model was uploaded onto the ProSa webpage, as well as the Verify 3D webpage for analysis of the structure compared to that of the template structure which was used for modelling. Both online software packages provide information regarding the model quality, energy distribution and 3D-1D amino acid residue position and conformation.

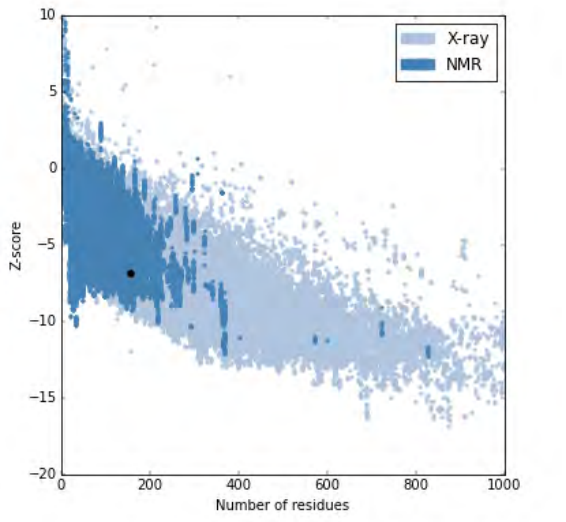


Figure 27: Model ProSa Z-Score plot. The ProSa Z-Score plot shows a Z-Score of -7.69 falling within the NMR spectrum of native proteins of a similar size.

Figure 27 shows that the model Z-Score falls within the NMR range of a native protein of a similar size which has been determined experimentally from protein chains currently found within the PDB. The Z-Score speaks to the energy separation between native folds and the average of the misfolds. The model can be assumed to be of good quality because the Z-Score falls within the range expected of a protein containing 159 amino acids.

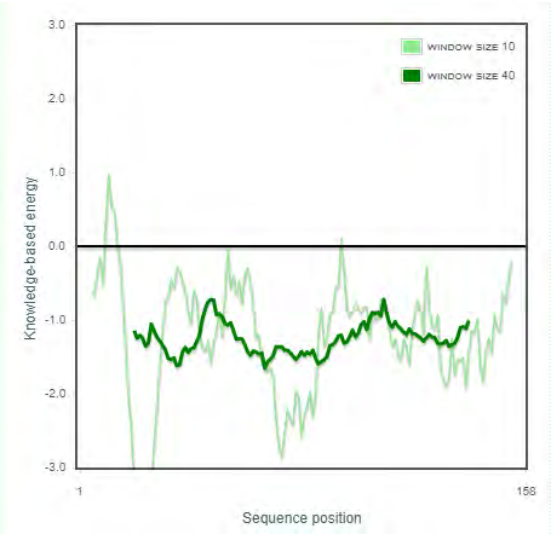


Figure 28: Model Knowledge-based energy chart. The Dark green line shows a smoothed 40-residue average energy and the light green line shows a smoothed 10-residue average energy line.

From Figure 28, one can see that knowledge-based energy charts, which plots energies as a function of amino acid sequence position, shows a favourable energy distribution for the *S. enterica* HPPK model quality as the line falls below zero. The dark green line on the plot is calculated by using the average energy over each 40-residue fragment which was smoothed and assigned to the central residue of the fragment and the light green line uses a smaller window size of 10 residues.

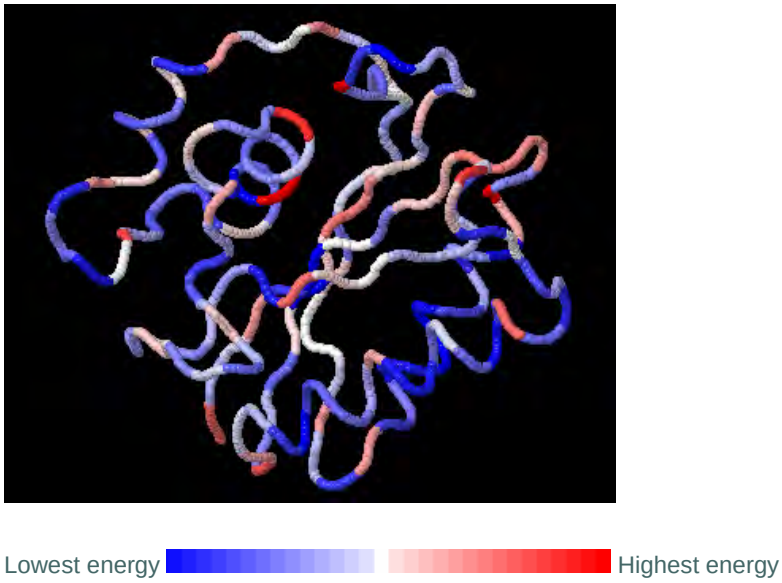


Figure 29: Model molecule energy visualization using Jmol software. The molecular residue energies are from low (blue) to high (red).

Higher energy regions, shown in red in Figure 29, are attributed to areas within the model that were not modelled very well and proved to be outliers. These areas serve as sources for further exploration to improve the quality of the model. Figure 29 shows the modelled *S. enterica* HPPK,

which contains areas of high energy, but the areas are few and lower energy areas predominant, thus further improvements to the model are not necessary.

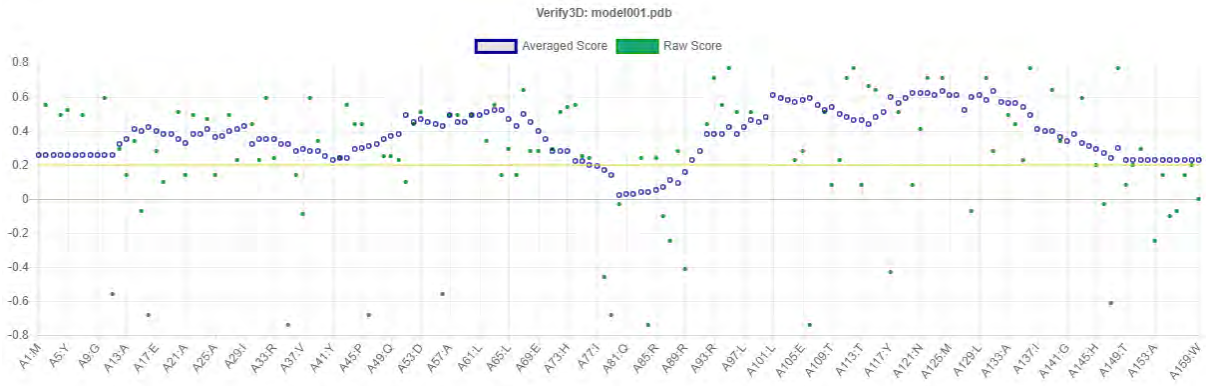


Figure 30: Verify3D plot. The plot compares the atomic protein model with its amino acid sequence as measured by a 3D profile.

Verification that the model, Figure 30, showed 96.20% of the amino acid residues have scored ≥ 0.2 using the Verify 3D software. We notice that the 3D model environment aligns well with the amino acid sequence giving the model a realistic morphology. This is done by assigning a structural class based on the location and environment of each residue position and by comparing the results to known good structures. The residue environment corresponds to 3 parameters: 1) the local secondary structure; 2) the area of the residue that is buried and 3) the fraction of side-chain area covered by polar atoms.

2.3.3 Visualisation of the template (*Y. pestis* HPPK) and target (*S. enterica* HPPK) using Discovery Studio.

This was done as a method to visualise the enzymes to further determine the validity of the model. The enzyme structures were uploaded into Discovery Studio in two separate windows and were selected and placed into one window where they were superimposed.

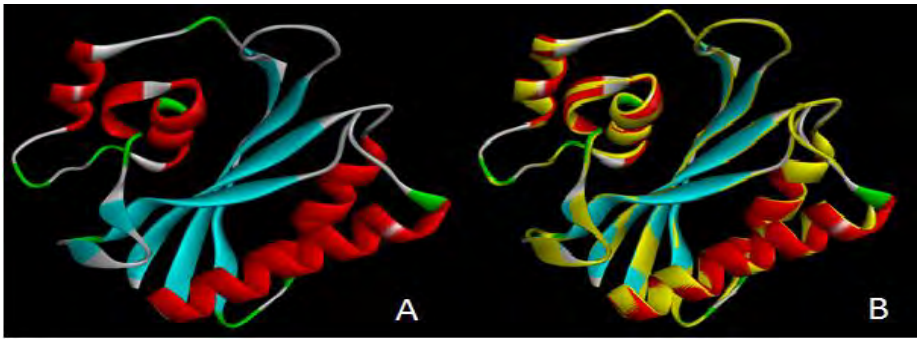


Figure 31: Visualised *S. enterica* HPPK (A) and *S. enterica* HPPK superimposed on *Y. pestis* HPPK (B) using Discovery Studio. This visualisation allows one to see how the two enzymes superimpose to show fit after verification has been completed.

From Figure 31, one can see that the two enzymes form a perfect fit. The areas shown in yellow are used to show the *S. enterica* HPPK enzyme as it fits onto the *Y. pestis* HPPK. The modelling of *S. enterica* HPPK, after being verified and tested proved to be a successful model which can be used for further experiments, namely *in silico* docking. All verification data points to the fact that the model and template used show a high level of similarity between each other.

2.3.4 Preparation of *E. coli* HPPK

The HPPK enzyme of *E. coli* was obtained from the PDB, with reference number 3hcx. This enzyme needed to be stripped of all bound molecules to allow for docking studies to be performed. Using Discovery Studio, the crystal structure was uploaded and all molecule sections which was not amino acid was selected and deleted. This renders the enzyme in an apo state.

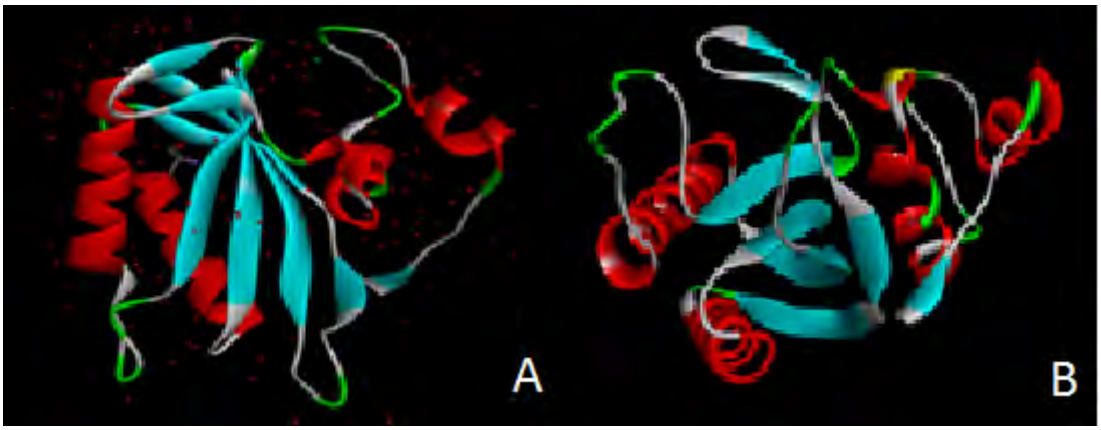


Figure 32: Showing *E. coli* HPPK with bound ligands (A), and *E. coli* HPPK in its apo state (B)

In Figure 32B, all molecules which are not amino acids had been removed. These molecules include water and a bound ligand, both of which resulted from the crystallisation process. The apo-HPPK was successfully determined and subsequently used for *in silico* docking studies.

2.3.5 *In Silico* Screening

In Silico screening of chemical compounds provides a rapid number of chemical entities in a short period which can further be examined using *in vitro* studies to verify the effectiveness of the method. Here several chemical compounds which are favourable drug-like compounds were chosen based on the Lipinski rule of 5 (ZINC 15) and screened against a largely untested enzyme, namely HPPK.

The 5107 compounds which were selected were added into the Autodock Vina ® software, along with the enzymes in their apo state. The critical binding sites on the enzymes were marked and the software was set to dock the compounds into the enzyme's catalytic binding site and the ATP binding site. The results of the docking were set to be displayed into a Microsoft Excel file for binding

energy and a separate file for binding distance. The two files were merged into one to determine the compounds which showed to bind most effectively. Low binding energy and short binding distance are what is most desirable, after which the specific type of bond obtained is used for further analysis of binding energy and distance and final choice of compounds. Binding distance is determined by the distance between the compound and the selected residues and averaged, and the binding energy is determined by how strongly the compound binds to the selected residues and averaged.

Table 1: *E. coli* HPPK top 3 ZINC 15 ligands and *S. enterica* HPPK top 3 ZINC 15 ligands, showing binding energy and binding distance between the ligand and the binding pocket.

	E. coli HPPK		*S. enterica* HPPK		
Compound	Binding Energy (kcal/mol)	Binding Distance (Å)	Compound	Binding Energy (kcal/mol)	Binding Distance (Å)
ZINC 0634089	-9.3	5.69062	ZINC 0082136	-9.4	4.95876
ZINC 0310450	-10.5	5.054349	ZINC 6792518	-9.4	4.95876
ZINC 4499339	-9.3	5.37888	ZINC 4498423	-9.4	4.84331

Table 1 shows the binding energies and binding distance of the ZINC compounds as they interact with the HPPK model *in silico*. All binding energies are low and binding distances small, indicating good interactions. The lower the binding energy the stronger the bond created between the ligand and the enzyme, likewise the smaller the binding distance between the ligand the protein the more closely the interaction will be and hence the interaction will indicate that the ligand interactions with the enzyme are strong (Seeliger and De Groot, 2010)(Nguyen *et al.*, 2020). These parameters indicate that the energy required to over the ligand-enzyme interact would be very large thus supporting the likelihood of the ligands acting as strong potential inhibitors of the enzyme (Pozzan, 2006).

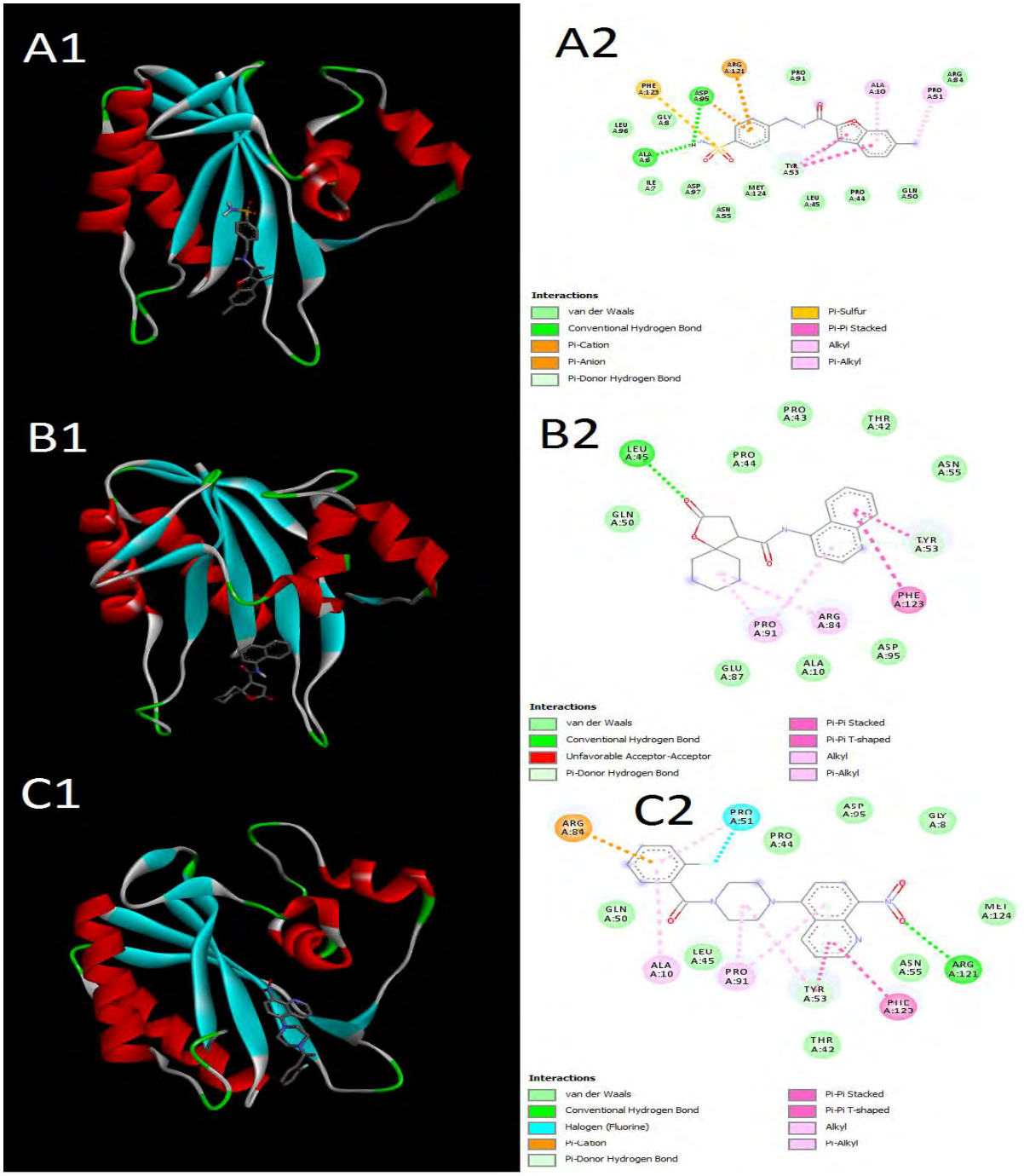


Figure 33: Top 3 ZINC 15 ligands bound to *E. coli* HPPK. A1, B1 and C1 represent the homology model interaction with the ligands. A2, B2 and C2 show the specific amino acid residues which interact with the ligands. A represents ZINC 0634089, B represents ZINC 0310450 and C represents ZINC 4499339.

Figure 33 A1 (ZINC 0634089), B1 (ZINC 010450) and C1 (ZINC 4499339) show the top ligands which docked to the *E. coli* HPPK, visualising the enzyme and the ligand. Figure 33 A2 (ZINC 0634089), B2 (ZINC 0310450) and C2 (ZINC 4499339) show the ligands interacting with the amino acid residues as well as the type of interactions that occur. Interacting amino acids with bond type and binding center are shown in Table 4 for *E. coli*.

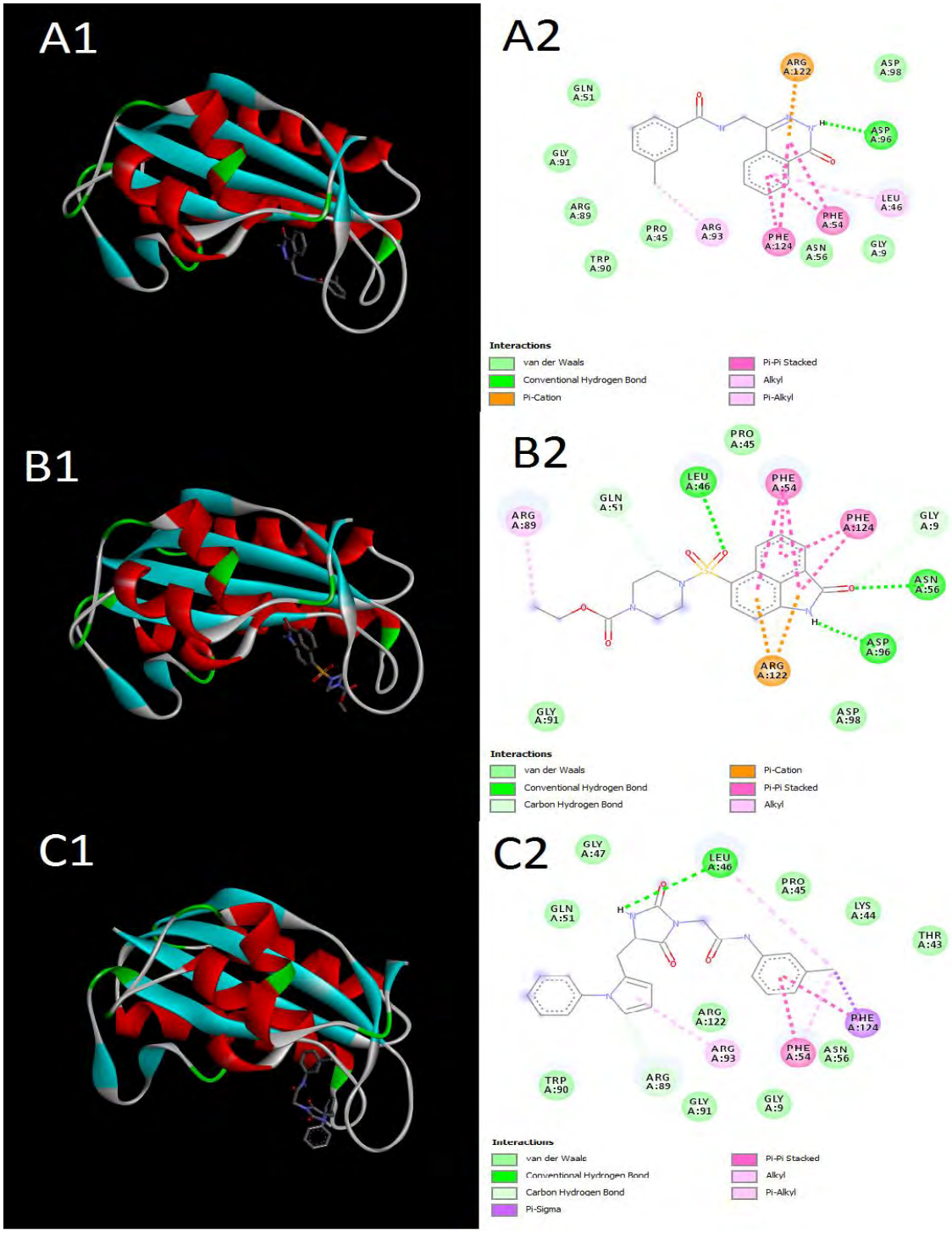


Figure 34: Top 3 ZINC 15 ligands bounds to *S. enterica* HPPK. A1, B1 and C1 represent the homology model interaction with the ligands. A2, B2 and C2 show the specific amino acid residues which interact with the ligands. A represents ZINC 0082136, B represents ZINC 6792518 and C represents ZINC 4498423.

The interaction occurring with *S. enterica* HPPK can be seen in Figure 34 A1 (ZINC 0082136), B1 (ZINC 6792518) and C1 (ZINC 4498423) with amino acid interactions shown in Figure 34 A2 (ZINC 0082136), B2 (ZINC 6792518) and C2 (ZINC 4498423). Interacting amino acids with bond type and binding center are shown in Table 5 for *S. enterica*.

Table 2: Top 3 ligands bound to both E. coli HPPK and S. enterica HPPK, showing binding energies and binding distance.

	E. coli HPPK		*S. enterica* HPPK	
Compound	Binding Energy (kcal/mol)	Binding Distance (Å)	Binding Energy (kcal/mol)	Binding Distance (Å)
ZINC 4671178	-9.8	5.43384	-9.6	4.9717
ZINC 4476281	-9.9	5.38942	-9.5	5.38876
ZINC 0923924	-9.9	5.40189	-9.1	4.29413

Table 6 shows the binding energies and binding distance of the ZINC compounds as they interact with the HPPK model *in silico*. All binding energies are low and bind distances are small, indicating good interactions.

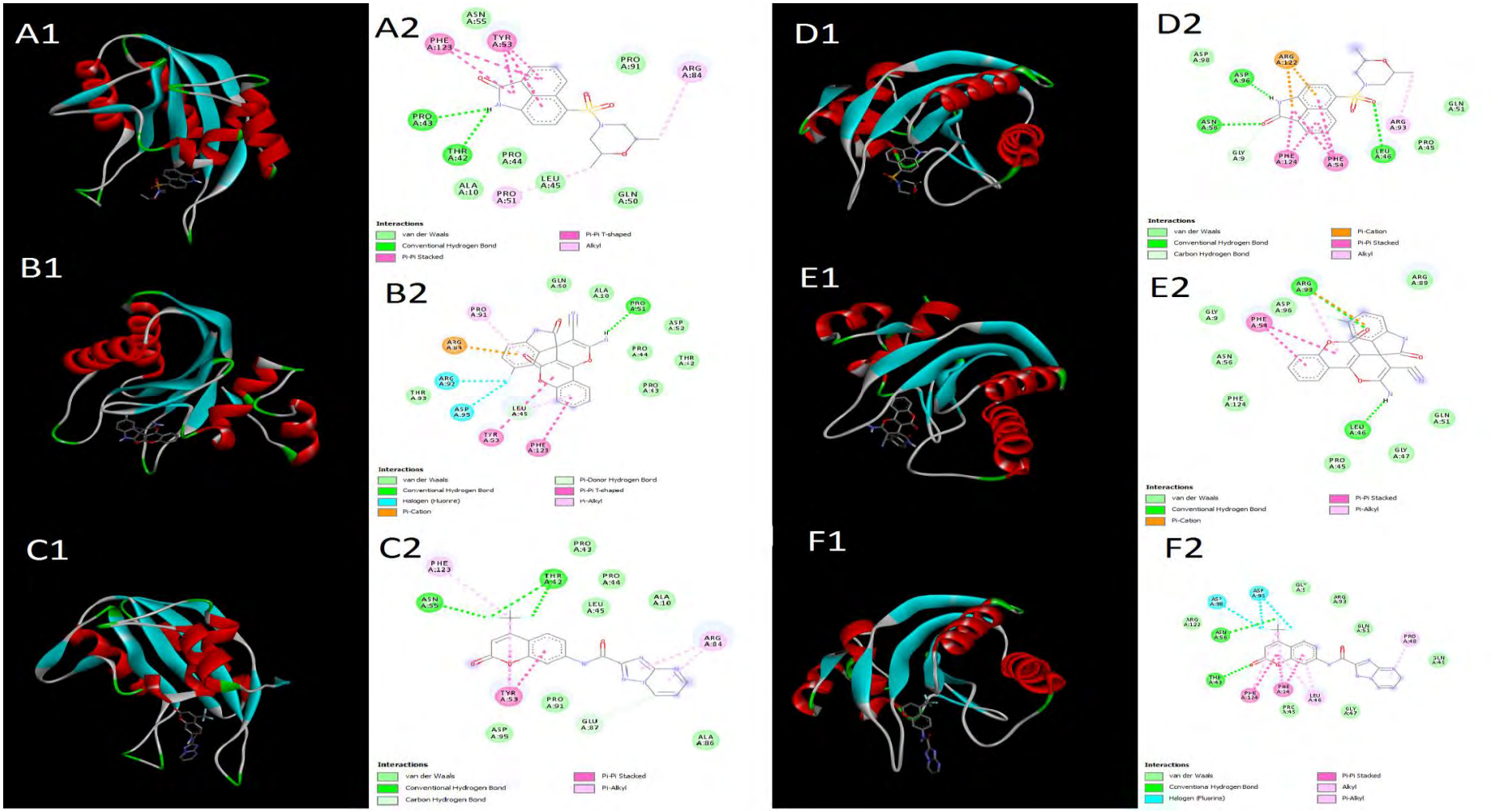


Figure 35: Top 3 ZINC 15 ligands which to both *E. coli* HPPK and *S. enterica* HPPK. A1-C2 are the ligands bound to *E. coli* HPPK and D1-F2 are the same ligands bound to *S. enterica* HPPK. In image 1 represents the ligand bound to the homology model of the respective organism enzyme and 2 represents the interacting amino acid residues with the ligand. A and D represent ZINC 4671178, B and E represent ZINC 4476281 and C and F represent ZINC 0923924. A, B and C show *E. coli* HPPK and D, E and F shows *S. enterica* HPPK.

Figure 35 A1 (ZINC 4671178), B1 (ZINC 4476281) and C1 (ZINC 0923924) show the top ligands which docked to the *E. coli* HPPK, visualising the enzyme and the ligand. Figure 35 A2 (ZINC 4671178), B2 (ZINC 4476281) and C2 (ZINC 0923924) show the ligands interacting with the amino acid residues as well as the type of interactions that occur. The interaction occurring with *S. enterica* HPPK can be seen in Figure 35 D1 (ZINC 4671178), E1 (ZINC 4476281) and F1 (ZINC 0923924) with amino acid interactions shown in Figure 35 D2 (ZINC 4671178), E2 (ZINC 4476281) and F2 (0923924) with the same ZINC compounds as that of *E. coli*. Interacting amino acids with bond type and binding center are shown in Table 4 and Table 5 for *E. coli* and *S. enterica* respectively.

Table 3: Amino acid residues forming part of the catalytic and ATP binding centers.

Organism	Binding Center	Amino Acid Residues
E. coli	Catalytic	Gly:8, Thr:42…Pro:44, Leu:45, Tyr:53, Asn:55, Leu:71, Gln:75, Glu:78, Arg:83, Arg:84, Trp:90, Leu:96, Asp:97…Ile:99,Thr:113, His:116…Tyr:117, Arg:121, Phe:123
	ATP	Leu:71, Gln:75, Glu:78, Arg:83, Arg:84, Arg:93, Asp:97…Ile:99, Thr:113, His:116…Tyr:117, Arg:121
S. enterica	Catalytic	Gly:9, Thr:43…Leu:46, Phe:54, Asn:56, Asp:71, Leu:75, Glu:78, Arg:85, Trp:90, Arg:93, Arg:96, Asp:98…Ile:99, Thr:113, His:116…Tyr:117, Arg:122, Phe:124
	ATP	Leu:71, Gln:75, Glu:78, Arg:83, Arg:85, Arg:93, Asp:98…Ile:99, Thr:113, His:116…Tyr:117, Arg:122

Table 3 defines the amino acid residues forming the catalytic and ATP binding centers of *E. coli* and *S. enterica* HPPK, showing the amino acids which interact directly with the HPPK substrate and ATP.

Table 4: In silico docking results of E. coli. Showing ZINC number, amino acid interaction, binding area and bond type.

ZINC Compound	Amino Acid Residue	Binding Center	Bond Type
0634089 (Figure 33 A1+2)	Ala A:6	Other	Hydrogen Bond

	Asp A:95	Other	Hydrogen Bond + Pi-Cation
	Arg A:121	Catalytic and ATP	Pi-Cation
	Phe A:123	Catalytic	Pi-Sulfur
	Tyr A:53	Catalytic	Pi-Pi Stacked
	Ala A:10	Other	Alkyl and Pi-Alkyl
	Pro A:51	Other	Alkyl and Pi-Alkyl
	Leu A:96	Catalytic	van der Waals (vdW)
	Gly A:8	Catalytic	vdW
	Ile A:7	Other	vdW
	Asp A:97	Catalytic and ATP	vdW
	Asn A:55	Catalytic	vdW
	Met A:124	Other	vdW
	Leu A:45	Catalytic	vdW
	Pro A:44	Catalytic	vdW
	Gln A:50	Other	vdW
	Arg A:84	Catalytic	vdW
	Pro A:91	Other	vdW
0310450 (Figure 33 B1 + 2)	Pro A:44	Catalytic	vdW
	Pro A:43	Catalytic	vdW
	Thr A:42	Catalytic	vdW
	Asn A:55	Catalytic	vdW

	Arg A:84	Catalytic and ATP	Alkyl + Pi-Alkyl
	Tyr A:53	Catalytic	Pi-Donor + Pi-Pi Stacked + Pi-Pi T-Shaped
	Leu A:45	Catalytic	Hydrogen Bond
	Gln A:50	Other	vdW
	Asp A:95	Other	vdW
	Ala A:10	Other	vdW
	Glu A:87	Other	vdW
	Gln A:50	Other	vdW
	Pro A:91	Other	Alkyl + Pi-Alkyl
	Phe A:123	Other	Pi-Pi Stacked + Pi-Pi T-Shaped
4499339 (Figure 33 C1 + 2)	Pro A:44	Catalytic	vdW
	Gly A:8	Catalytic	vdW
	Thr A:42	Catalytic	vdW
	Tyr A:53	Catalytic	vdW
	Phe A:123	Catalytic	Pi-Pi Stacked + Pi-Pi T-Shaped
	Arg A:84	Catalytic + ATP	Pi-Cation
	Arg A:121	Catalytic + ATP	Hydrogen Bond
	ASP A:95	Other	vdW
	Met A:124	Other	vdW

	Asn A:55	Other	vdW
	Leu A:45	Other	vdW
	Gln A:50	Other	vdW
	Pro A:91	Other	Alkyl + Pi-Alkyl
	Ala A:10	Other	Alkyl + Pi-Alkyl
	Tyr A:53	Other	PI-Donor + Pi-Pi Stacked + Pi-Pi T-Shaped + Alkyl + Pi-Alkyl
	Pro A:51	Other	Halogen (Fluorine)
4671178 (Figure 35 A1 + 2)	Asn A:55	Catalytic	vdW
	Leu A:45	Catalytic	vdW
	Pro A:44	Catalytic	vdW
	Arg A:84	Catalytic + ATP	Alkyl
	Tyr A:53	Catalytic	Pi-Pi Stacked + Pi-Pi T-Shaped
	Phe A:123	Catalytic	Pi-Pi Stacked + Pi-Pi T-Shaped
	Pro A:91	Other	vdW
	Gln A:50	Other	vdW
	Ala A:10	Other	vdW
	Pro A:51	Other	Alkyl
	Tyr A:42	Other	Halogen Bond

	Pro A:43	Catalytic	Hydrogen Bond
4476281 (Figure 35 B1 + 2)	Pro A:43	Catalytic	vdW
	Thr A:42	Catalytic	vdW
	Pro A:44	Catalytic	vdW
	Leu A:45	Catalytic	Pi-Donor + Pi-Alkyl
	Phe A:123	Catalytic	Pi-Pi Stacked
	Tyr A:53	Catalytic	Pi-Pi Stacked
	Arg A:84	Catalytic + ATP	Pi-Cation
	Asp A:52	Other	vdW
	Ala A:10	Other	vdW
	Gln A:50	Other	vdW
	Thr A:93	Other	vdW
	Pro A:91	Other	Pi-Alkyl
	Asp A:95	Other	Halogen (Fluorine)
	Arg A:92	Other	Halogen (Fluorine)
	Pro A:51	Other	Hydrogen Bond
0923924 (Figure 35 C1 + 2)	Asp A:95	Other	vdW
	Pro A:91	Other	vdW
	Ala A:86	Other	vdW
	Ala A:10	Other	vdW
	Leu A:45	Catalytic	vdW

	Pro A:44	Catalytic	vdW
	Pro A:43	Catalytic	vdW
	Phe A:123	Catalytic	Pi-Alkyl
	Arg A:84	Catalytic + ATP	Pi-Alkyl
	Tyr A:53	Catalytic	Pi-Pi Stacked
	Glu A:87	Other	Carbon Hydrogen
	Thr A:42	Catalytic	Hydrogen Bond
	Asn A:55	Catalytic	Hydrogen Bond

From Table 4 compounds bound to both the catalytic bind center as well as the ATP binding center of *E. coli* HPPK with bond occurring with amino acids that do not form part of the substrate or ATP interacting amino acids, yet still fall within the area in which the substrate and ATP dock.

Table 5: *In silico* docking results of *S. enterica*. Showing ZINC number, amino acid interaction, binding area and bond type.

ZINC Compound	Amino Acid Residue	Binding Center	Bond Type
0082136 (Figure 34 A1 + 2)	Asp A:96	Catalytic	Hydrogen Bond
	Arg A:122	Catalytic + ATP	Pi-Cation
	Phe A:124	Catalytic	Pi-Pi Stacked
	Phe A:54	Catalytic	Pi-Pi Stacked
	Arg A:93	Catalytic + ATP	Pi-Pi Stacked
	Leu A:46	Catalytic	Alkyl and Pi-Alkyl
	Gln A:51	Other	Alkyl and Pi-Alkyl
	Gly A:91	Other	van der Waals (vdW)

	Arg A:89	Other	vdW
	Trp A:90	Catalytic	vdW
	Pro A:45	Other	vdW
	Asn A:56	Catalytic	vdW
	Gly A:9	Catalytic	vdW
	Asp A:98	Catalytic + ATP	vdW
6792518 (Figure 34 B1 + 2)	Leu A:46	Catalytic	Hydrogen Bond
	Asn A:56	Catalytic	Hydrogen Bond
	Asp A:96	Catalytic	Hydrogen Bond
	Arg A:89	Other	Carbon Hydrogen + Alkyl
	Gln A:51	Other	Carbon Hydrogen
	Leu A:46	Catalytic	Carbon Hydrogen
	Gly A:9	Catalytic	Carbon Hydrogen
	Arg A:122	Catalytic + ATP	Pi-Cation
	Phe A:54	Catalytic	Carbon Hydrogen + Pi-Pi Stacked
	Phe A:124	Catalytic	Pi-Pi Stacked
	Pro A:45	Catalytic	vdW
	Asp A:98	Catalytic	vdW
	Gly A:91	Other	vdW
4498423 (Figure 34 C1 + 2)	Leu A:46	Catalytic	Hydrogen Bond + Carbon Hydrogen + Pi-

			Alkyl
	Arg A:89	Other	Carbon Hydrogen
	Phe A:54	Catalytic	Carbon Hydrogen + Pi-Pi Stacked + Pi-Alkyl
	Phe A:124	Catalytic	Pi-Sigma Bond + Pi-Pi Stacked
	Arg A:93	Catalytic + ATP	Pi-Alkyl
	Gln A:51	Other	vdW
	Gly A:47	Other	vdW
	Pro A:45	Catalytic	vdW
	Lys A:44	Catalytic	vdW
	Thr A:43	Catalytic	vdW
	Asn A:56	Catalytic	vdW
	Gly A:9	Catalytic	vdW
	Arg A:122	Catalytic + ATP	vdW
	Gly A:91	Other	vdW
	Trp A:90	Catalytic	vdW
4671178 (Figure 35 D1 + 2)	Asn A:56	Catalytic	Hydrogen Bond
	Asp A:96	Catalytic	Hydrogen Bond
	Leu A:46	Catalytic	Hydrogen Bond
	Gly A:9	Catalytic	Carbon Hydrogen
	Arg A:122	Catalytic + ATP	Pi-Cation

	Phe A:124	Catalytic	Pi-Pi Stacked
	Phe A:54	Catalytic	Pi-Pi Stacked
	Arg A:93	Catalytic + ATP	Alkyl
	Asp A:98	Catalytic + ATP	vdW
	Gln A:51	Other	vdW
	Pro A:45	Catalytic	vdW
4476281 (Figure 35 E1 + 2)	Arg A:93	Catalytic + ATP	Hydrogen Bond + Pi-Cation + Pi-Alkyl
	Leu A:46	Catalytic	Hydrogen Bond
	Phe A:54	Catalytic	Pi-Pi Stacked
	Gly A:9	Catalytic	vdW
	Asn A:56	Catalytic	vdW
	Phe A:124	Catalytic	vdW
	Pro A:45	Catalytic	vdW
	Gly A:47	Other	vdW
	Gln A:51	Other	vdW
	Arg A:89	Other	vdW
	Asp A:96	Catalytic	vdW
0923924 (Figure 35 F1 + 2)	Asn A:56	Catalytic	Hydrogen Bond
	Thr A:43	Catalytic	Hydrogen Bond
	Asp A:98	Catalytic + ATP	Halogen (Fluorine)
	Asp A:96	Catalytic	Halogen (Fluorine)

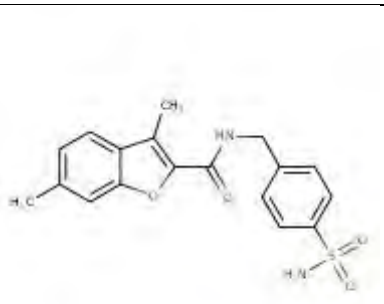
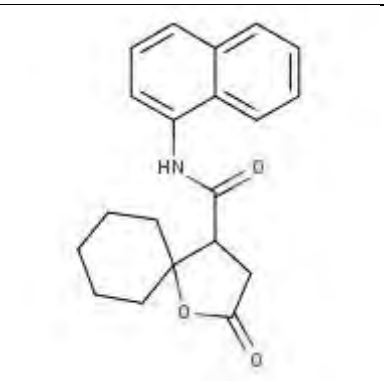
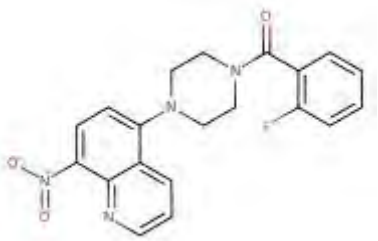
	Phe A:124	Catalytic	Pi-Pi Stacked
	Phe A:54	Catalytic	Pi-Pi Stacked
	Pro A:48	Other	Alkyl + Pi-Alkyl
	Leu A:46	Catalytic	Alkyl + Pi-Alkyl
	Gly A:9	Catalytic	vdW
	Arg A:93	Catalytic + ATP	vdW
	Gln A:51	Other	vdW
	Gln A:49	Other	vdW
	Gly A:47	Other	vdW
	Pro A:45	Catalytic	vdW

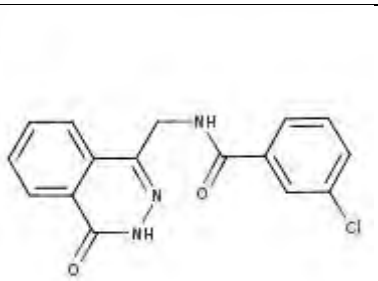
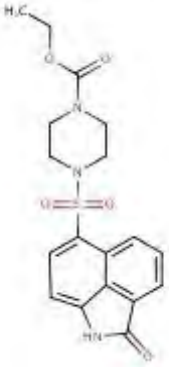
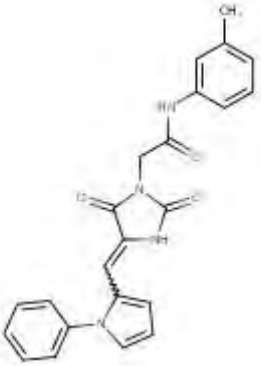
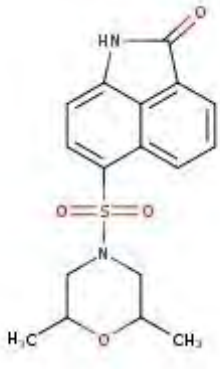
From Table 5 compounds bound to both the catalytic binding center as well as the ATP binding center of *S. enterica* HPPK with bond occurring with amino acids that do not form part of the substrate or ATP interacting amino acids, yet still fall within the area in which the substrate and ATP dock.

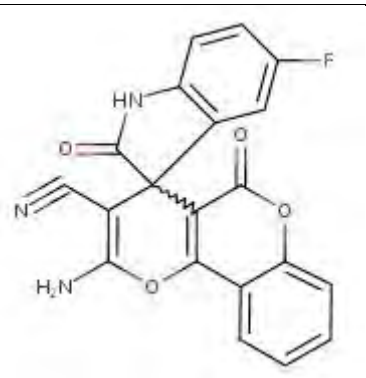
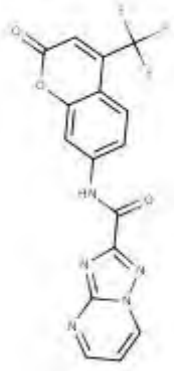
All compounds are seen to have a short binding distance and bonds with low binding energy. Bond types that show good bonding are also observed - these are conventional hydrogen bonds, pi-interactive forces and van der Waals forces. Halogen bonds also occurred in some compounds. Some amino acids have formed more than one type of interaction with the selected binding pocket, which suggests enhanced binding to the enzymes and potentially a stronger inhibitor. The majority of the amino acids which interacted with the ZINC compounds interact with the amino acids forming part of the catalytic or ATP binding centers, specifically those interacting with the substrate and ATP. The others form part of the binding centers but are not directly involved with the substrate or ATP interaction.

Table 6, below, shows the structures and IUPAC names of the docked compounds, as sourced from the Molport® webpage, which was used to procure the compounds for further *in vitro* experimentation.

Table 6: Zinc ID, IUPAC name and structure of the top ligands docked against *S. enterica* HPPK and *E.* coli HPPK enzymes.

Zinc ID	IUPAC Name	Structure
ZINC00634089	3,6-dimethyl-N-[(4-sulfamoylphenyl)methyl]-1-benzofuran-2-carboxamide	
ZINC00310450	N-(naphthalen-1-yl)-2-oxo-1-oxaspiro[4.5]decane-4-carboxamide	
ZINC04499339	5-[4-(2-fluorobenzoyl)piperazin-1-yl]-8-nitroquinoline	

ZINC00082136	3-chloro-N-[(4-oxo-3,4-dihydrophthalazin-1-yl)methyl]benzamide	
ZINC6792518	Ethyl-4-({3-oxo-2-azatricyclo[6.3.1.0^{4,12}]dodeca-1(11),4(12),5,7,9-pentaen-9-yl}sulfonyl)piperazine-1-carboxylate	
ZINC04498423	2-{2,5-dioxo-4-[(1-phenyl-1H-pyrrol-2-yl)methylidene]imidazolidin-1-yl}-N-(3-methylphenyl)acetamide	
ZINC04671178	9-[(2,6-dimethylmorpholin-4-yl)sulfonyl]-2-azatricyclo[6.3.1.0^{4,12}]dodeca-1(11),4(12),5,7,9-pentaen-3-one	

ZINC04476281	2'-amino-5-fluoro-2,5'-dioxo-1,2-dihydro-5'H-spiro[indole-3,4'-pyrano[3,2-c]chromene]-3'-carbonitrile	
ZINC00923924	N-[2-oxo-4-(trifluoromethyl)-2H-chromen-7-yl]-[1,2,4]triazolo[1,5-a]pyrimidine-2-carboxamide	

2.4 Discussion

The modelling of the *S. enterica* HPPK enzyme from that of the crystal structure of the *Y. pestis* HPPK homologue showed positive and excellent results. The sequence similarity between the two enzymes is large with 66% of amino acids being the same. The sequence in the *Y. pestis* HPPK structure covers a large number of amino acids (96%) of those of *S. enterica* HPPK. The crystal structure used as a template had a resolution of 1.80 which allows for accurate modelling of the enzyme, as resolution determines the level of detail which can be observed. Both proteins have the same number of amino acid residues and perform the same function in their respective native organisms, thus we can presume that the modelled structure will reflect its native structure under physiological conditions in *S. enterica*. The Dope Z-Score and the Z-Score was sufficient to ensure that the modelled enzyme could be used in further docking studies. The amino acid sequence identity between *E. coli* HPPK and *S. enterica* HPPK is 74.84% (Appendix A), as compared to the 66% identity between *Y. pestis* HPPK and *S. enterica* HPPK. Unfortunately, PRIMO did not identity an *E. coli* HPPK crystal structure with a high enough resolution to use for modelling. Had a crystal structure with a better resolution been available the resulting model could have been of a higher quality. The possibility exists that an alternative organism contains an HPPK homologue with higher sequence similarity to that of *S. enterica* HPPK to provide a better template for modelling, but such a crystallized structure is not available at present.

A crystallized version of the *S. enterica* HPPK would have been better for use in *in silico* docking and may have yielded more compelling results. Nonetheless, the currently used model showed good docking results and even overlap of compounds between itself and the results obtained for *E. coli* HPPK, which is to be expected as they have such a large sequence similarity. The top molecules that were docked also show similarity in that they have an aromatic ring structure with a tail. This can also be expected as most receptor pockets have an area for binding a large structure followed by a tail. Looking at the natural substrate of HPPK, HMDP, it consists of a fused ring structure with short chains, which is to be joined to a pyrophosphate moiety obtained from ATP. ATP is a long chain molecule, and as the catalytic binding sites for the substrate and ATP are next to each other, it can be expected that a fused ring structure with a long chain attached to it would bind well. The program was set to dock molecules within the binding pockets for HMDP and ATP. This clearly shows an overlap between the catalytic or substrate-binding site and the ATP binding site, which is to be expected as the function of the enzyme is to obtain a pyrophosphate moiety from ATP and bind it to HMDP. The ligands docked also show overlap between the residues they bound to and those of the binding pocket, as this was the area set to dock, but other residues have also been used in their binding.

Alternatively, as an approach to find novel binding areas the program could have been set to dock blindly to any part of the enzyme which could have resulted in very different results. An allosteric site binding can be seen as a redistribution of the conformational states by the binding of an effector molecule at a distal site, other than the enzymes catalytic or orthosteric site, which will usually result in structural perturbations (Grover, 2013). These results also occur as a result of mutations, changes in pH, temperature, ionic strength and covalent modifications such as phosphorylation. Currently, computation methods for the identification of allosteric sites are underway, which would predict ligand-protein interactions and binding, with the use of molecular dynamics and network-centric perspectives (Amamuddy *et al.*, 2020). Temiz and Bahar (2002), as well as Kurt et al. (2003), had noticed that the HIV protease and reverse transcriptase enzymes had different forms when in the apo state and when in the ligand-bound state, which caused the receptors to have distinct dynamic or communication networks. An advantage of targeting allosteric sites is the fact that allosteric sites have a low degree of sequence conservation, making species-specific drug design a major benefit (Ayyildiz *et al.*, 2020). The allosteric sites have been highlighted as having the potential to have fewer adverse effects. An investigation for the treatment of schizophrenia using a drug that binds to an allosteric site on muscarinic (M1) receptors altering the neurotransmission of acetylcholine is investigating the lowered adverse effect profile presented allosteric site binding of drugs (Wenthur *et al.*, 2014). If enzymes do possess a high degree of sequence conservation, such as can be found between *E. coli* HPPK and *S. enterica* HPPK, the use of allosteric sites could present the opportunity to design a broad-spectrum drug molecule. The current literature proves promising with regards to the future of allosteric sites for use as a treatment option and is worth further exploration in the antimicrobial sphere.

AutoDock Vina was launched in 2010 and has been a trustworthy tool employed in the drug discovery pipeline since then, producing numerous excellent results (Nguyen *et al.*, 2020). It is straightforward to use and allows for several input and output settings. Since its introduction, several programs have become available which have algorithms that allow the software to take on a more physiologically simulated environment in which docking is performed. These include programs such as Maestro Schrödinger, 1-Click Docking, BSP-SLIM, CABS-dock, FlexPepDock and GalaxyPepDock to name a few (Pagadala, Syed and Tuszynski, 2017). The approach which most programs use is shape matching, systematic searching (which is an exhaustive search and also uses fragmentation and conformational ensembles) and stochastic algorithms (which include Monte Carlo methods, genetic algorithms, Tabu search methods and swarm optimisation methods) (Wang *et al.*, 2016). The scoring methods used are force field, empirical and knowledge-based. More recently methods such as quantum mechanical and semi-empirical quantum mechanical based scoring

algorithms have been introduced to better capture the binding affinity trend and native pose identification. A shortcoming of the software is that it does not allow for possible changes in enzyme structure, which could result in a biological setting due to factors such as temperature, pH and osmolality to mention a few (Nguyen *et al.*, 2020). Regardless of its shortcomings, it has still provided results in docking and has grouped the molecules in binding energy and binding distance. This provides a clear picture as to the predicted strength of particular interaction between the receptor and the ligand.

Docking studies also provided the different types of bonds which would occur between the receptor and the ligand. This one preferentially looks for conventional hydrogen bonds as these represent a strong bond. Other bonds which are of importance are pi-interactive forces, halogen bonds and also van der Waals forces, all of which show strong binding affinities. Halogen bonds have generated debate and differing opinions, in past *in silico* docking programs have been unable to verifiably estimate their bond strengths when compared to hydrogen bonds or van der Waals forces (Kurczab, Kucwaj-Brysz and Śliwa, 2020). As a result of difficulties faced with halogen bonds, new approaches have been a quantum-mechanical-based molecular docking approach and also using algorithms that look for amino acids more likely to bond with halogens (Koebel *et al.*, 2016). This was done by calculations of the energies of ligands bound to proteins, either containing halogens or not and those same ligands but in a non-halogenated form. Halogens do play a factor in drug optimisation, where they are used to make molecules more lipophilic allowing them to cross cell membranes more rapidly (Tyagi *et al.*, 2020). This is important in enabling absorption and tissue distribution for compounds in oral formulations, as well as access to the internal environment of organisms where the intended drug targets are found. Thus in my opinion molecules possessing a halogen atom would be advantageous further down the line for drug optimisation.

The ZINC library of compounds was used, specifically the ZINC 15 subset. The molecules were further refined to take into account Lipinski's rule of 5, resulting in molecules that have a high likelihood of being good drug molecules. Results obtained show that the molecules yielded promising potential inhibitors *in silico* and thus would make good drug candidates for further exploration (Lipinski, 2000). This would also result in minimal optimisation if any of the molecules show promising results in target-based assays. While Lipinski's rule of 5 provides a good guideline for drug-like compounds, it is not absolute and exceptions are known. The rule of 5 is an attempt to maximise the chances of obtaining a drug-like molecule, which is orally bioavailability, looking at a vast set of known drug compounds. Lipinski (2000), stated the number of drug-like compounds is quite limited and the chances of obtaining hit compounds from the screening of millions of compounds are very small.

This has brought about the examination of chemical compounds of the basis of the properties they possess, specifically as they relate to use in a biological setting (Tyagi *et al.*, 2020). ADME (absorption, distribution, metabolism and elimination) properties and how they would translate to drug ability is a key feature, as all drugs are intended for biological use. Thus Lipinski proposed the rule of 5 stating that for a chemical compound to be biologically active the following set of parameters need to be met: less than 5 H-bond donors, less than 10 H-bond acceptors, a molecular weight of less than 500 and a calculated LogP of not more than 5. However, the rule of 5 is not absolute, and it has been noted that as many as half of all drug targets presumed to be involved in disease will not be modulated by a drug that follows the rule of 5 (Doak *et al.*, 2014). This is clear from the fact that binding sites of protein-protein interactions are better modulated with a drug that goes beyond the chemical space of the rule of 5, and in-depth studies have revealed that cell-permeable and orally-absorbed drugs also exist in the space beyond the rule of 5 (Tyagi *et al.*, 2020). One could thus possibly look into expanding the number of molecules used for the docking, as this could result in unlikely molecules showing up which could produce better docking into the receptor site. A larger set of molecules docked could also likely produce molecules that could possess more desirable bonds when interacting with the receptor. Larger and more complex molecules could cover a larger number of amino acid residues with the receptor pocket, as currently the compounds chosen as the most successful, do not cover the entire binding pocket of the enzyme (Doak *et al.*, 2014).

Further experimentation with the enzyme in an *in silico* environment could include molecular dynamic studies, which would help better understand the mechanisms of the enzyme, and also the importance of metal coordination ions within the enzyme. One could also attempt to predict the ideal environment in which the enzyme would work optimally which would greatly assist in the design and implementation of *in vitro* assays. Large scale phylogenetic trees and sequence alignments would assist in determining similar enzymes and allow the generation of compounds worth possible broad-spectrum activity against a wide variety of pathogens.

In conclusion, the bioinformatics studies performed in this section have been successful and has assisted in developing a working model of the HPPK enzyme of *S. enterica*. This model was docked, along with that of *E. coli* HPPK, against several drug-like molecules, which allowed the identification of compounds predicted to bind effectively to the enzymes, with strong affinity and good bond types. The molecules are varied, but also show some similarity in size and shape. These molecules were used for *in vitro* studies described in subsequent chapters of the dissertation.

Chapter 3

Protein Expression and Purification

3.1 Introduction

Protein expression and purification are used to obtain a protein of interest for the characterisation of its function and structure, possible interactions of the protein and use in the development of novel agents for disease modification and prevention, to name a few. The process was once left only to experts in the field, but simplification and advances made in technology have resulted in the widespread use, adaptation and application of the process (Gräslund *et al.*, 2008). Each project using recombinant proteins will follow a specific protocol to isolate the protein of interest. Since proteins are unique and their uniqueness is further complexified by the strategies chosen for production, there is no one-size-fits-all approach to protein expression and purification, and empirical experimentation may be required to find the most appropriate method for a particular protein of interest (Nallamsetty and Waugh, 2007).

In an extensive review study of over 10 000 expressed and purified proteins by Gräslund et al. (2008) the process, consisting of several steps, can be summarised as follows: 1) The complementary DNA (cDNA) is obtained either by amplification of genomic DNA (prokaryotic genes, or eukaryotic genes lacking introns) or full-length cDNA in which the sequence has been verified (eukaryotes) or by the synthesis of the complete gene; 2) The cDNA is then cloned into an expression vector with the use of ligation-independent cloning (LIC) adding a T7 RNA polymerase-driven expression and N-terminal oligohistidine tag (a cleavage site needs to be specified for a sequence-specific protease, which will allow the removal of this tag; 3) the recombinant protein is then expressed in a chosen expression organism, generally at a temperature at which the organism will thrive. The possibility exists that the expression organism may have codon biases differing from that of the protein's native organism, in which case supplementation with the appropriate tRNA genes need to occur, or host organism codon optimisation can occur. 4) The protein will then be solubilised and purified in a buffered solution of a monovalent salt appropriate to the process, at sufficient strength. 5) Preliminary purification should be done using immobilised metal affinity chromatography (IMAC), after which size-exclusion chromatography and eventually ion-exchange chromatography can be done to further purify and polish the protein. 6) The removal of the affinity tag should be done to minimise non-native sequences in the recombinant protein.

Artificial synthesis of genes refers to the creation of genes with the use of a DNA template sequence, thus assembling nucleotides *de novo*. This allows for the synthesis of genetic material which proves difficult to obtain or even the creation of completely new and atypical genes (Hughes, Miklos and

Ellington, 2011). The process consists of two main steps. The first is the production of oligonucleotide fragments which are usually shorter than 200 base pairs, known as solid-phase DNA synthesis or DNA printing. The second step then connects these short oligonucleotide fragments using different assembly protocols. The oligonucleotides are chemically synthesised from either normal or modified nucleotides to protect the amines, hydroxyl or phosphate group from improper interactions (van den Brulle *et al.*, 2008).

Once the gene is produced the polymerase chain reaction (PCR) is commonly used to amplify the quantity of the gene for insertion into a plasmid vector. PCR can also be done using a DNA template as a starting point. Essentially, PCR introduces restriction enzymes sites at the terminals of the DNA amplicon using primers, to enable cleavage of the desired gene (Valones *et al.*, 2009). The same (or sticky end compatible) restriction enzymes are used to digest the plasmid vector, which results in the gene connecting to the plasmid vector forming a recombinant DNA molecule with the application of ligase enzymes. For protein expression, plasmids are optimised to ensure that only essential DNA and the gene of interest remain, thus minimising the chance of mutations from taking place (Marsic *et al.*, 2008). Typically an expression vector will contain the following essential elements, namely: the gene of interest expression cassette (which will include a promotor and a gene-termination signal), a bacterial origin of replication and a selected antibiotic resistance cassette for a particular host. Additional elements of the vector may include multiple cloning sites, epitope tags, secretion signals, protease recognition sites and internal ribosome entry sites (Withers-Martinez *et al.*, 1999). For successful expression, a DNA sequence is added to the vector, called a promotor system, such as T7 or *tac* and they are induced by isopropyl β-D-1-thiogalactopyranoside (IPTG) or an autoinduction method which manipulates carbon sources during the growth phase of the host (Chen, 2012). The host organism which will be used for expression also plays a role in the optimisation process. As the majority of recombinant proteins are expressed using already optimised *E. coli*, the plasmids are further optimised to be suitable to the *E. coli* intracellular environment thus increasing the chances of successful expression of the recombinant proteins. The plasmid DNA is introduced into the host organism by transformation, and the organism is then cultured to increase the number of cell biomass and thus plasmid or expressed protein yields (van den Brulle *et al.*, 2008).

For the expression of a recombinant protein, there are a variety of host organisms available throughout the biological kingdoms, including bacteria, yeast, mammalian and insect cells (Terpe, 2006). Although this large variety of expression hosts exists, the most common host is *E. coli*, which has been optimised for this purpose. Some strains of the bacterium have had all non-essential DNA

removed for research purposes (Lesley, 2001). Despite its extensive use as a research bacterium, *E. coli* can pose difficulties with expression, such as misfolding of proteins, the formation of inclusion bodies, the potential of DNA mutations and little to no protein activity or cell death due to protein toxicity, to name a few (Assenberg *et al.*, 2013). The aforementioned makes the analytical scale expression of proteins important. The analytical scale expression followed by gel electrophoresis analysis will provide insights into the expression profile of a protein with clues to any problems which could have occurred, e.g. low or absent expression of the protein, expression of truncated or degraded protein, expression of the protein in an insoluble form or inhibition of *E. coli* growth due to protein toxicity. The initial factor which is to be accessed is the solubility of the recombinant protein. This is done by separation of the pellet and supernatant by centrifugation after the expression culture has been lysed, and analysis of the protein composition of the fractions by the use of sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) (Nallamsetty and Waugh, 2007). Protein solubility will often motivate the process of expression optimisation for use during subsequent large-scale expression. Optimisation methods include, but are not limited to, the temperature at which the protein is expressed, the duration of time the protein is allowed to express for, the concentration of promotor inducer (IPTG), possible inclusion of an overexpressed chaperone in the host organism or increasing the expression culture volume (Han *et al.*, 2007) (Nguyen *et al.*, 2004).

The large-scale expression is performed to produce a relatedly large quantity of the recombinant protein and is performed using the optimal condition as determined by analytical scale expression. After cells lysis a large number of cellular components will be in the lysate including the protein of interest, thus purification needs to be performed. Separation of the protein of interest and other cellular components usually makes use of protein size, physicochemical properties, binding affinities and biological activity for separation to take place (Lesley, 2001). Based on its efficiency and convenience, affinity chromatography, which takes advantage of a fusion tag attached to the protein of interest is commonly used to separate it from other cellular components (used in this study). Other techniques include size exclusion chromatography, hydrophobic interaction chromatography, ion-exchange chromatography, free-flow-electrophoresis and high-performance liquid chromatography (HPLC) (Esposito and Chatterjee, 2006) (Han *et al.*, 2007). Affinity chromatographic purification requires that a genetic sequence be inserted into the plasmid vector which will result in the protein being expressed with the tag. The tagged portion of the protein will then be attracted to and bind to the specific marker on a resin which can later be washed off and the protein of interest can be freed. Commonly used tags include chitin-binding protein (CBP), maltose-binding protein

(MBP), streptavidin tag (Strep-tag), glutathione-S-transferase (GST), polyhistidine tag (His-tag), calmodulin tag and polyglutamate tag, to name a few (Assenberg *et al.*, 2013) (Chen, 2012).

In this study, all proteins were His-tagged and purified by the use of IMAC. Amino acid side chains can interact with transitional metals to form co-ordination covalent bonds, as seen with histidine (Hoffmann and Roeder, 1991). Proteins containing amino acids with an affinity for metal ions will be retained on a column containing a resin or beads which have immobilised metals on them. The metals could include cobalt, nickel or copper for the purification of histidine-containing proteins or peptides, or iron, zinc or gallium for the purification of phosphorylated proteins or peptides. For proteins to be eluted the pH of the wash buffer can be changed or the concentration of the imidazole, which acts as a metal ion ligand, can be increased. IMAC allows for fast, strong and specific binding of proteins and also has a high yield, thus making it the preferred method for protein purification (Assenberg *et al.*, 2013).

The concentration of the purified expressed proteins can be determined by the use of a Bradford assay. It is a colorimetric protein assay using Coomassie Brilliant Blue G-250, which measures absorbance at 595nm. The dye occurs in a red form under acid conditions, blue under basic conditions and is green at a neutral pH (Gräslund *et al.*, 2008). The Coomassie dye forms a strong noncovalent complex with the carboxyl group of the proteins via van der Waals forces and electrostatic interactions with the amino group of the protein. Ionisable groups on the protein accept electrons from the dye causing a disruption of its native state exposing hydrophobic pockets within the tertiary structure of the protein, binding non-covalently to non-polar regions of the dyes which will eventually cause a stabilization towards the blue colour, indicating a protein-dye complex has formed (Nallamsetty and Waugh, 2007). The intensity of the colour indicates the amount of protein present, which is confirmed by an absorbance reading and with the aid of constructing a standard curve.

Three enzymes belonging to the folic acid pathway were expressed using *E. coli* bacteria. The enzymes form part of the bacterial *de novo* biosynthetic pathway (*S. enterica* HPPK and *E. coli* HPPK) and the protozoan auxotrophic mechanism (*T. brucei* PTR1). All proteins were His-tagged and were expressed in their full length, namely *S. enterica* HPPK (17 kDa), *E. coli* HPPK (17 kDa), and *T. brucei* PTR1 (21 kDa). The goal of this chapter was to:

1) Demonstrate that the enzymes were able to be expressed in *E. coli* cells;
2) Optimise the expression of the enzymes;

3) Purify the enzymes from the optimised expression and determine the concentrations for further plate-based enzyme activity assays;

3.2 Methods and Materials

3.2.1 Transformation of T7 Express *lys*Y *E. coli* cells

The method used was adapted from Cutler, (2003). The coding sequences of *T. brucei* PTR1, *S. enterica* HPPK and *E. coli* HPPK, (Appendix B), were codon-optimised for *E. coli* expression, synthesised and cloned between the *Nh*el and *XH*ol restriction sites of the expression plasmid pET-28a (Appendix C, Figure 58) by GenScrpit (Hong Kong). The plasmid enables the IPTG-inducible expression of N-terminal His-tagged proteins and the selection of transformed *E. coli* cells using kanamycin. The plasmids were supplied as 4 µg lyophilised stocks and dissolved in 20 µl water to enable *E. coli* transformation. Competent T7 Express *lys*Y *E. coli* cells (New England Biolabs) were thawed on ice of which 50 µl were transferred to a sterile microfuge tube. The plasmid (0.5 µl) of the plasmid was added to the microfuge tube and mixed by briefly flicking the bottom on the microfuge tube. The mixture was incubated on ice for 30 min, whereafter it was placed in a heating block pre-heated 42.5 °C for 60 sec before being placed on ice for a further 5 min. 500 µl of Luria Broth (LB) was added to the microfuge tube and incubated at 37 °C for 1 hour. 50 µl of the suspension was plated onto LB-Agar plates containing 50 µg/ml kanamycin and incubated at 37 °C overnight in an inverted position.

3.2.2 Long term storage of transformed bacterial colony

The method used was adapted from Cutler, (2003). One colony containing the plasmid was picked from the agar using a micropipette tip and inoculated into LB broth containing 100 µg/ml kanamycin and incubated with shaking overnight at 37 °C. Previously sterilised glycerol was added to 850 µl of the cultured bacterial stock to reach a final concentration of 17 % (v/v) and stored at -80 °C.

3.2.3 Analytical scale protein expression

The method used was adapted from Cutler, (2003). Three ml of previously autoclaved LB broth was added to autoclaved McCartney bottles and kanamycin was added to a final concentration of 100 µg/ml. A scraping of the cryopreserved bacterial glycerol stock was added to the McCartney bottle and incubated overnight at 37 °C with shaking. Two McCartney bottles were prepared by adding 8 ml of LB broth containing 100 µg/ml kanamycin; one labelled induced and the other uninduced. The uninduced bottle was used as a control. Each bottle with the broth was inoculated with 400 µl of the overnight culture and incubated at 37 °C with shaking for 1.5 hours. The optical density (OD) of each culture was determined at 600 nm using a spectrophotometer, to ensure that the OD_{600} fell between 0.5-0.9. To the induced McCartney bottle, a 1 M IPTG stock solution was added to create a final concentration of 1 mM. Both cultures were incubated at 37 °C for 3 hours with shaking. The expression cultures were decanted into Beckman-Coulter JA 20 rotor tubes and centrifuged at 4 °C and 5000 *xg* for 10 min using an Avanti® J-E Centrifuge. The supernatant was discarded and the

pellets were washed using 20 mM Tris-HCL (pH 8.0) and re-centrifuged at 4 °C and 5000 *xg* for 10 min. The supernatant was discarded and the pellets were frozen at -20 °C. The pellets were thawed on ice for 12 min and resuspended in 1 ml 20 mM Tris-HCL (pH 8.0). A 10 mg/ml lysozyme stock solution was added to create a final concentration of 1 µg/ml, swirled briefly and incubated on ice for 30 min, swirling every 10 min. Each tube was sonicated twice for 45 sec at 60 Hz using a Vibra-Cell® sonicator, with a 1 min resting period on ice between the sonication steps. The lysed *E. coli* cells were centrifuged at 5000 *xg* for 10 min at 4 °C before the supernatant (soluble fraction) and pellet (insoluble fraction) were analysed using SDS-PAGE (the insoluble material was resuspended in a volume of water equal to the volume of the soluble fraction prior to SDS-PAGE analysis).

3.2.4 SDS-PAGE Analysis

The method used was adapted from Cutler, (2003). Samples were analysed using a 12 % (w/v) SDS-PAGE gel consisting of a resolving gel (4 ml [30 % (w/v)] acrylamide, 2.5 ml lower gel buffer [1.5 M Tris-HCL, 0.4 % (w/v) SDS, pH 8.8], 3.5 ml water, 35 µl 10 % (w/v) ammonium persulfate and 7 µl TEMED [N,N,N´,N´-Tetramethyl-ethylenediamine]) and a 4 % stacking gel (0.7 ml [30 % (w/v)] acrylamide, 1.25 ml stacking gel buffer [0.5 M Tris-HCL, 0.4 % (w/v) SDS, pH 6.8], 3 ml water, 25 µl 10 % (w/v) ammonium persulfate and 6 µl TEMED]. 60 µl of each sample (soluble and insoluble fractions of the induced and uninduced *E. coli*) was added to sterile microfuge tubes and 20 µl of 4 X SDS sample buffer (10 ml stacking buffer, 8 ml glycerol, 0.8 g SDS, 0.8 ml 2-mercapto-ethanol, 0.2 mg bromophenol blue, 1.2 ml water) was added. The samples were placed in boiling water for 5 min. The samples were loaded onto the gel with lane 1 containing Color Prestained Protein Standard (New England Biolabs). The analysis was done using electrophoresis in a Mini-PROTEAN Tetra System in SDS running buffer (25 mM Tris-HCL, 0.2 M glycine, 3.5 mM SDS) for 1.5 hours with a constant voltage of 120 V. The gels were removed and stained using a Coomassie stain (45 % [v/v] water, 45 % [v/v] methanol, 10 % [v/v] acetic acid, 0.25 % [w/v] Coomassie Brilliant Blue R-250) overnight at room temperature with gentle shaking. Once staining was complete, the staining solution was decanted and the gels were destained using a destaining solution (44 % [v/v] water, 44 % [v/v] methanol and 11 % [v/v] acetic acid) at room temperature with gentle shaking. The solution was changed every 20 min until the protein bands were visible. Photographing of the gels was done using the ChemiDoc® XRS+ gel documentation system (Bio-Rad).

3.2.5 Large-scale protein expression.

The method used was adapted from Cutler, (2003). A scraping of the bacterial glycerol stocks was placed in McCartney bottles containing 5 ml LB broth, to which a 10 mg/ml kanamycin stock was added to a final concentration of 0.05 mg/ml. The cultures were incubated overnight at 37 °C with shaking. 5 ml of the overnight culture was inoculated into 250 ml LB broth containing kanamycin and

incubated at 37 °C with shaking until the OD_{600} had reached between 0.5-0.9. To the expression culture, a 1 M ITPG stock was added to a final concentration of 1 µM and the culture was incubated at 12 °C for 24 hours. Cultures were decanted into autoclaved Beckman-Coulter J-14 tubes and centrifuged at 5000 *xg* and 4 °C for 10 min using an Avanti® J-E Centrifuge. The supernatant was discarded and the pellet was washed and resuspended with 20 ml of cold binding buffer (50 mM Tris-HCL containing 20 mM imidazole, pH 8.0). The cultures were re-centrifuged at 5000 *xg* at 4 °C for 10 min, the supernatant was discarded and the pellet was stored overnight at -20 °C.

3.2.6 Protein Purification using Ni-NTA column for His-tagged Proteins

The method used was adapted from Spriestersbach *et al.* (2015). The IPTG-induced *E. coli* pellet was thawed on ice for 30 min and then resuspended in cold binding buffer to which 2 ml of a 10 mg/ml lysozyme stock solution was added and incubated on ice for a further 30 min swirling gently every 10 min. The bacterial suspension was sonicated for 1 min at 60 Hz at 4 °C and rested for 1 min on ice before the sonication was repeated, using a Vibra-Cell® sonicator. The lysate was centrifuged at 14 000 *xg* for 30 min at 4 °C using a JA14 rotor. The supernatant was decanted into 50 ml tubes and stored on ice. Filtration of the supernatant was performed using a 0.45 µm and then a 0.2 µm filter. A 120 µl sample of the filtered lysate was collected for SDS-PAGE analysis. A Qiagen Fast Start kit Ni-NTA column was opened slightly and its contents were shaken very slightly. The storage solution was allowed to run-through, after which the column was washed with filter sterilised water and equilibrated using the binding buffer. The previously filtered lysate was allowed to run through the column after which a 120 µl sample was collected for SDS-PAGE analysis. The column was washed twice using the 5 ml binding buffer and 120 µl samples of each wash were collected for SDS-PAGE analysis. The protein was eluted from the column by applying 3 ml elution buffer (50 mM Tris-HCL, 0.5 M imidazole, pH 8.0) and 120 µl of the elution was collected for analysis, while the remaining elution was stored at 4 °C. The column was rinsed using filter sterilised water and a 120 µl sample was collected to ensure all desired protein had been removed. Once washed the column was stored in a 50 % (v/v) ethanol solution at 4 °C.

3.2.7 Cleaning and Recharging the Ni-NTA column

The method used was adapted from Spriestersbach *et al.* (2015). Once protein purification was complete, the Ni-NTA column was rinsed by filling the column with filter sterile water and allowing it to run through. Stripping buffer (20 mM sodium phosphate, 500 mM NaCl, 50 mM EDTA, pH 7.4) was added to the column and allowed to run through. The column was rinsed with water and 0.1 M $NiSO_4$ was loaded onto the column to replace the lost Ni. The column was rinsed with water and then filled with a 50 % (v/v) ethanol solution and stored at 4 °C. The cleaning and recharging of the column were done at room temperature.

3.2.8 Desalting of Purified Proteins.

The method used was adapted from Spriestersbach *et al.* (2015). Desalting of the proteins was done to remove imidazole using an 8.3 ml Sephadex® G-25 PD-10 Desalting Column (GE Healthcare). The storage solution was allowed to run through before 5 full volumes of Tris Protein Storage Buffer (TPSB) (25 mM Tris-HCL, 15 mM NaCl, 0.1 mM $MgCl_2$, 0.03 mM BSA [Bovine Serum Albumin], 0.2 mM 2-mercapto-ethanol) was applied to the column to equilibrate it. Subsequently, 3 ml of the Ni-NTA column eluate containing the protein was loaded onto the column. To elute the protein 5 ml TPSB was added to the column. A 120 µl aliquot of the eluate was collected for SDS-PAGE analysis. The desalted protein eluate was mixed with glycerol to achieve a final concentration of 40 % (v/v) and stored at -80 °C. The desalting column was washed with filter sterile water twice and filled for storage with a 50 % (v/v) ethanol solution. The desalting procedure was carried out at 4 °C.

3.2.9 Protein concentration determination using Bradford assay

The method used was adapted from Spriestersbach *et al.* (2015). To construct a standard curve, a set of BSA protein standards were prepared in a serial dilution from 1.5-0.125 mg/ml. To a sterile multi-pipette reservoir, 5 ml of Bradford's reagent was added and allowed to equilibrate to room temperature. The respective samples (protein standards, collected purified protein samples, equilibration buffer and elution buffer) were added to a clear 96 well flat-bottom plate, 5 µl in each well, in triplicate, and 250 µl Bradford's reagent was added. Incubation of the plate took place at room temperature for 5 min before the absorbance was read at 595 nm using a SpectraMax® M3 plate reader (Molecular Devices). A standard curve was constructed from the average of the protein standards readings after subtracting background readings obtained with buffer samples alone and used to determine the concentration of the protein samples.

3.3 Results

3.3.1 Analytical scale Protein Expression

3.3.1.1 Analytical scale Protein Expression at 37 °C

For determination of the expression and solubility of the *E. coli* HPPK, *S. enterica* HPPK and *T. brucei* PTR1 proteins, analytical scale expression of the recombinant proteins was performed in T7 Express *lys*Y *E. coli* cells.

Soluble supernatants and insoluble pellets of IPTG-induced and uninduced cultures were prepared and analysed by Coomassie-stained SDS-PAGE gels. Results for *E. coli* HPPK and *S. enterica* HPPK are shown in Figure 36 and *T. brucei* PTR1 in Figure 37.

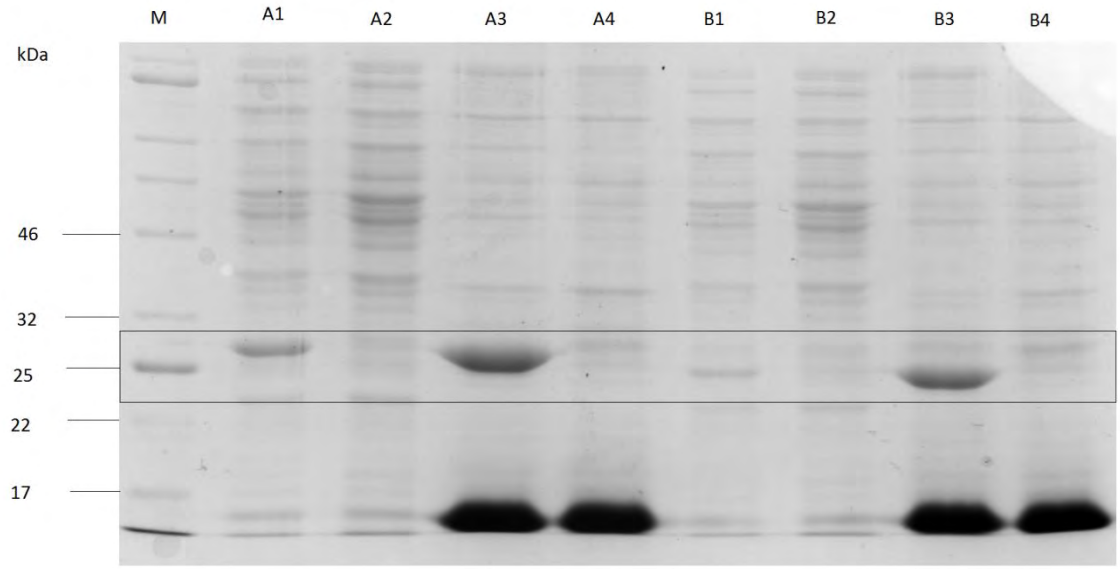


Figure 36: SDS-PAGE analysis of *E. coli* HPPK (A) and *S. enterica* HPPK (B) analytical scale expression conducted at 37 °C for 3 hours. Analytical scale expression samples were analysed by SDS-PAGE on a 12 % gel which was stained with Coomassie. M: Molecular mass marker; Lane 1: Induced soluble fraction; Lane 2: Uninduced soluble fraction; Lane 3: Induced insoluble fraction; Lane 4: Uninduced insoluble fraction.

Figure 36, A (*E. coli* HPPK) and B (*S. enterica* HPPK) show that the soluble fractions contained the majority of protein bands visible throughout the full molecular weight range, indicating that the *E. coli* cells were successfully lysed to release most of their protein content. Compared to the uninduced soluble fractions (lane 2), more intensely stained bands were observed in the vicinity of the 25 kDa marker in the induced soluble fractions (lane 1) which is desired as this is the predicted size of the HPPK proteins (17 kDa) based on their amino acid sequence with the addition of the His-tag. Lane 3 of Figure 36, A and B, corresponds to the induced insoluble fractions and contains the majority of the expressed proteins, which suggests that, although the HPPK proteins are efficiently

expressed in *E. coli*, they are predominantly insoluble. Although sufficient HPPK was present in the soluble fractions to warrant an attempt at purification, further optimisation for analytical scale expression of the HPPK proteins was performed in an attempt to improve the ratio of soluble to insoluble protein.

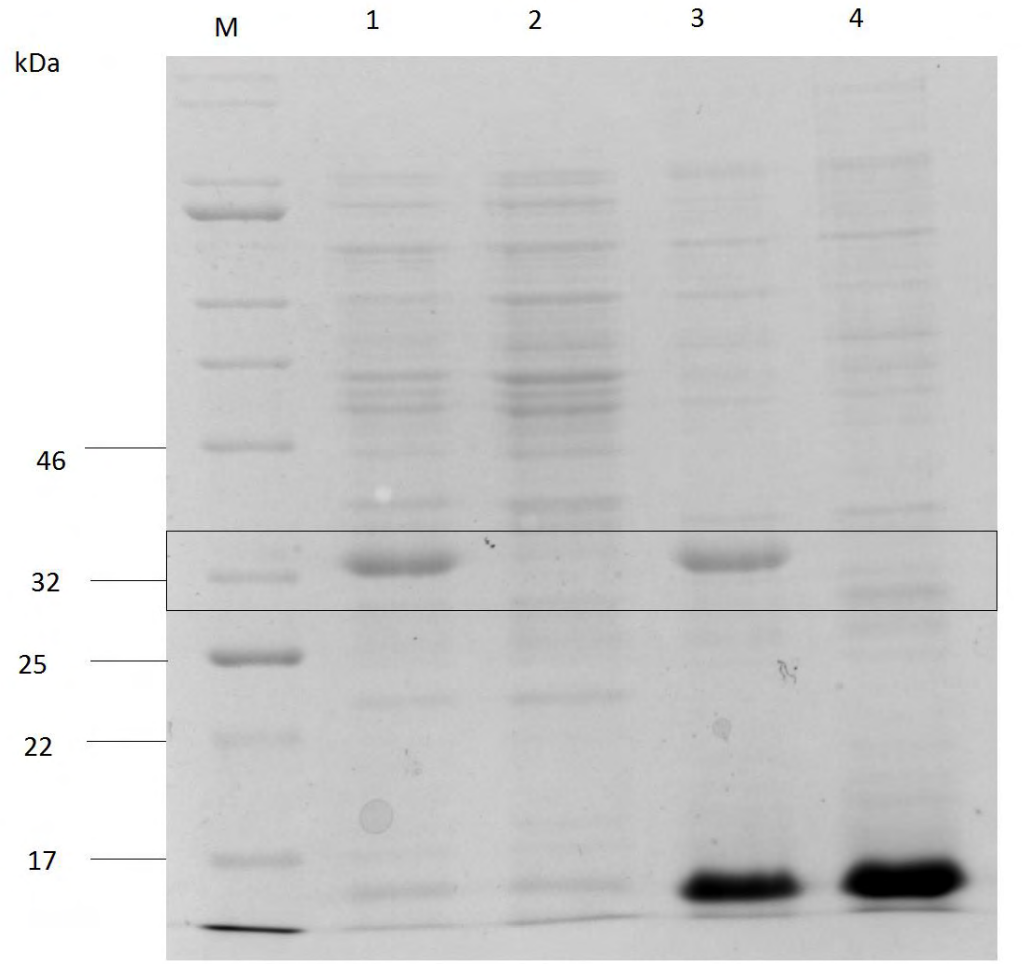


Figure 37: SDS-PAGE analysis of T. brucei PTR1 analytical scale expression conducted at 37 °C for 3 hours. Analytical scale expression samples were analysed by SDS-PAGE on a 12 % gel which was stained with Coomassie. M: Molecular mass marker; Lane 1: Induced soluble fraction; Lane 2: Uninduced soluble fraction; Lane 3: Induced insoluble fraction; Lane 4: Uninduced insoluble fraction.

From Figure 37 we note that the induced soluble fraction contained a protein of approximately 32 kDa that was absent in the uninduced fraction and corresponded to the predicted size of TPR1 (21 kDa) plus the His-tag. In contrast to both HPPK proteins, it was predominantly soluble, yet to further enhance the proportion of soluble protein, analytical scale expression was conducted at lower temperatures.

3.3.1.2 Analytical scale Protein Expression at 28 °C

Analytical scale expression was performed as described previously, with the difference that after the addition of 1 mM IPTG, cultures were incubated at 28 °C for 12 hours. After incubation, the experiment continued as described, whereafter SDS-PAGE analysis occurred (previously described). Results for *E. coli* HPPK and *S. enterica* HPPK are shown in Figure 38 and *T. brucei* PTR1 in Figure 39.

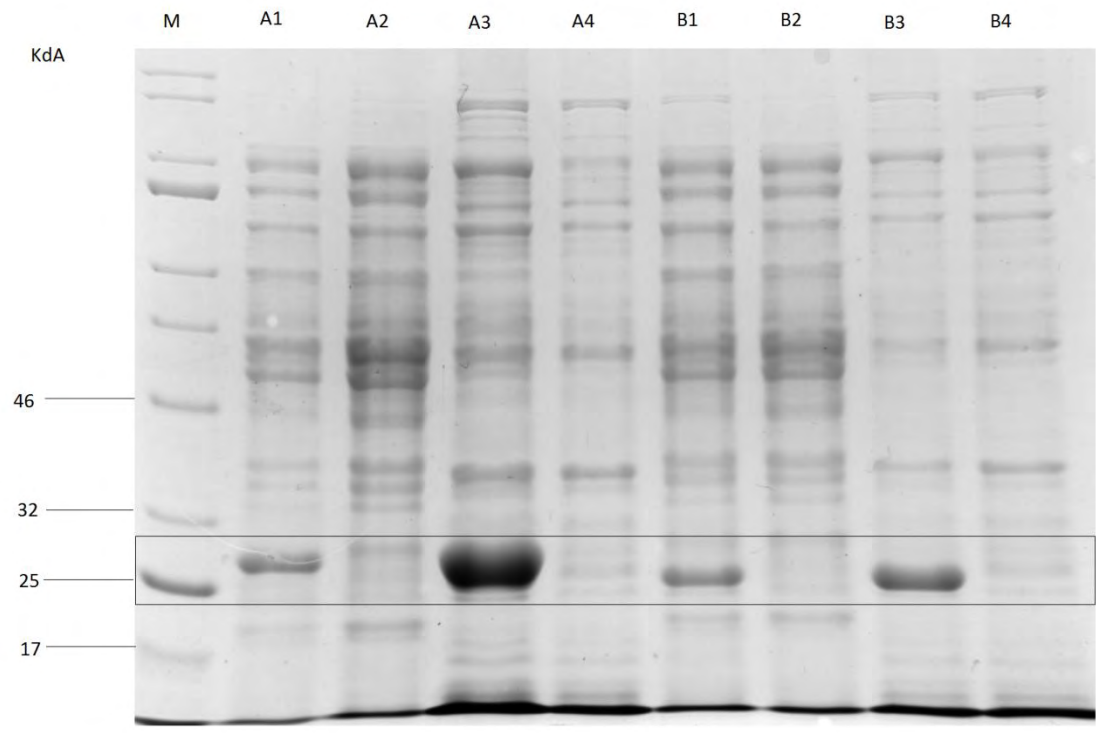


Figure 38: SDS-PAGE analysis of *E. coli* HPPK (A) and *S. enterica* HPPK (B) analytical scale expression conducted at 28 °C for 12 hours. Analytical scale expression samples were analysed by SDS-PAGE on a 12 % gel which was stained with Coomassie. M: Molecular mass marker; Lane 1: Induced soluble fraction; Lane 2: Uninduced soluble fraction; Lane 3: Induced insoluble fraction; Lane 4: Uninduced insoluble fraction.

From Figure 38, A and B, the results obtained from the decrease in incubation temperature to 28 °C and increase in incubation time to 12 hours were very similar to those obtained at 37 °C for 3 hours. Although unique protein bands corresponding to the predicted size of the HPPK proteins were present in the induced soluble fractions (lane1), the majority of the HPPK proteins remained insoluble as can be seen by the presence of a more intensely stain band present in the induced insoluble fractions (lane 3). In an attempt to improve the ratio of soluble to insoluble HPPK proteins, further analytical scale optimisation was performed.

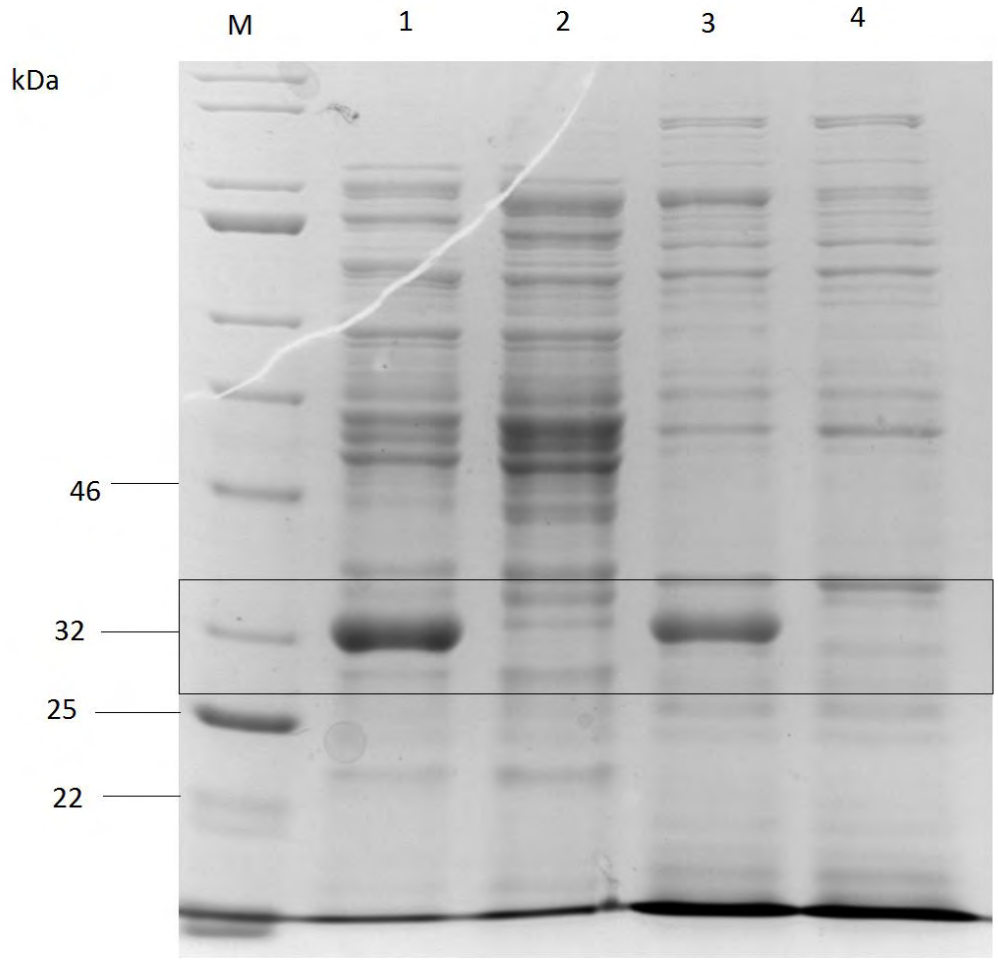


Figure 39: SDS-PAGE analysis of *T. brucei* PTR1 analytical scale expression conducted at 28 °C for 12 hours. Analytical scale expression samples were analysed by SDS-PAGE on a 12 % gel which was stained with Coomassie. M: Molecular mass marker; Lane 1: Induced soluble fraction; Lane 2: Uninduced soluble fraction; Lane 3: Induced insoluble fraction; Lane 4: Uninduced insoluble fraction.

From Figure 39 lane 1, the induced soluble fraction contained a unique intensely stained band indicating the presence of a protein at approximately 32 kDa. The same band was present in the induced insoluble fraction (lane 3) but to a lesser extent. The ratio of soluble to insoluble *T. brucei* PTR1 had increase sufficiently (indicated by the intense protein band at 32 kDa in lane 1 and a less intensely stained band in lane 3) and thus no further protein expression optimisation was attempted and these conditions were adopted for large scale expression and purification of *T. brucei* PTR1

3.3.1.3 Analytical scale Protein Expression at 12 °C

Further improvement of the solubility ratio of the recombinant HPPK proteins was performed as described previously. After the addition of 1 mM IPTG, cultures were incubated at 12 °C for 24 hours

and the experimental procedure described previously was followed. SDS-PAGE analysis was performed and the results for *E. coli* HPPK and *S. enterica* HPPK are shown in Figure 40.

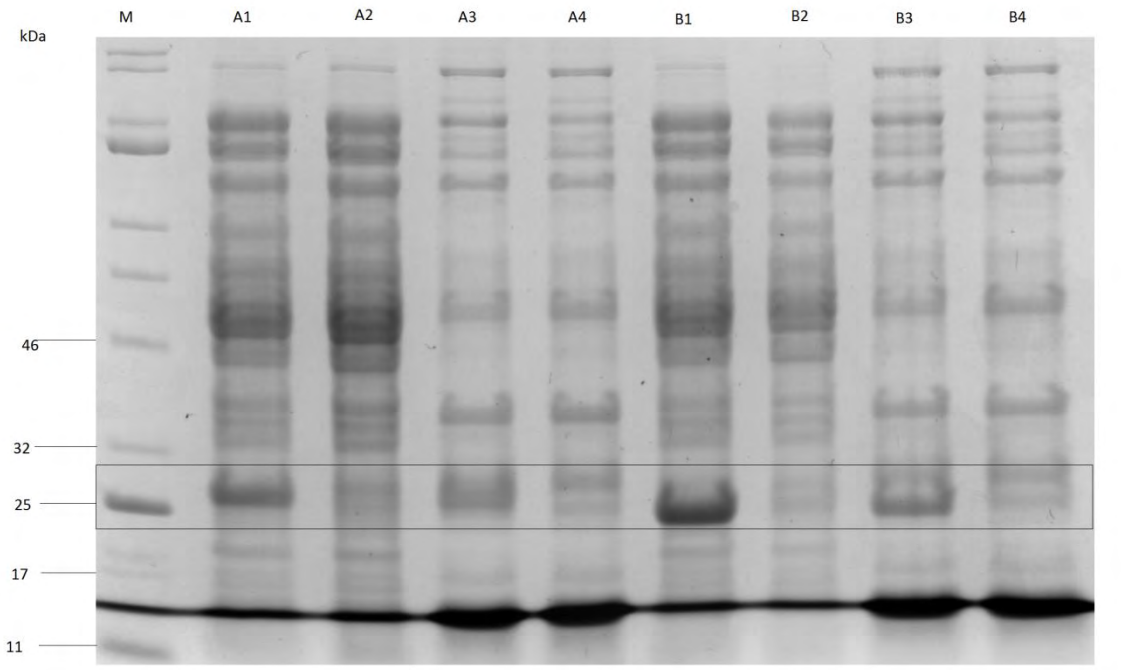


Figure 40: SDS-PAGE analysis of *E. coli* HPPK (A) and *S. enterica* HPPK (B) analytical scale expression conducted at 12 °C for 24 hours. Analytical scale expression samples were analysed by SDS-PAGE on a 12 % gel which was stained with Coomassie. M: Molecular mass marker; Lane 1: Induced soluble fraction; Lane 2: Uninduced soluble fraction; Lane 3: Induced insoluble fraction; Lane 4: Uninduced insoluble fraction.

From Figure 40, A and B, lane 1, an improved solubility of the HPPK proteins is indicated by the presence of a more intensely stained band with an approximate molecular weight of 25 kDa. In lanes 2 and 4, faintly stained bands can be seen occurring in the vicinity of 25 kDa, which is most likely low levels of uninduced HPPK expression. The ratio of soluble to insoluble HPPK had increased sufficiently compared to expression profiles obtained at higher temperatures and thus no further protein expression optimisation was performed for *E. coli* HPPK and *S.enterica* HPPK and these conditions were adopted for large scale expression and purification.

From the analytical scale expression experiments it can be deduced that, compared to standard induction conditions of 37 °C for 3-5 hours, for HPPK a lower temperature and longer incubation time (12 °C for 24 hours) will result in a larger yield of soluble protein, while for PTR1 intermediate conditions (28 °C for 12 hours) will result in a larger yield of soluble protein.

3.3.2 Protein Purification

Once optimisation of protein expression conditions was completed, large-scale expression and purification were performed to obtain high quantities of the proteins.

Soluble fractions obtained from IPTG-induced *E. coli* were applied to Ni-NTA columns and the eluates buffer-exchanged using desalting columns. The supernatants, Ni-NTA column flow-throughs, washes and eluates and desalting column eluates were analysed by SDS-PAGE. Results for *E. coli* HPPK and *S. enterica* HPPK are presented in Figure 41 and *T. brucei* PTR1 in Figure 42.

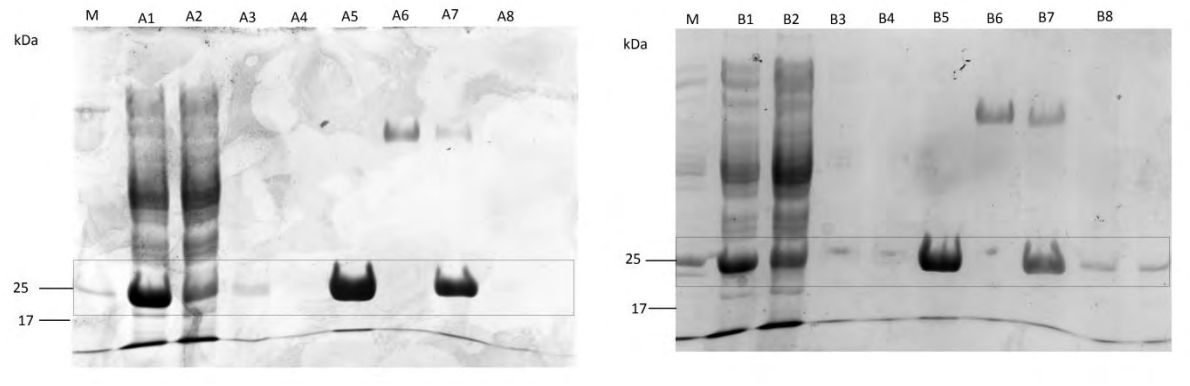


Figure 41: SDS-PAGE analysis of E. coli HPPK (A) and S. enterica HPPK (B) purified protein samples. *E. coli* HPPK and *S. enterica* HPPK was purified from a filtered lysate of *E. coli* cells induced with 1 mM IPTG at 12 °C for 24 hours and desalted by size exclusion chromatography. Samples were analysed using a 12 % SDS-PAGE gel stained with Coomassie. M: Molecular mass marker; Lane 1: Filtered Lysate; Lane 2: Column flow-through; Lane 3: First wash; Lane 4: Second wash; Lane 5: Elution; Lane 6: Ni-NTA column rinse; Lane 7: Desalted protein; Lane 8: Desalting column rinse.

In Figure 41, A and B, Lane 1 we can observe that the cells were completely lysed due to the presence of a high number of visible bands across the molecular weight spectrum, including a highly stained band in the vicinity of 25 kDa, which is the predicted size of the HPPK protein plus the His-tag. In the column flow-through (Lane 2), there appears a similar less intense band at the 25 kDa mark, either representing HPPK that failed to bind to the column or endogenous *E. coli* proteins migrating in the same position. The presence of a single highly stained band in Lane 5 indicates that the protein had been successfully purified in high quantity without contaminating *E. coli* proteins or evidence of extensive protein degradation. Lane 7 shows a loss of intensity of the staining which could indicate that some protein had been lost during the desalting process, but no proteins are visible in Lane 8, which is the desalting column wash. The presence of a protein in the Ni-NTA column rinse, shown by Lane 6 is somewhat anomalous, as no proteins in that area are visible in the elution or the desalting fraction, and could be due to the protein having such a large size that the initial run-through of the Ni-NTA column could not have removed it, but with a subsequent wash, it was removed.

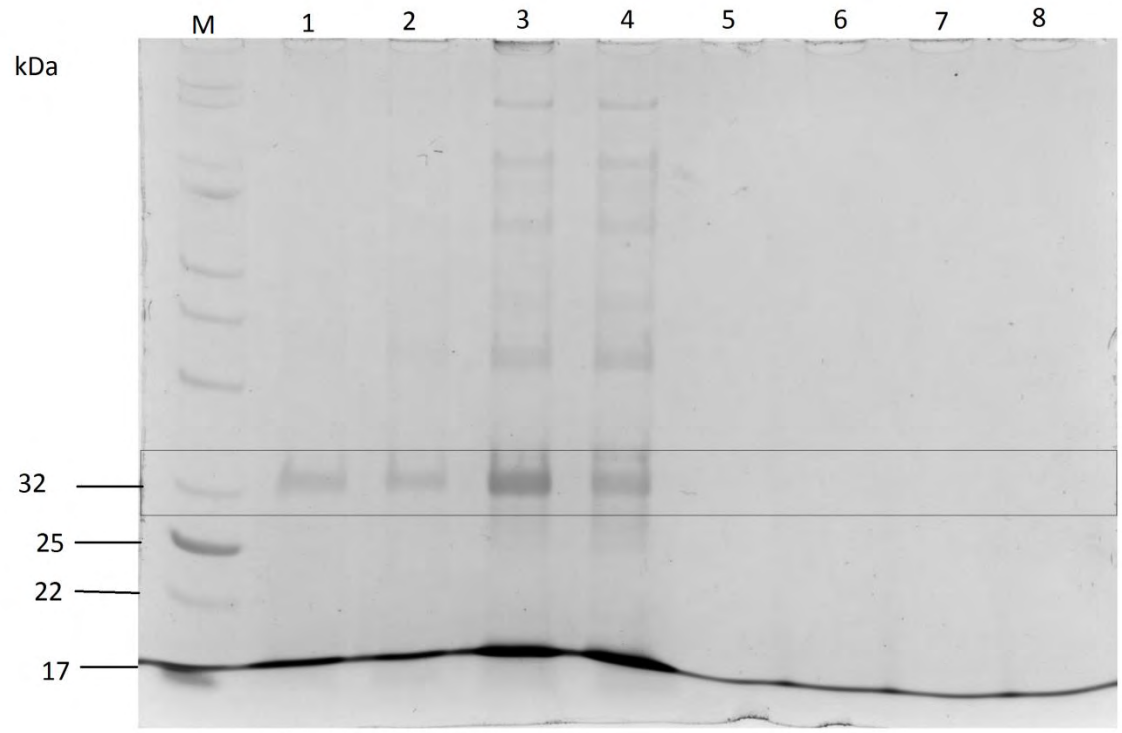


Figure 42: SDS-PAGE analysis of *T. brucei* PTR1 purified protein samples. *T. brucei* PTR1 was purified from a filtered lysate of *E. coli* cells induced with 1 mM IPTG at 28 °C for 12 hours and desalted by size exclusion chromatography. Samples were analysed using a 12 % SDS-PAGE gel stained with Coomassie. M: Molecular mass marker; Lane 1: Elution; Lane 2: Desalted protein; Lane 3: Filtered Lysate; Lane 4: Column flow-through; Lane 5: First wash; Lane 6: Second wash; Lane 7: Ni-NTA column rinse; Lane 8: Desalting column rinse.

From Figure 42, although samples were not accurately loaded onto the gel, purification of the 32 kDa *T. brucei* TPR1 protein plus His-tag yielded similar results to that obtained with the two HPPK proteins (Figure 41), except that the yield of eluted and desalted protein was less based on band thickness and staining intensity (lanes 1 and 2). Some evidence of unbound *T. brucei* TPR1 was also present in the column flow-through (lane 4).

3.3.3 Protein Concentration determination and storage

For proteins to be stored and later used in assays, the concentration of the desalted proteins needed to be determined.

Desalting had been carried out to ensure that any remaining imidazole which was used in the Ni-NTA affinity chromatography elution buffer was removed and also to exchange the elution buffer to TPSB. For protein concentrations, a Bradford assay was used and a standard curve was constructed using known concentrations of BSA. After the addition of Bradford reagent to the desalted protein samples and TPSB buffer alone (background control), absorbance was measured at 595 nm. Protein

concentrations are shown in Table 7 and the Bradford assay standard curve in Figure 43 and theoretical mg/L yield calculated. For long term storage, glycerol was added to the protein samples to a final concentration of 40 % (v/v) and the samples were stored at -20 °C.

Table 7: Purified Protein concentrations as calculated using Bradford assay.

Protein	Concentration (mg/ml)	Theoretical Yield (mg/L *E. coli* culture)
E. coli HPPK	0.496	5.95
S. enterica HPPK	0.543	6.52
T. brucei PTR1	0.216	2.59

From Table 7, the concentrations obtained from the purification shows a good yield of recombinant protein, sufficient for further experimental use. The theoretical yield shows that high scale expression would result in a large quantity of recombinant protein, in the event that a large quantity of protein is required.

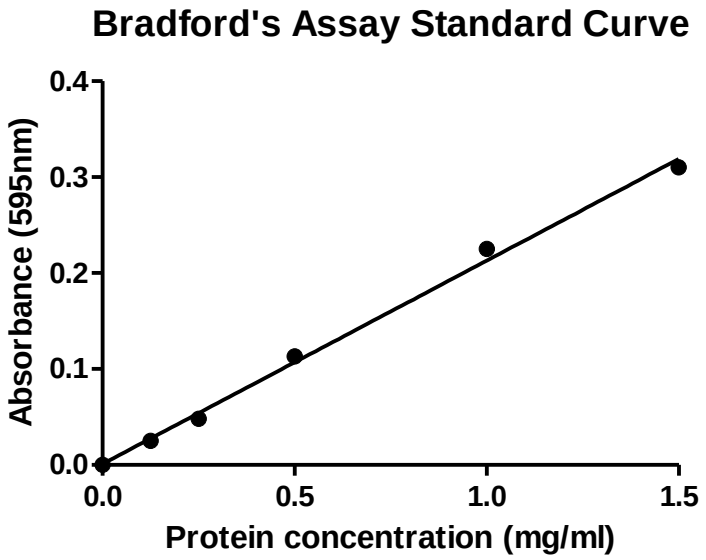


Figure 43: Bradford Assay Standard Curve. BSA protein concentrations of 0; 0.125; 0.25; 0.5; 1 and 1.5 mg/ml were used to perform Bradford assay and the absorbance of Bradford reagent was measured at 595 nm after 5 min of incubation.

3.4 Discussion

The expression vector used in this study contained a T7 promotor system, a *lacI* expression cassette and a lac operon for IPTG-inducible expression, and the host strain used was the T7 Express *lys*Y cells. The *E. coli* strain allows for very high levels of protein expression and stronger control overexpression. In the bound state T7 lysozyme and T7 RNA polymerase prevent attachment to the *lac* operon thus preventing protein expression from occurring in the absence of IPTG (Nallamsetty and Waugh, 2007). All proteins expressed and purified in this study had codons that were optimised for expression using *E. coli* cells and were synthesised by Genscript. The use of a bacterial expression system proved to be effective in expressing the proteins in a relatively short time and at a low cost, as stated by Terpe (2006). From analytical scale expression analysis using SDS-PAGE, the presence of the HPPK and PTR1 proteins in the transformed *E. coli* samples was confirmed. The expression was higher in induced fractions and at the expected protein size for each protein, thus indicating that the expression vector was successfully incorporated and functional within the bacterium. This suggests that optimisation of the codons for *E. coli*, using frequencies of occurrences of synonymous codons specific to the *E. coli* strain, proved to have been a favourable option. There was also no overt indication of cell death due to protein toxicity which would result in low levels of protein being expressed (Sørensen and Mortensen, 2005). Several expression systems, as well as promotor systems, exist, although bacterial systems are the most commonly used. *E. coli* is the favoured expression system to use for recombinant protein expression, as it is relatively inexpensive, provides fast results and has been well characterised and studied to allow for easy troubleshooting. This study found that the use of *E. coli* as an expression system proved to be fruitful, and trouble-shooting required in expression did not prove to be difficult (Pellizza *et al.*, 2018).

A problem that was identified in the analytical scale expression experiments was that all the proteins (*E. coli* HPPK, *S. enterica* HPPK and *T. brucei* PTR1) expressed at 37 °C for 3 hours were predominantly insoluble, as was seen from the SDS-PAGE analysis. The solubility of a protein will be influenced by factors such as the amino acid composition and sequence, molecular weight, and conformation. Furthermore, environmental factors such as buffer ionic strength, types of solvents, pH, temperature and processing conditions will also play a role in protein solubility. The misfolding of proteins, which is the main reason for insolubility, due to an increased production rate will lead to insoluble protein aggregates (Nallamsetty and Waugh, 2007). According to Ventura (2016) and Trimpin and Brizzard (2009) optimisation strategies that can be used to improve protein solubility include: 1) reducing the rate of protein synthesis, which can be done by lowering the growth temperature, using a weaker promoter (trc in the place of T7), using a lower copy number plasmid or lowering the inducer concentration. 2) changing the growth medium by adding prosthetic groups or

co-factors, adding buffers to control any pH fluctuations or the addition of glucose to repress the lac promoter, the addition of substances to increase the osmotic pressure. 3) co-expression with chaperones or foldases, which will promote proper folding and interact transiently with folding intermediates or by accelerating steps in folding. 4) periplasmic expression which will result in the protein being expressed into the periplasm which will allow for the proper formation of the disulphide bonds and reduced proteolysis. 5) altering the host strain to one with a more oxidising cytoplasmic environment. 6) addition of fusion partners which fuses the N-terminus to the C-terminus of a fusion partner to improve solubility. 7) expression of a fragment or functional domain of the protein rather than the full-length protein which will reduce the size of the protein, improving the ability of the organism to express the protein. Lowering the temperature of expression and increasing the expression time as a method of optimisation is both a valid, easy and inexpensive optimisation strategy, thus temperature and time of 28 °C for 12 hours was used to improve the solubility of the proteins, but this also did not provide an improved expression profile in the case of HPPK. The expression temperature and time were further decreased to 12 °C for 24 hours, which provided a more optimal expression of the HPPK recombinant proteins.

Since soluble *T. brucei* PTR1 predominated over insoluble PTR1 at 28 °C, thus no further optimisation was attempted and these conditions were adopted for large scale expression and purification. For *T. brucei* PTR1 Bello et al. (1994) used an alternative strain of BL21 *E. coli* cells (pLysS), with the same induction system and expression at 37 °C to produce PTR1. The pH at which Bello et al. (1994) performed harvesting and purification was lower at pH 7, while in this study pH 8 was used. Other differences between the method followed by Bello et al. (1994) are the use of 2-mercaptoethanol in the wash buffer, protein precipitation by ammonium sulphate and separation on a DEAE-cellulose column. The expression and purification methods used by others such as Cavazzuti et al. (2008), Dawson et al. (2006), Linciano et al. (2017), Nare, Hardy, and Beverley (1997), Shanks et al. (2010) and Sienkiewicz, Ong, and Fairlamb (2010) all derived their methods from that used by Bello et al. (1994)

The HPPK expression had challenges in the initial optimisation but final optimisation proved successful and yielded the protein of interest. There were minor bands present in the HPPK position in the uninduced samples, which are likely endogenous *E. coli* proteins migrating at the same molecular weight, but could also be an indication of a low level of uninduced HPPK expression. This could indicate minor basal leaking of the protein expression had occurred. The use of a strain derived from the BL21 cells should prevent any basal leaking from occurring. Alternatively, a pLys S or a pLys E or pBAD plasmid which has tight regulation could also prevent the leaking from occurring

(Han *et al.*, 2007). The methods used in this study are very similar to those used by Dennis et al. (2018) although on a larger scale than used in this study and at a higher temperature and shorter-expression time. Bing et al. (1999) performed a different method and at a much larger scale. The expression temperature was higher than that used by Dennis et al. (2018) and a concentration step by diafiltration was added before purification by ion-exchange chromatography. Bing et al. (1999) also precipitated the protein using ammonium sulfate. Other authors such as Su and Cukier (2009), Yang et al. (2005), Chhabra et al. (2013), and Yun et al. (2014) used various methods for the expression and these methods have some overlap with the method used here and those used by Dennis et al. (2018) and Bing et al. (1999).

The optimised expression conditions were used for large scale expression of the proteins and purification was performed via Ni-NTA chromatography. All recombinant proteins had a polyhistidine tag and the successful binding of the proteins onto the IMAC column was noted as the recombinant proteins were present on the gels in the lanes containing the eluate and desalted samples (Tate *et al.*, 2003). Thus purification of the proteins was successful. The concentration was determined using Bradford assay with the aid of a standard curve. This proved successful and no difficulties were encountered during this process.

In conclusion, using expression conditions determined during analytical scale expression experiments, all three the target proteins of this study were successfully prepared and purified soluble and intact, at high yields and free of *E. coli* protein contamination, suitable for use in follow-up enzyme activity experiments. Further confirmation of the identity of the proteins could be obtained by the use of western blotting with anti-His tag antibodies.

Chapter 4

PTR1 and HPPK enzyme activity assays

4.1 Introduction

The PTR1 enzyme has been studied for decades and is a validated drug target, with numerous studies having been performed on the enzyme. The general approach used for enzymatic testing is through an absorbance assay. PTR1 forms part of the dehydrogenase reductase family and is an NADPH-dependent enzyme that reduces biopterin to dihydrobiopterin (H_2B) and H_2B to tetrahydrobiopterin (H_4B) (see Figure 20, Chapter 1 page 29) (Dawson *et al.*, 2006). The reduction of biopterin and H_2B is accompanied by oxidation of NADPH to $NADP^+$. Thus monitoring the decrease in NADPH levels is a validated and tested method of monitoring the enzyme reaction as the oxidation product of NADPH does not absorb light at 340 nm. The absorbance assay measures light absorbance in the UV-Vis spectrum (200 nm - 800 nm). Molecular light absorbance is due to the presence of delocalised π-electrons, which are excited from their lower energy conformation or ground state to a higher energy state or excited state (Østergaard, 2016). As the light at a specific wavelength is passed through a sample, a portion of the light will be absorbed and the rest will be transmitted, indicated by a loss of intensity in the incident light. Absorbance assays can be used in a quantitative manner in which the concentration of a sample is determined by the transmission ratio, that is the loss of intensity, or in a qualitative manner to determine the wavelength at which peak absorbance of a compound will occur (Mäntele and Deniz, 2017).

As the enzyme reaction progresses the concentration of NADPH will decrease and thus the absorbance at 340 nm will decrease. Inhibitor studies should have the opposite effect in that they will prevent the enzyme from consuming NADPH and therefore a higher concentration of NADPH will be present with a higher absorbance reading (Chin *et al.*, 2009).

Resazurin (7-Hydroxy-3*H*-phenoxazin-3-one 10-oxide) is a phenoxazine dye that has very little fluorescence properties but is reduced to resorufin (7-Hydroxy-3*H*-phenoxazin-3-one) which is highly fluorescent (Kim and Jang, 2018). The colour conversion is from a blue/purple colour for resazurin to a pink colour for resorufin. The conversion is irreversible. Cell viability assays use the reduction of resazurin for the determination of aerobic respiration of metabolically active cells (Kumar *et al.*, 2018)(Präbst *et al.*, 2017). Within cells, NADPH dehydrogenase or NADH dehydrogenase will cause the reduction of resazurin using NADPH or NADH as the reductant, thus making resazurin an effective reagent for detecting the presence of live cells in samples, e.g. when performing assays to detect cytotoxic compounds. The reduction of resazurin to resorufin by nicotinamide nucleotides may also be exploited to detect the presence or concentration of NADPH/NADH in samples. It has

also been noted that resazurin can be used to determine mitochondrial activity. Due to the high level of sensitivity resazurin has been used in concentrations as low as 2.0 pmol, making it an inexpensive reagent to use for assays (O'Neill *et al.*, 2014). Two means of measuring resorufin levels are available, namely as a fluorescence assay or as an absorbance assay.

Fluorescence assays determine the fluorescence of a molecule based on its fluorescent properties. For a molecule to fluoresce it has to absorb light at a specific wavelength and emit it as a different wavelength and this wavelength difference is referred to as the Stokes shift (Østergaard, 2016). The wavelength that is absorbed is usually shorter than the wavelength that is emitted. When a photon hits a fluorophore (molecules possessing fluorescence properties) energy is transmitted to the electrons which will become excited. The energy is rapidly lost by the electrons by the emission of photons. The determination of fluorescence involves detection of the release of the photons (fluorescence emission) while the fluorophore-containing sample is illuminated with photons that excite the electrons (fluorescence excitation) (Taneja and Tyagi, 2007). Many assays use reagents that result in a fluorescent signal when a reaction occurs, e.g. the conversion of a non-fluorescent substrate to a fluorophore, as in the case of resorufin. Resorufin has strong fluorescent properties and this makes it an inexpensive means for fluorescence detection. Generally, for fluorescence detection of resofurin, the emission is set at 585-590 nm and the excitation is set at 530-570 nm, thus incorporating a range for optimal detection (Präbst *et al.*, 2017). As mentioned, the resazurin assay protocol can also be used as an absorbance assay, as the π-electrons are present in the molecule, thereby creating the delocalised electron effect desired in an absorbance assay. The absorbance is measured at 570 nm and 600 nm (O'Neill *et al.*, 2014).

HPPK forms part of the *de novo* folate biosynthetic pathway. It is responsible for catalysing the transfer of a pyrophosphate moiety from ATP to HMDP and produces AMP and HPPP as products (see Figure 12, Chapter 1 page 21). It has been established that the presence of 2 Mg^{2+} ions is required to act as a coordination ion which assists in creating the correct geometry of the HPPK enzyme for the one-step reaction of pyrophosphoryl transfer (Bermingham and Derrick, 2002). The enzyme has been studied in-depth to some degree, and while inhibitors of the enzyme have been identified, none have been successfully developed to reach clinical trials.

The HPPK enzyme uses ATP as a substrate and produces AMP as a product, thus to quantitatively measure enzyme activity, either ATP or AMP could be monitored (Blaszczyk *et al.*, 2004). The firefly (*Photinus pyralis*) luciferase enzyme is an oxidative enzyme responsible for the bioluminescence, consuming ATP and using it to convert luciferin to oxyluciferin and release photons. By adding it and the substrate luciferin to a reaction, it can thus be used to determine ATP levels after an enzyme

that consumes ATP, such as HPPK, has run its course (Cheng *et al.*, 2012). Luciferases are a large group of enzymes from many diverse protein families. They are all classified as oxidoreductases, indicating they act on a single donor incorporating molecular oxygen with their substrate (Iragues *et al.*, 2018). No light excitation is required for luciferase bioluminescence and therefore minimal autofluorescence will occur. This is an advantage over fluorescence assays, where background fluorescence readings can obscure or reduce the sensitivity with which particular fluorophores can be detected (Blaszczyk *et al.*, 2004).

The commercially available Kinase-Glo® Luminescent Kinase assay uses luciferase from *Photinus pyralis* (America firefly), which catalyses the reaction of ATP with D-luciferin and oxygen to produce oxyluciferin, AMP, carbon dioxide and pyrophosphate as products and releases a photon of light, which is detected by a luminometer (Promega corporation, 2015). The resulting light has an emission maximum of 562 nm. Due to luciferase acting on a single donor and no autofluorescence, the luciferase assay is one of the most sensitive assays, followed by fluorescence, with absorbance assays being the least sensitive. The sensitivity of the luciferase enzyme allows for potential problems with the assay which will cause false high readings. Any ATP derived from sources such as skin cells, excretions or ATP from microorganism activity will be used by the luciferase enzyme, thus making the need for stringent control over the workspace and methodical implementation crucial to minimise as many potential sources of contamination as possible (Iragues *et al.*, 2018)(Cheng *et al.*, 2012). In the case of the HPPK enzyme, an alternative means to detect enzyme activity may be to measure the AMP produced, as opposed to the ATP that is consumed. In assays such as the commercial AMP –Glo® assay, the enzymatic reaction is stopped, remaining ATP is removed and enzymes are added that convert AMP to ADP and ADP to ATP. Subsequently, the usual firefly luciferase assay can be used to detect the ATP, which should correlate with the original amount of AMP present (Promega Corporation, 2015).

As the need for advancement in the area of High-throughput screening arose to address the advances made in the large number of compounds which chemical libraries are composed of, so the need for advancement in quality and accuracy of the assays used increased (Iversen *et al.*, 2006). The design of assays for optimised speed, efficiency, signal detection and low reagent use along with a high degree of quality has led to the development of statistical tests tailored to consider quality (or signal robustness). The Z′-factor compares a negative control and positive control of an assay and allows the operator to determine which control to use within an experiment. Statistical measures used to calculate the Z′-factors of a particular assay are the means (μ) of the two controls and the standard deviation (σ) of the two controls (Zhang, Chung and Oldenburg, 1999)(Atmaramani,

Pancrazio and Black, 2020). A score of 1 would imply an ideal perfect assay where no further optimisation is required, while a score of less than 0 would indicate that the overlap between the two controls is too high to decisively discern between them. A score falling between 0 and 0.5 is indicative of a marginal assay with a greater need for optimisation and a score between 0.5 and 1 indicates an excellent assay in which optimisation can occur but the closer the value to 1 the less the need for optimisation becomes (Zhang, Chung and Oldenburg, 1999).

The experiments in this Chapter aimed to:

i) Establish and optimise a *T. brucei* PTR1 NADPH absorbance assay to confirm enzyme activity and assess its utility as a screening tool.

ii) Screen compounds that were previously shown to interact with the enzyme in *in silico* screening and cause parasite death in living *T. brucei* parasites for Kimuda *et al.*, (2019) inhibitory activity against the PTR1 enzyme.

iii) Determine the IC_{50} of any compound which showed a 50 % or greater enzyme inhibition at 50 µM inhibitor concentration.

iv) Establish and demonstrate that resazurin can be used as an alternative to 340 nm absorbance readings to detect NADPH levels and PTR1 enzyme activity for compound screening.

v) Establish and optimise an HPPK luminescence assay to confirm enzyme activity and assess its utility as a screening tool.

vi) Screen compounds determined to interact with the enzymes *in silico* as described in Chapter 2 of this dissertation.

vii) Determine the IC_{50} of any compound which showed a 50 % or greater enzyme inhibition at 50 µM inhibitor concentration.

viii) Determine if any compounds act as false substrates for the enzymes.

4.2 Methods and Materials

4.2.1 *T. brucei* PTR1 activity NADPH absorbance assay

The method used in this study was based on the work of Bello *et al.* (1994) with alterations. The purified His-tagged *T. brucei* PTR1 was diluted to a final concentration of 20 µM in TPSB (25 mM Tris-HCl, 15 mM NaCl, 0.1 mM $MgCl_2$, 0.03 mM BSA, 0.2 mM 2-mercapto-ethanol). A 1 mM NADPH stock solution was prepared in 20 mM Tris-HCl for assay use and stored at -20 °C. A 1 mM 6-biopterin stock solution was prepared in dimethyl sulfoxide (DSMO) for assay use and stored at -20 °C. In a microcentrifuge tube, a reaction mix was prepared to provide final assay concentrations of 50 µM NADPH, 50 µM 6-biopterin, and 0.8 µM *T. brucei* PTR1 in a final volume of 500 µl 20 mM Tris-HCl. The reaction mix was allowed to incubate at room temperature for 30 min before the solution was distributed into a clear flat bottomed 96 well assay plate (NEST, China). The absorbance was measured immediately after distribution into plates at 340 nm using a SpectraMax® M3 plate reader.

4.2.2 *T. brucei* PTR1 absorbance kinetic assay

The method used was based on the work of Teixeira *et al.* (2019)From the stock solutions, a reaction mix was prepared to contain 50 µM NADPH, 50 µM 6-biopterin and 0.1 µM *T. brucei* PTR1 in a 500 µl volume 20 mM Tris-HCl. The reaction mix was distributed into a clear flat bottomed 96 well assay plate and placed in the SpectraMax® M3 plate reader set to measure absorbance at 340 nm over a 60 min period, at 5 min intervals. The reaction temperature was set to 25 °C.

4.2.3 *T. brucei* PTR1 absorbance screening assay

A screening assay was prepared by using the stock solutions, previously described by Bello *et al.* (1994), in a reaction mix containing 50 µM NADPH, 50 µM 6-biopterin and 0.1 µM *T. brucei* PTR1 in a 500 µl volume of 20 mM Tris-HCl. Compounds used for inhibitor studies were determined using *in silico* docking studies performed by Kimuda, Laming, Hoppe, & Bishop, (2019) and were purchased from Molport ®. 10 mM stock solutions of inhibitors were prepared in DMSO and added to the reaction mix to a final assay concentration of 50 µM. The reaction was allowed to incubate at 25 °C for 40 min before being distributed into a clear flat bottomed 96 well assay plate and the absorbance was read at 340 nm using a SpectraMax® M3 plate reader.

A kinetic assay was performed with inhibitor 1 and inhibitor 5 using the concentrations described above. The reaction mix was distributed into clear flat bottomed 96 well assay plates and placed in the SpectraMax® M3 plate reader, reading absorbance at 340 nm over 40 min at 5 min intervals. The reaction temperature was set at 25 °C.

4.2.4 *T. brucei* PTR1 Resazurin assay

A 500 µM Resazurin assay stock solution was prepared in phosphate-buffered saline (PBS) for further dilution into assay reaction mix. In a microcentrifuge tube, a PTR1 reaction mix was prepared by the addition of 50 µM NADPH, 50 µM 6-biopterin and 0.1 µM *T. brucei* PTR1 and allowed to incubate at room temperature for 40 min. After incubation, 5 µl of the resazurin assay stock was added creating a final concentration of 5 µM and the reaction was allowed to continue for 30 min at room temperature. The reaction mix solution was distributed into clear flat bottom 96 well assay plates and fluorescence was measure with excitation at 560 nm and emission at 590 nm on a SpectraMax® M3 plate reader.

4.2.5 HPPK Luminescence assay

HPPK activity was determined by measuring unprocessed ATP using the Promega Kinase-Glo® Luminescent assay kit (Promega Corporation, 2015). Assay stock solutions were diluted to a concentration of 10 µM ATP, 10 µM HMDP and 10 µM HPPK enzyme (prepared as described in Chapter 3) using reaction buffer (200 mM Tris-HCl, 20 mM βME, 0.02 % BSA, 20 mM $MgCl_2$, 0.02 % Tween 20, 10 % DMSO) and stored at 4 °C. From assay stock solutions, a reaction mix was made containing a final assay concentration of 0.1 µM HPPK enzyme, 1.4 µM HMDP and 1.5 µM ATP. The reaction was allowed to incubate at 30 °C for 30 min, after which 200 µl Kinase-Glo® reagent was added and the reaction proceeded for a further 10 min at room temperature before being distributed into a white flat bottomed 96 well assay plates. Luminescence was read using a SpectraMax® M3 plate reader set to read as an endpoint assay at all wavelengths with an integration time of 1 second. Inhibitors were determined in Chapter 2 of the dissertation using bioinformatics *in silico* screening and were purchased from Molport®. For inhibitor studies, an assay stock solution of 10 mM was prepared in DMSO. Inhibitors were added to the reaction mix described above to achieve a final assay concentration of 50 µM. An IC_{50} was performed using a 2-fold serial dilution series with starting assay concentration of 100 µM using the above-described assay procedure.

4.2.6 Z'-factor statistical analysis

Z'-factors were calculated to use as an index for the quality of the assay as established by Zhang, Chung, and Oldenburg (1999), using the following formula:

$$Z' = 1 - \frac{3(\sigma p + \sigma n)}{\mu p - \mu n}$$

σp –standard deviation of the positive control; σn – standard deviation of the negative control; µp – mean of the positive control; µn – mean of the negative control.

Positive and negative controls for each experiment are indicated in the results section.

4.2.7 Compounds used in this study

All compounds were purchased as 2 mg stocks from MolPort (Riga, Latvia). These included the predicted PTR1 inhibitors (MolPort ID 019-765-188, 001-638-900, 001-740-726, 002-011-639 and 019-741-839, designated as RUBi004, 007, 014, 016 and 018 in this study) and HPPK inhibitors (MolPort ID 000-629-405, 002-825-627, 001-994-804, 002-186-309, 019-762-754, 001-603-649, 000-850-220, 000-489-967 and 000-482-306 for ZINC 4671177, 4498423, 0310450, 4499339, 6792518, 0923924, 4476280, 0082136 and 0634089 respectively). Compounds were dissolved in DMSO to 10 mM and stored as aliquots at -20°C.

4.3 Results

4.3.1 *T. brucei* PTR1 Absorbance Assay

4.3.1.1 Enzyme Activity Assay

A preliminary assay was performed to determine if the purified enzyme was active before any optimisation. A reaction mixture was prepared to contain 0.8 µM *T. brucei* PTR1 enzyme, 50 µM NADPH, 50 µM 6-biopterin in 20 mM Tris-HCl. The reaction was allowed to incubate at room temperature for 30 min before the absorbance was measured at 340 nm. A negative (NADPH) control was set up which contained all the reaction components except the PTR1 enzyme, as well as a background control lacking NADPH. Results are shown in Figure 44.

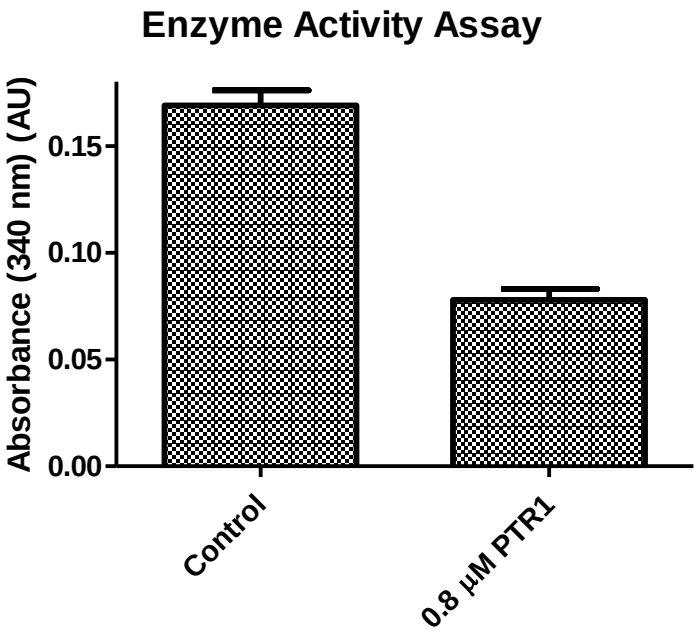


Figure 44: *T. brucei* PTR1 enzyme activity assay. The reaction contained 50 µM NADPH, 50 µM 6-biopterin and 0.8 µM *T. brucei* PTR1 and after 30 min of incubation at room temperature, the absorbance was read at 340 nm. Negative control lacked enzyme and 6-biopterin. Results are depicted with the background (Tris-HCl lacking the enzyme, substrate and NADPH) values subtracted and are shown as the mean of technical triplicate readings ± standard deviation.

The results in Figure 44 show that the *T. brucei* PTR1 enzyme was active and that the enzyme reaction showed a decrease in absorbance compared to that of the NADPH control. After subtracting background absorbance reading, Abs_{340} decreased by 53.95 % when comparing the enzyme reaction to the negative control. Percentage change was calculated by subtracting the enzyme reaction from the NADPH control, dividing this difference by the NADHP control and multiplied by 100. An

additional control that could be carried out is a reaction containing enzyme without biopterin, to account for substrate-independent oxidation of NADPH by the enzyme.

To determine the suitability of an enzyme assay for compound screening (inhibitor identification), a statistical value that is often used is the Z'-factor (Zhang, Chung and Oldenburg, 1999).

For Z´-factor calculations, the NADPH standard was used as the positive control as no enzyme activity is measured and the enzyme reaction was used as the negative control (after subtracting the background values) and produced a value of 0.61 which is above the 0.5 value that is regarded as the minimum for a highly suitable assay.

4.3.1.2 NADPH Optimisation

In an attempt to further optimise the *T. brucei* PTR1 absorbance assay the initial step was taken to optimise the NADPH concentrations used in the assay. The assay was prepared using 0.8 µM *T. brucei* PTR1 enzyme, 50 µM 6-biopterin, with differing concentrations of NADPH in 20 mM Tris-HCl and allowed to incubate at room temperature for 30 min before the absorbance was read at 340 nm. For each concentration of NADPH used a negative control lacking enzyme was included, as well as one background measurement of a reaction mixture lacking NADPH. Results are shown in Figure 45.

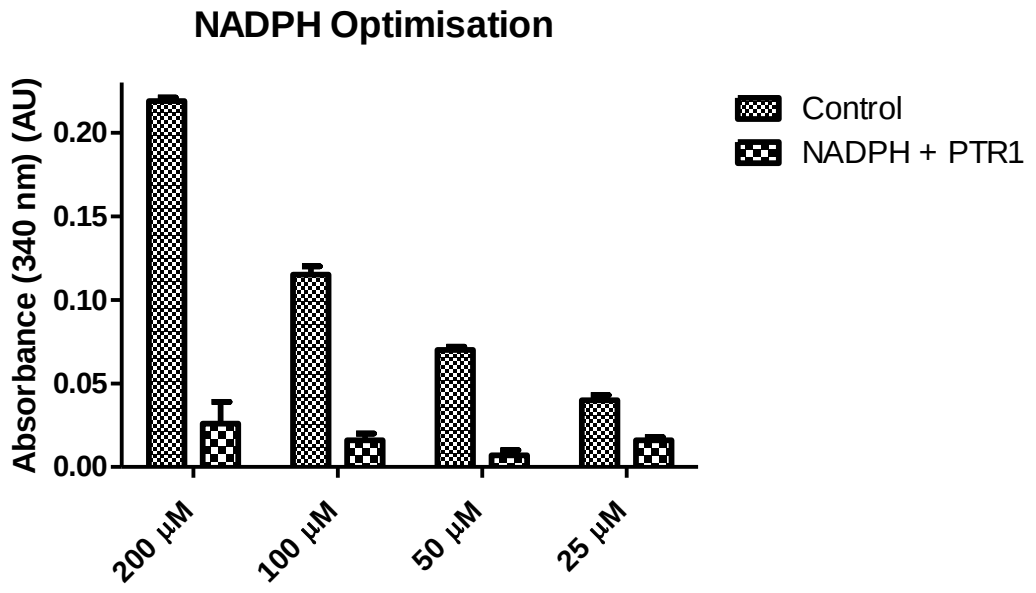


Figure 45: *T. brucei* PTR1 absorbance assay with differing concentrations of NADPH. The concentration of NADPH used during the assay followed a serial dilution from 200 µM to 25 µM and included the initially used 50 µM concentration. The reaction contained 50 µM 6-biopterin and 0.8 µM *T. brucei* PTR1, incubated at room temperature for 30 min. Results are depicted with background absorbance values subtracted and are shown as the mean of technical triplicate readings ± standard deviation.

From Figure 45, it is noted that absorbance values decrease with decreasing NADPH concentrations and that a significant change in absorbance (ΔAbs_{340}) is noted occurring between the control (NADPH) and the enzyme reaction (NADPH + PTR1). At 50 µM NADPH concentration, the enzyme reaction consumes the majority of the NADPH. The assay was statistically analysed and is discussed below

Table 8: *T. brucei* PTR1 absorbance assay with differing NADPH concentration showing Δ Abs_{340} and Z´-factors. For the determination of the Z'-factor, calculations used NADPH control as a positive control and the full enzyme reaction as a negative control.

NADPH conc.	Δ Abs_{340}	Z´-factor
200 µM	0.193 (88.13 %)	0.76
100 µM	0.099 (86.38 %)	0.75
50 µM	0.062 (89.47 %)	0.76
25 µM	0.025 (61.16 %)	0.38

From Table 8, the biggest Δ Abs_{340} occurs in the presence of 50 µM NADPH (89.47 %) with the most statistical significance occurring at 50 µM when comparing the individual NADPH controls to the reactions. The Z'-factor calculated at 200 µM was equal to that calculated at 50 µM, which was an improvement on all other Z'-factors. The ΔAbs_{340} calculated at each NADPH concentration proved to have a lower percentage change than that at 50 µM NADPH concentration. The above statistical analysis indicates that the optimal concentration to use is 50 µM, which was originally used in the assay. This is slightly lower than that used by Bello *et al.* (1994), Shanks *et al.* (2010) and Cavazzuti *et al.* (2008) who used 100 µM. Kazemi *et al.* (2010) used 40 µM NADPH.

4.3.1.3 Enzyme concentration optimisation

The NADPH concentration used in the assay was set at 50 µM and to further optimise the assay it was decided to determine whether lower enzyme concentrations would still produce robust enzyme activity results in the assay. The assay was run in a reaction mix using 50 µM NADPH, 50 µM 6-biopterin and a serial dilution of *T. brucei* PTR1 enzyme 0.8 µM to 0.025 µM. The reaction was allowed to incubate at room temperature for 30 min before the absorbance was measured at 340 nm. An NADPH control (reaction lacking enzyme) as well as a background control lacking NADPH was included. Results are shown in Figure 46.

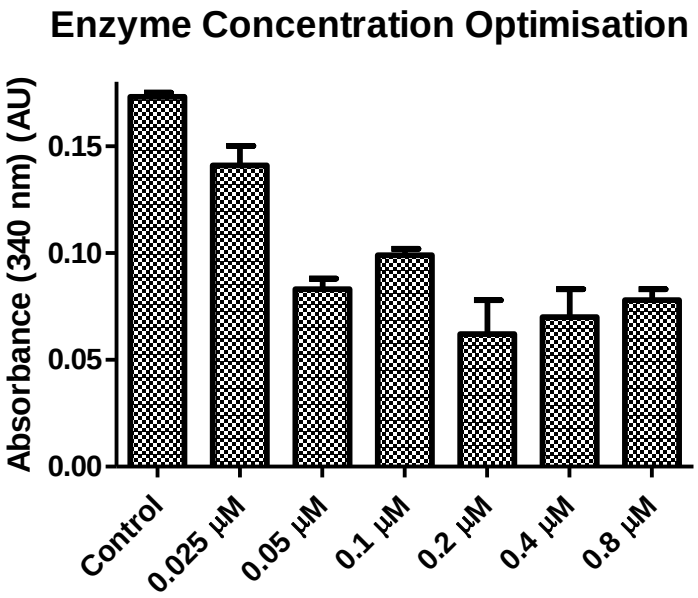


Figure 46: *T. brucei* PTR1 absorbance assay enzyme optimisation. The reaction was performed using 50 µM 6-biopterin, 50 µM NADPH and a serial dilution series from 0.8 µM to 0.025 µM *T. brucei* PTR1 enzyme. Incubation occurred at room temperature for 30 min. A negative NADPH (50 µM) was used. Results are depicted with background absorbance values subtracted and are shown as the mean of technical triplicate readings ± standard deviation.

From Figure 46, the smallest ΔAbs_{340} occurs at the lowest enzyme concentration, with the largest change occurring at 0.2 µM - 0.8 µM. Although the mean Abs_{340} change appeared to decrease from 0.2 to 0.4 and 0.8 uM, there was no significant difference between the values ($p>0.05$). The ΔAbs_{340} at 0.1 µM is significant, although the higher concentrations and 0.05 µM are higher. Statistical analysis was performed to investigate the percentage change and Z'-factors for each enzyme concentration.

Table 9: *T. brucei* PTR1 absorbance assay with differing enzymes concentration showing Δ Abs_{340} and Z´-factors. For the determination of Z'-factors, calculations used NADPH control as a positive control and the full enzyme reaction as a negative control.

T. brucei PTR1 conc.	Δ Abs_{340}	Z´-factor
0.025 µM	0.032 (18.50 %)	0.05
0.05 µM	0.090 (52.02 %)	0.78
0.1 µM	0.074 (42.97 %)	0.84

0.2 µM	0.111 (64.16 %)	0.53
0.4 µM	0.103 (59.54 %)	0.57
0.8 µM	0.095 (55.12 %)	0.81

From Table 9, the most significant ΔAbs_{340} occurred at 0.2 µM concentration and the lowest change at 0.025 µM, indicating that 0.2 µM would be the ideal concentration to use. From the Z'-factors, 0.2 µM has a Z'-factor of 0.53 while 0.1 µM has an improved Z'-factor of 0.84 which is the highest Z'-factor of the different concentrations indicating that 0.1 µM would result in an improvement of assay conditions. Although the ΔAbs_{340} is lower (42.97 %) than compared to the initial enzyme assay (53.95% and that obtained during NADPH optimisation (89.47 %), the Z'-factor (0.84) has improved from the initial enzyme reaction (0.61) and the NADPH optimisation (0.76), thus indicating 0.1 µM to be an improve enzyme concentration to use.

4.3.1.4 *T. brucei* PTR1 Absorbance assay Kinetic analysis

To ensure that the assay in the optimised format was of good quality a final step was taken to evaluate the enzyme efficiency over time and optimise the reaction time. A reaction mix was prepared using 0.1 µM *T. brucei* PTR1, 50 µM NADPH and 50 µM 6-biopterin in 20 mM Tris-HCl. The reaction was allowed to progress at room temperature for 60 min, in the Spectramax M3 plate reader and the absorbance was measure at 340 nm at five min intervals. An NADPH control (reaction lacking enzyme) as well as a background control lacking NADPH was included. Results are shown in Figure 47.

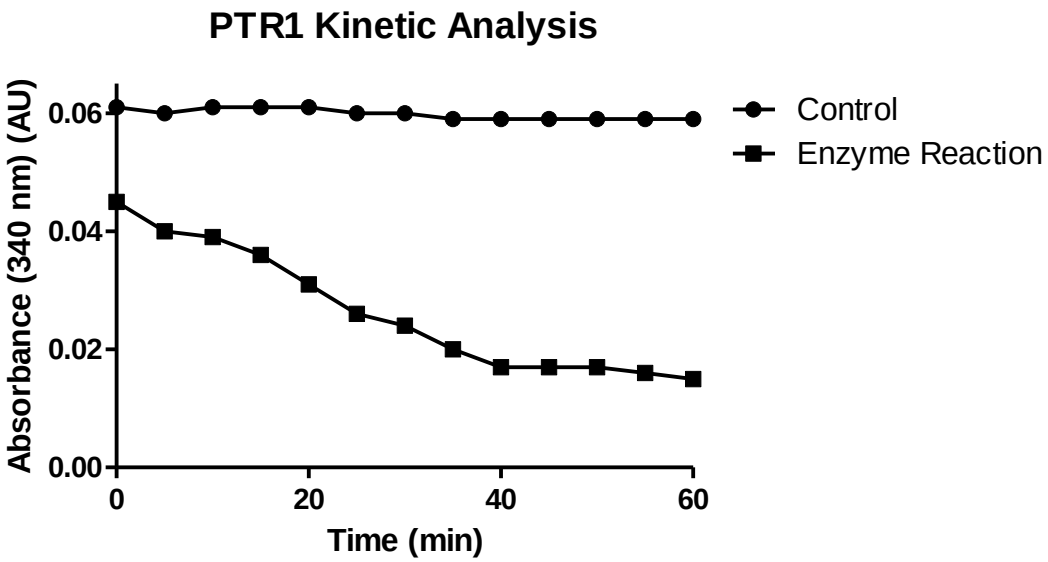


Figure 47: Evaluation of time on the *T. brucei* PTR1 absorbance assay. The reaction was performed using 50 µM 6-biopterin, 50 µM NADPH and 0.1 µM *T. brucei* PTR1 enzyme. The reaction was monitored at Abs_{340} in a Spectramax M3 plate reader over 60 min, with reading intervals of 5 min. A negative NADPH (50 µM) control was used, as well as a background control. Results are depicted with background absorbance values subtracted and are shown as the mean of technical triplicate readings ± standard deviation (standard deviations are very low, hence not visible in the graph).

Figure 47, shows the NADPH control and all other data points well below it. A problem that was noted was that the absorbance reading for the enzyme reaction does not start at the same point as that of the NADPH control, which could be ascribed to NADPH being consumed in the reaction while the plate was prepared and transferred to the reader (a plate reader with reagent injectors to avoid this problem was not available). The enzyme reaction appears to have run to completion after 40 min, with negligible changes occurring from 40 min to 60 min.

Table 10: Z´-factor score and ΔAbs_{340} at different time points at a fixed *T. brucei* PTR1 concentration (0.1 µM) after 60 min. For the determination of the Z'-factor, calculations used NADPH control as a positive control and the full enzyme reaction as a negative control.

Time (min)	ΔAbs_{340}	Z´-Factors
0	0.016 (26.35 %)	0.74
5	0.020 (33.41 %)	0.82
10	0.023 (36.96 %)	0.90
15	0.025 (41.57 %)	0.76

20	0.030 (49.11 %)	0.74
25	0.034 (56.14 %)	0.78
30	0.036 (59.83 %)	0.83
35	0.039 (66.82 %)	0.89
40	0.042 (71.64 %)	0.94
45	0.042 (71.70 %)	0.88
50	0.042 (71.11 %)	0.83
55	0.043 (72.27 %)	0.86
60	0.044 (75.13 %)	0.88

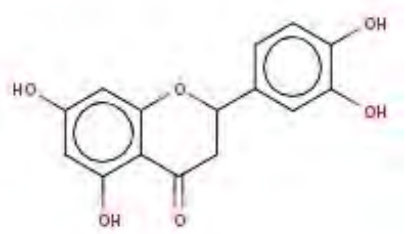
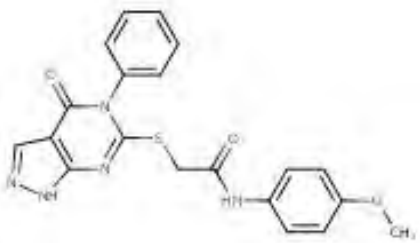
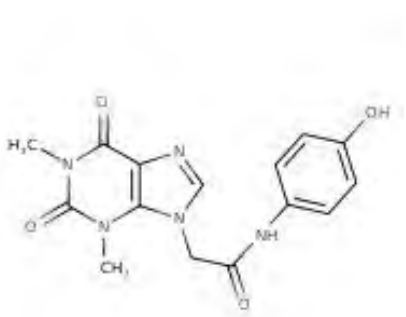
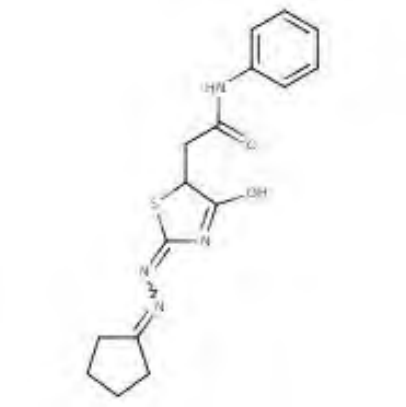
Table 10, shows the different Z'-factors and ΔAbs_{340} occurring throughout the experiment. The optimal Z'-factor was noted to occur at 40 min while showing an improved and high change in absorbance. As mentioned the ΔAbs_{340} occurring at 45 min – 60 min is not very different from that measure at 40 min (71.64 %) and the Z'-factors are also lower than those occurring at 40 min (0.94). This indicates that the optimal reaction time is 40 min.

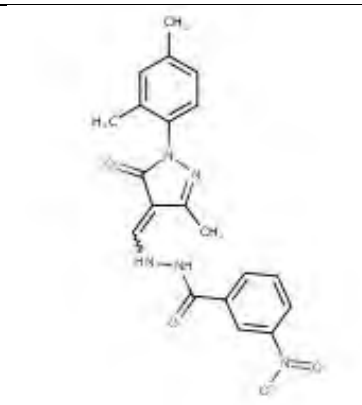
4.3.1.5 *T. brucei* PTR1 absorbance inhibitor screening assay.

The purpose of this study was to screen compounds identified by bioinformatic *in silico* screening performed by Kimuda et al., (2019). The optimal reaction conditions were determined from optimisation experiments yielding improved Z'-factor results and used for inhibitor screening. In a reaction mix 50 µM NADPH, 50 µM 6-biopterin and 0.1 µM *T. brucei* PTR1. The reaction was allowed to incubate at room temperature for 40 min before the absorbance was measured at 340 nm. An NADPH control (reaction lacking enzyme) as well as a background control lacking NADPH was included. Results are shown in Figure 48.

Table 11: Inhibitors of *T. brucei* PTR1 as determined by Kimuda et al., (2019), structures and names obtained from Molport. The table represents hit compounds from the study, using an alternative screening library to that used in Chapter 2. RUBi numbers are used as they are used in the paper

Inhibitor	RUBi number	Structure	IUPAC name

Inhibitor 1	RUBi014		2-(3,4-dihydroxyphenyl)-5,7-dihydroxy-3,4-dihydro-2H-1-benzopyran-4-one
Inhibitor 2	RUBi007		N-(4-methoxyphenyl)-2-({4-oxo-5-phenyl-1H,4H,5H-pyrazolo[3,4-d]pyrimidin-6-yl}sulfanyl)acetamide
Inhibitor 3	RUBi016		2-(1,3-dimethyl-2,6-dioxo-2,3,6,9-tetrahydro-1H-purin-9-yl)-N-(4-hydroxyphenyl)acetamide
Inhibitor 4	RUBi018		2-[2-(2-cyclopentylidenehydrazin-1-ylidene)-4-hydroxy-2,5-dihydro-1,3-thiazol-5-yl]-N-phenylacetamide

Inhibitor 5	RUBi004		N'-{[1-(2,4-dimethylphenyl)-3-methyl-5-oxo-4,5-dihydro-1H-pyrazol-4-ylidene]methyl}-3-nitrobenzohydrazide

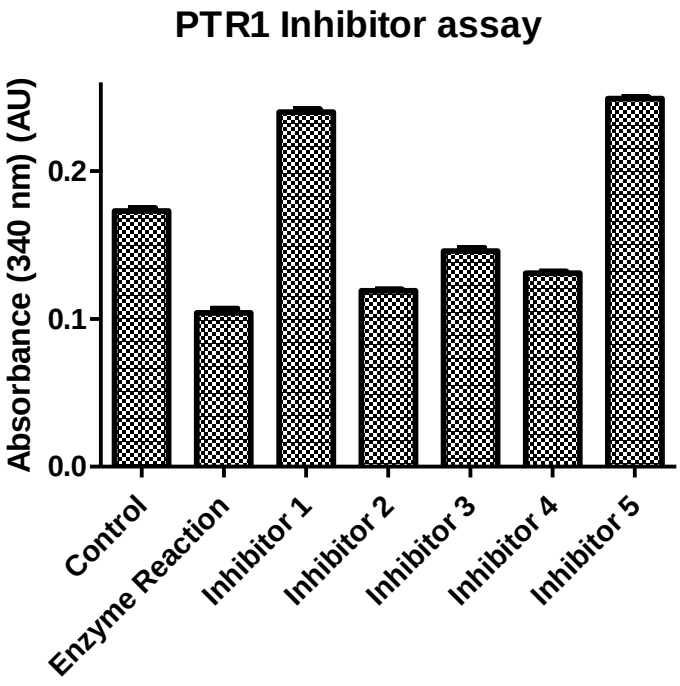


Figure 48: *T. brucei* PTR1 absorbance assay in the presence of inhibitors. The reaction was performed using 50 µM 6-biopterin, 50 µM NADPH and 0.1 µM *T. brucei* PTR1 enzyme with the addition of 50 µM inhibitor. The reaction was allowed to incubate at room temperature for 40 min and the absorbance was read at 340 nm. A negative NADPH (50 µM) control was used, as well as a background control. Results are depicted with background absorbance values subtracted and are shown as the mean of technical triplicate readings ± standard deviation.

From Figure 48, the presence of inhibitors does increase the absorbance values, indicating that the presence of the inhibitors decreases enzyme activity. Inhibitor 1 and inhibitor 5 has absorbance readings higher than the control which indicates that the compounds themselves possess absorbance reading at 340 nm, thus determining their inhibitory effects proves difficult.

Table 12: ΔAbs_{340} value of inhibitors compared to the enzyme reaction. ΔAbs_{340} was calculated by subtraction reaction values from NADPH control.

Individual Assay components	ΔAbs_{340}
Enzyme Reaction	0.069 (40.08 %)
Inhibitor 1	Results in a negative change
Inhibitor 2	0.054 (31.02 %)
Inhibitor 3	0.027 (15.61 %)
Inhibitor 4	0.042 (24.47 %)
Inhibitor 5	Results in a negative change

From Table 12, the ΔAbs_{340} of the enzyme reaction was 40.08 %, which was decreased from the kinetic assay, but not very different from the enzyme optimisation experiment (42.97 %) which could have resulted from changes in ambient conditions within the experimental environment. The inhibitors showed a decrease in the ΔAbs_{340} when compared to the enzyme reaction. The smallest ΔAbs_{340} is noted with inhibitor 3, which shows a ΔAbs_{340} of 15.61 %, slightly more than a 50 % enzyme inhibition. Inhibitor 4 shows a 24.47 % ΔAbs_{340} (less than 50 % enzyme inhibition) and Inhibitor 2 shows a 31.02 % change (less than 50 % enzyme inhibition). These results indicate that inhibitor 3 provides the greatest enzyme inhibition followed by Inhibitor 4 and Inhibitor 2 shows the lowest enzyme inhibition. The inhibitory activities of the compounds may be underestimated. Additional background controls should include reactions carried out in the presence of compound but lacking enzyme, to account for possible contributions made by the compounds to the absorbance at 340 nm.

4.3.1.6 *T. brucei* PTR1 Inhibitor 1 and 5 Kinetic Analysis

In an attempt to investigate the inhibitory effects of Inhibitor 1 and Inhibitor 5 which produced Abs_{340} readings higher than that of the negative (no enzyme) control, a kinetic assay was performed. The assay conditions were used as described for the inhibitor study assay, with the exception that the assay was distributed into a clear flat bottomed 96 well assay plate and placed into the SpectraMax® M3 plate reader set to read the absorbance at 340 nm over 40 min with 5 min reading intervals. The assay temperature was set at 25 °C. An NADPH control (reaction lacking enzyme) as well as a background control lacking NADPH was included. Results are shown in Figure 49.

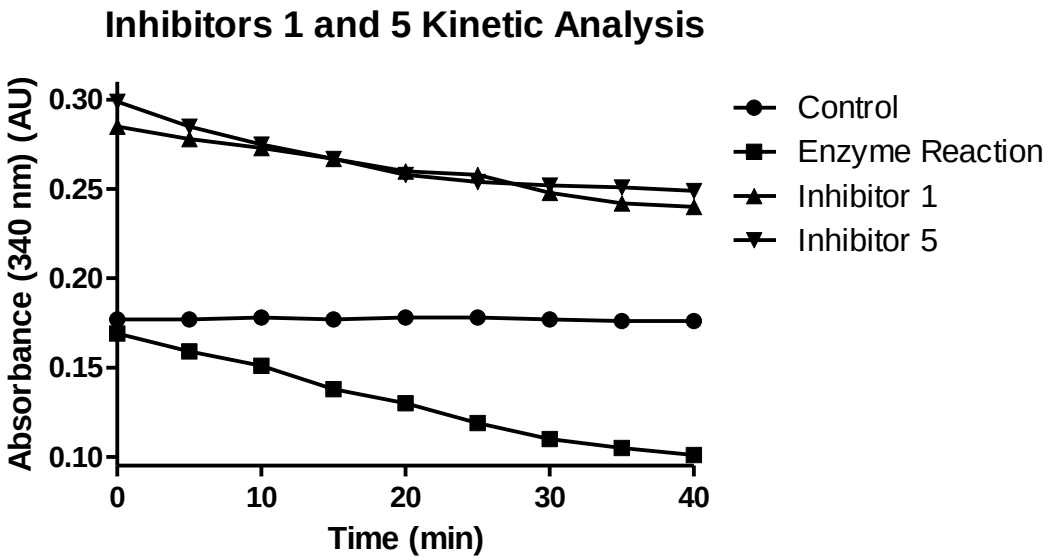


Figure 49: Kinetic analysis of Inhibitor 1 and Inhibitor 5 of *T. brucei* PTR1. The reaction was performed using 50 µM 6-biopterin, 50 µM NADPH and 0.1 µM T. brucei PTR1 enzyme with the addition of 50 µM inhibitor. The reaction was run in SpectraMax® M3 plate reader over 60 min, with reading intervals of 5 min. A negative NADPH (50 µM) control was used, as well as a background control. Results are depicted with background absorbance values subtracted and are shown as the mean of technical triplicate readings ± standard deviation.

Figure 49, shows a decrease in the absorbance of the enzyme reaction and the reaction with inhibitors. It is noted that the inhibitors starting point is substantially higher than that of the NADPH control and so too is the final absorbance readings. The enzyme reaction starting point is very close to that of the NADPH control. The two inhibitors possess a very similar reaction profile, with inhibitor 1 starting at a higher level and ending at a lower level than inhibitor 5. This could be because these inhibitors possess similar absorbance profiles and similar inhibition properties.

Table 13: ΔAbs_{340} value of Inhibitor 1 and Inhibitor 5 compared to the enzyme reaction. The ΔAbs_{340} was calculated from the difference between the initial and final values of the inhibitors and enzyme reaction and the enzyme reaction.

Individual Assay Components	ΔAbs_{340}
Enzyme Reaction	0.062 (40.24 %)
Inhibitor 1	0.045 (15.91 %)
Inhibitor 5	0.051 (16.83 %)

Table 13 measures the ΔAbs_{340} of the Inhibitors and the enzyme reaction. Enzyme activity remains constant with a ΔAbs_{340} of 40.24 %. The compounds absorb at 340 nm which can be seen from the kinetic curve starting and ending at higher readings than the negative (NADPH) control. The compounds yielded negative results during the initial end-point assay, thus a kinetic assay was performed. Despite this, comparing the ΔAbs_{340} of Inhibitor 1 (15.91 %) and Inhibitor 5 (16.83 %) to that of the enzyme reaction (40.24 %), the inhibitors do possess inhibitory effects, inhibiting PTR1 activity by approximately 60 %.

4.3.2 Resazurin Assay

In an attempt to improve on the results of the *T. brucei* PTR1 absorbance assay, it was opted to use resazurin to observe the changes in NADPH concentration. Resazurin is used to detect NADPH and NADH levels as it is reduced from a blue/purple colour to a pink coloured compound resorufin, which is highly fluorescent.

4.3.2.1 Resazurin Fluorescence Assay

To determine whether using resazurin is viable as a detection method, the assay was set up as determined by the optimised NADPH absorbance screening assay and allowed to incubate at room temperature for 40 min. After incubation 5 µl of a diluted resazurin stock was added to the solution to give a final assay concentration of 5 µM and the assay was further incubated at room temperature for 30 min. The assay measured the fluorescence of resorufin formed by NADPH-mediated reduction of resazurin with excitation set at 560 nm and emission at 590 nm. Results are shown in Figure 50.

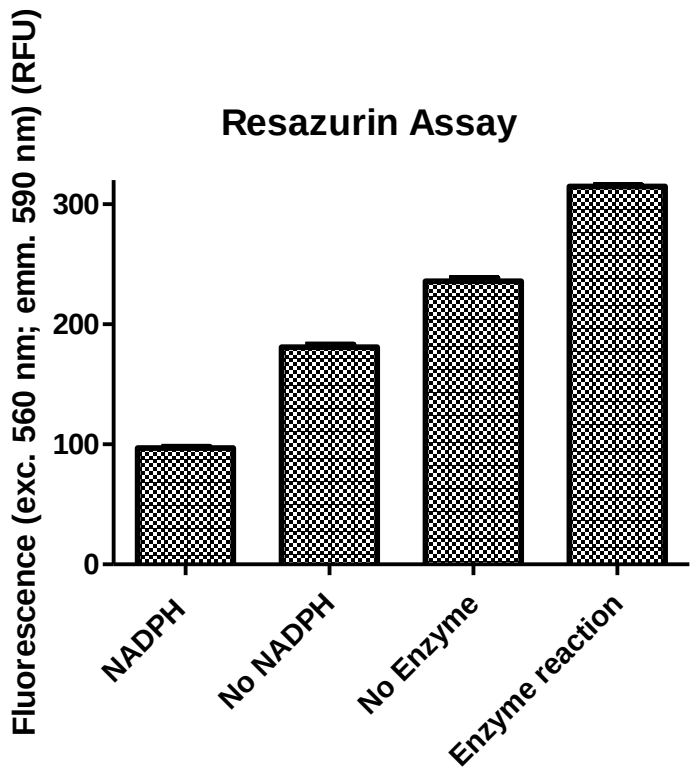


Figure 50: Resazurin fluorescence assay of *T. brucei* PTR1 with controls. The assay was performed using 50 µM NADPH, 50 µM 6-biopterin and 0.1 µM *T. brucei* PTR1 and incubated at room temperature for 40 min after which 5 µM resazurin was added and incubated for a further 30 min. The fluorescence was measured with excitation at 560 nm and emission at 590 nm using a SpectraMax® M3 plate reader. The bar show NADPH with other reaction reagents; No NADPH, Full enzyme reaction, 6-biopterin and *T. brucei* PTR1, lacking NADPH; No Enzyme, Full enzyme reaction, 6-biopterin and NADPH, lacking *T. brucei* PTR1; and Enzyme reaction, Full enzymes reaction with all reagents, 5-biopterin, NADPH and *T. brucei* PTR1. Results are shown as the mean of technical triplicate readings ± standard deviation.

From Figure 50 it is noted that the results are opposite to what is expected. As NADPH converts resazurin to resorufin, the NADPH control should result in a high fluorescence level. The enzyme reaction would have consumed the majority of the NADPH by the time resazurin was added thus the full reaction should have resulted in the lowest fluorescence values, contrary to the high values depicted in Figure 36. Thus the experiment was unsuccessful.

4.3.2.2 Resazurin Fluorescence Assay at two lower resazurin concentrations.

In an attempt to establish the result observed in the initial resazurin assay and explore the impact the individual assay component has on resazurin, an assay was set up under the same conditions and concentrations as the above resazurin assay, using lower resazurin concentrations. Results are shown in Figure 51.

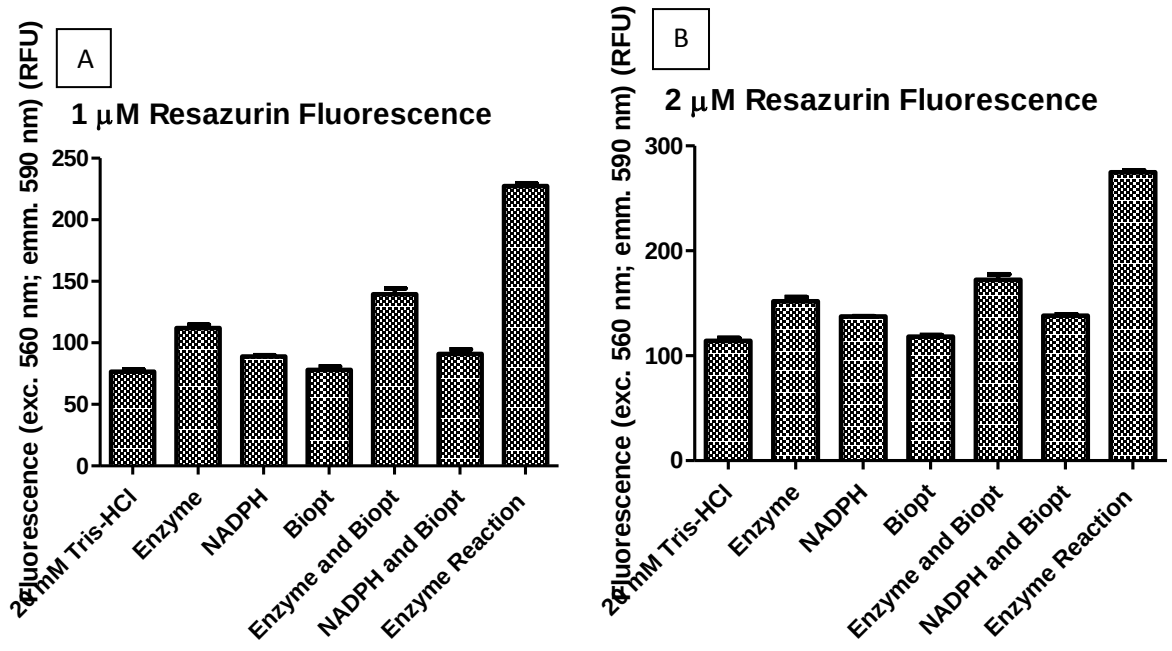


Figure 51: Resazurin fluorescence assay of *T. brucei* PTR1 assay components under differing resazurin concentrations 1 µM (A) and 2 µM (B). The assay was performed using 50 µM NADPH, 50 µM 6-biopterin and 0.1 µM *T. brucei* PTR1 and incubated at room temperature for 40 min after which 1 µM (A) and 2 µM (B) resazurin was added and incubated for a further 30 min. The fluorescence was measured with excitation at 560 nm and emission at 590 nm using a SpectraMax® M3 plate reader. Results are shown as the mean of technical triplicate readings ± standard deviation.

From Figure 51 the results were consistent with those obtained in Figure 50, in that the NADPH signal was lower than that obtained in the presence of the enzyme. The fluorescence values were higher when more resazurin is used. From the results obtained in Figure 51, it can be tentatively assumed that the PTR1 enzyme has some ability to reduce resazurin as the levels are consistently higher in the presence of the enzyme. It was also noted that it appears that 6-biopterin can reduce resazurin as well as when 6-biopterin was used with the enzyme the fluorescence values increased. The experiment was unsuccessful and thus cannot be used for the evaluation of PTR1 activity in a microplate assay.

4.3.3 HPPK Luminescence Assay

4.3.3.1 Initial *E. coli* HPPK and *S. enterica* HPPK Luminescence Assay

To determine the activity of the HPPK enzymes an initial assay was performed in a reaction mix containing 0.1 µM HPPK enzyme, 1.4 µM HMDP (substrate) and 1.5 µM ATP. The reaction was incubated at 30 °C for 30 min after which 200 µl Kinase-Glo® reagent was added and incubated for a further 10 min at room temperature. Once incubation was complete the solution was distributed

into a white 96 well assay plate and the luminescence was read. Refer to Section 4.2.5 under Methods and Materials. Results are shown in Figure 52.

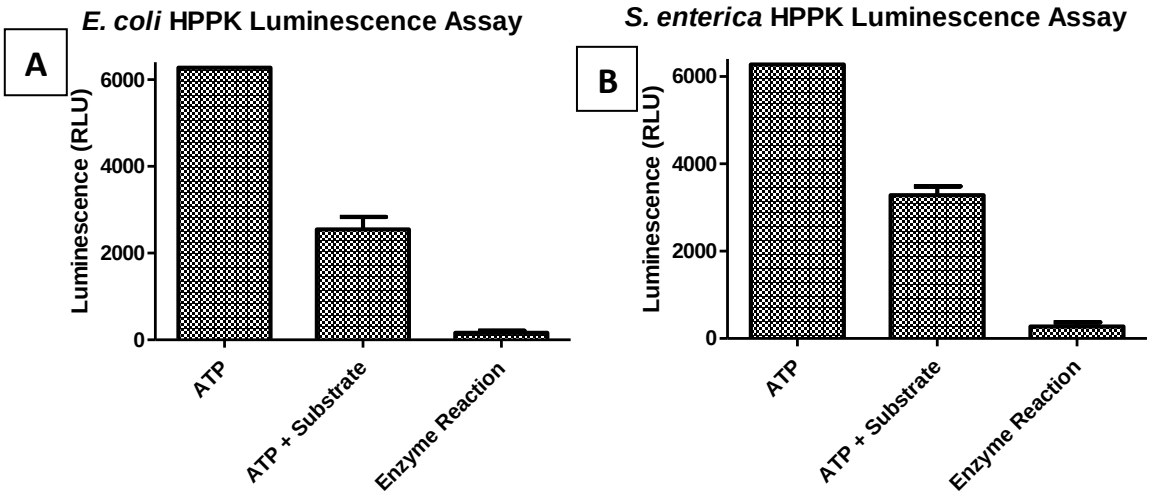


Figure 52: *E. coli* HPPK and *S. enterica* HPPK Luminescence assay for enzyme activity confirmation. The assay was performed using 0.1 µM HPPK, 1.4 µM HDMP and 1.5 µM ATP. The reaction was allowed to incubate at 30 °C for 30 min, after which 200 µl Kinase-Glo® reagent was added before the reaction was allowed to further incubate at room temperature of 10 min. The reaction luminescence was measured using a SpectraMax® M3 plate reader. A) shows the enzyme activity of *E. coli* HPPK and B) shows the enzyme activity of *S. enterica* HPPK, both with an ATP standard and a reaction without enzyme. ATP contained only ATP with reaction buffer; ATP + Substrate contained ATP and Substrate only with reaction buffer; and Enzyme reaction contained ATP, Substrate, HDMP and HPPK enzyme. Results are shown as the mean of technical triplicate readings ± standard deviation.

Figure 52 shows a large decrease in the luminescence reading after the full enzyme reaction had run to completion, thus indicating that the enzyme was active and the assay functional. When the reaction was run without the enzyme present there was a decrease in luminescence signal compared to the control of ATP alone, indicating that ATP was consumed (59.37 %). This indicates that the enzyme increases the rate of the reaction and that the reaction will proceed regardless of the presence or absence of the enzyme. To determine the quality of the assay for screening purposes and to ascertain if further optimisation of the assay needed to occur, the Z´-factors of the assay was calculated.

For Z´-factor calculations, the reaction without enzyme was used as the negative control and the enzyme reaction was used as the positive control.

Table 14: Z´-factors and ΔLum of the luminescence assay of HPPK

Enzyme	Z´-factors	ΔLum

E. coli HPPK	0.58	6109.202 (97.37 %)
S. enterica HPPK	0.70	5998.482 (95.61 %)

Table 14, indicated that the Z´-factors were sufficiently is above 0.5 to indicate a good assay for screening. The change in luminescence (ΔLum) values for the assay were very high which shows that the majority of the ATP was consumed during the enzyme reaction.

4.3.3.2 *E. coli* HPPK and *S. enterica* HPPK Luminescence Analysis Assay

To further explore the results obtained in Figure 38, the reaction was run using different parameters to assess assay functionality. All concentrations were kept the same at 0.1 µM HPPK enzyme, 1.4 µM HMDP and 1.5 µM ATP. The reaction was incubated at 30 °C for 30 min after which 200 µl Kinase-Glo® reagent was added and incubated for a further 10 min at room temperature. Once incubation was complete the solution was distributed into a white 96 well assay plate and the luminescence was read. Results are shown in Figure 53.

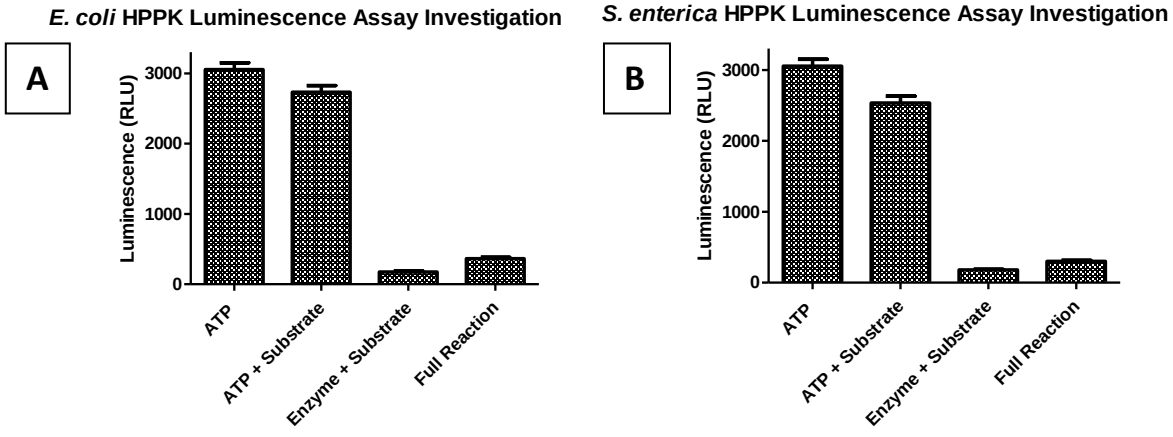


Figure 53: *E. coli* HPPK (A) and *S. enterica* HPPK (B) Luminescence assay for enzyme investigation. The assay was performed using 0.1 µM HPPK, 1.4 µM HDMP and 1.5 µM ATP. An ATP, substrate with ATP lacking enzyme and an enzyme and substrate lacking ATP controls were used for assay investigation. The reaction was allowed to incubate at 30°C for 30 min, after which 200 µl Kinase-Glo® reagent was added before the reaction was allowed to further incubate at room temperature of 10 min. The reaction luminescence was measured using a SpectraMax® M3 plate reader A) shows the enzyme activity of *E. coli* HPPK and B) shows the enzyme activity of *S. enterica* HPPK. ATP contained only ATP with reaction buffer; ATP + Substrate contained ATP and Substrate only with reaction buffer; Enzyme and Substrate contained only HPPK enzyme with Substrate in reaction buffer; and Enzyme reaction contained ATP, Substrate, HDMP and HPPK enzyme. Results are shown as the mean of technical triplicate readings ± standard deviation.

In Figure 53 we notice that in the absence of enzyme (ATP + Substrate) the luminescence readings are relatively high but still lower than that of the ATP standard with a 10.53 % ΔLum. This is expected as the reaction without the enzyme should have a high luminescence signal since little or no ATP should be consumed in the absence of the enzyme. Both HPPK enzymes depleted ATP by approximately 90 % compared to the ATP + substrate control, based on the ΔLum values.

Table 15: Z´-factors and ΔLum of luminescence assay of HPPK. The ΔLum measured the change in luminescence signal between the enzyme reaction and the ATP + substrate control.

Enzyme	Z´-factors	ΔLum
E. coli HPPK	0.86	2689.959 (88.09 %)
S. enterica HPPK	0.85	2753.328 (90.16 %)

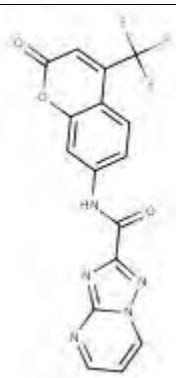
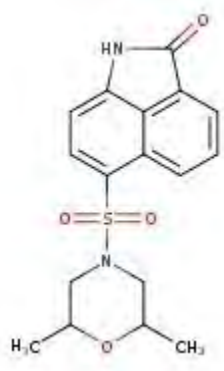
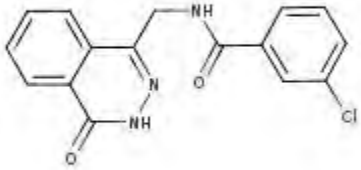
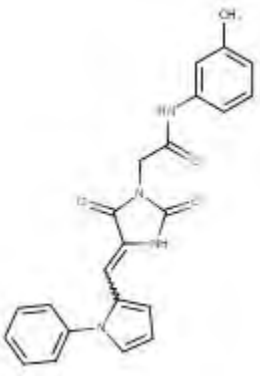
From Table 15, the Z´-factors had improved substantially and thus no optimisation of the assay was performed. The enzyme was active having a ΔLum of 88.09 % for *E. coli* HPPK and 90.16 % for *S. enterica* HPPK

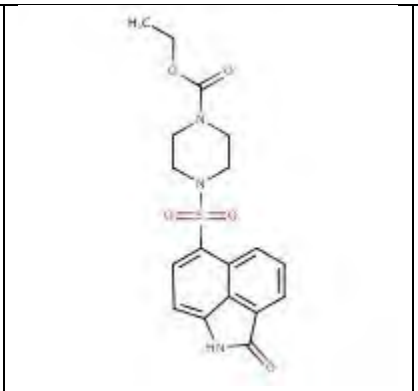
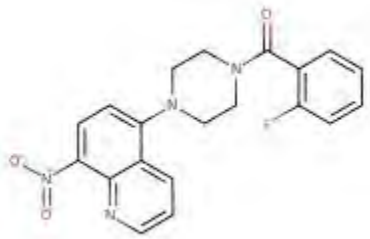
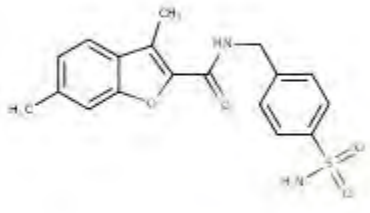
4.3.3.3 *E. coli* HPPK and *S. enterica* HPPK Inhibitor Screening

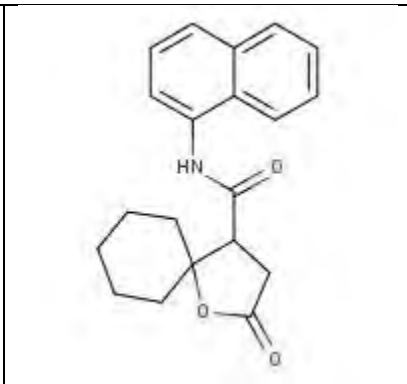
Once the assay was found to be of good quality for screening purposes, the compounds found in Chapter 2 of the dissertation was tested as inhibitors for the HPPK enzymes (Table 16). The assay was performed in a reaction mix containing 0.1 µM HPPK enzyme, 1.4 µM HMDP and 1.5 µM ATP to which inhibitors were added at a concentration of 50 µM. The reaction was incubated at 30 °C for 30 min after which 200 µl Kinase-Glo® reagent was added and incubated for a further 10 min at room temperature. Once incubation was complete the solution was distributed into a white 96 well assay plate and the luminescence was read. The results were represented graphically. All inhibitors were tested against both *E. coli* HPPK and *S. enterica* HPPK due to the large similarity of the enzymes. Inhibitors are shown in Table 16 and the inhibition results are shown in Figure 54 for *E. coli* and Figure 55 for *S. enterica*.

Table 16: Inhibitor names and structures retrieved from Molport and their ZINC number as used in Chapter 2. Inhibitors were randomly assigned a number from 1-9.

Inhibitor	Zinc Number	HPPK target	Structure	IUPAC Name

1	0923924	*E. coli and S. enterica*		N-[2-oxo-4-(trifluoromethyl)-2H-chromen-7-yl]-[1,2,4]triazolo[1,5-a]pyrimidine-2-carboxamide
2	4671178	*E. coli and S. enterica*		9-[(2,6-dimethylmorpholin-4-yl)sulfonyl]-2-azatricyclo[6.3.1.0^{4,12}]dodeca-1(11),4(12),5,7,9-pentaen-3-one
3	0082136	*S. enterica*		3-chloro-N-[(4-oxo-3,4-dihydrophthalazin-1-yl)methyl]benzamide
4	4498423	*S. enterica*		2-{2,5-dioxo-4-[(1-phenyl-1H-pyrrol-2-yl)methylidene]imidazolidin-1-yl}-N-(3-methylphenyl)acetamide

5	6792518	*S. enterica*		Ethyl4-({3-oxo-2-azatricyclo[6.3.1.0^{4,12}]dodeca-1(11),4(12),5,7,9-pentaen-9-yl}sulfonyl)piperazine-1-carboxylate
6	4499339	*E. coli*		5-[4-(2-fluorobenzoyl)piperazin-1-yl]-8-nitroquinoline
7	4476281	*E. coli and S. enterica*		2'-amino-5-fluoro-2,5'-dioxo-1,2-dihydro-5'H-spiro[indole-3,4'-pyrano[3,2-c]chromene]-3'-carbonitrile
8	0634089	*E. coli*		3,6-dimethyl-N-[(4-sulfamoylphenyl)methyl]-1-benzofuran-2-carboxamide

9	0310450	*E. coli*		N-(naphthalen-1-yl)-2-oxo-1-oxaspiro[4.5]decane-4-carboxamide

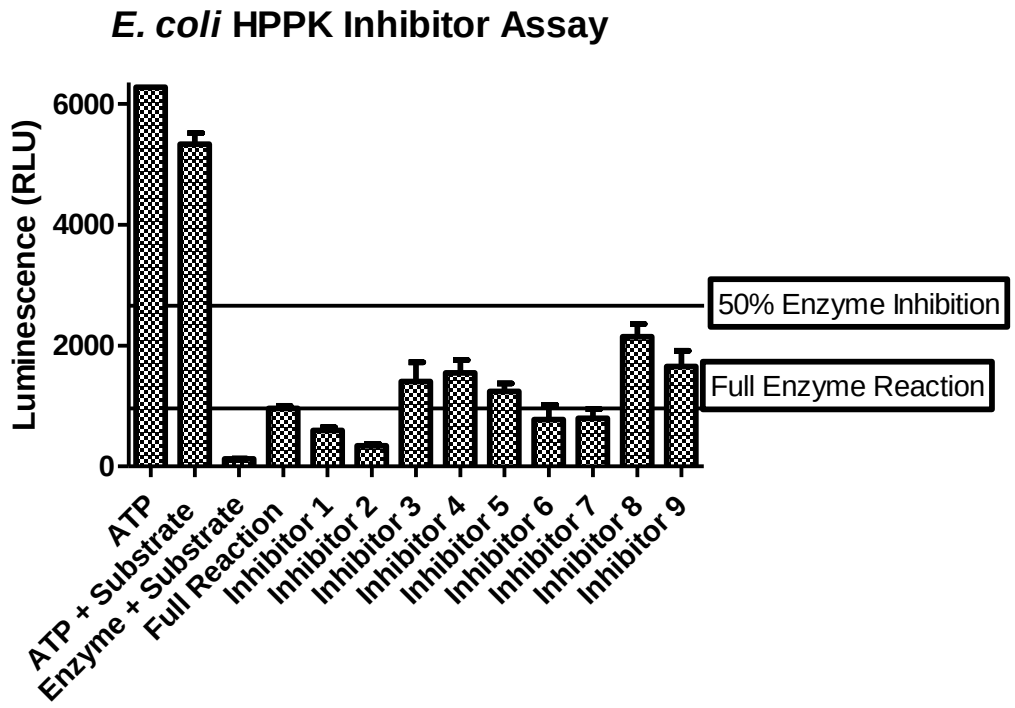


Figure 54: *E. coli* HPPK Inhibitory study. The assay was performed in a reaction mix containing 0.1 µM HPPK enzyme, 1.4 µM HMDP and 1.5 µM ATP to which 50 µM inhibitors were added. The reaction mix was incubated at 30 °C for 30 min after which 200 µl Kinase-Glo® was added and incubated for a further 10 min at room temperature. The reaction was distributed into a white 96 well assay plat and luminescence was read. ATP, ATP + Substrate and Enzyme + Substrate controls were set up. Results are shown as the mean of technical triplicate readings ± standard deviation.

Figure 54, shows the inhibitory effect of the nine inhibitors on the *E. coli* HPPK enzyme. The graph was labelled at the level at which the full enzyme reaction readings are and from that the 50 % enzyme inhibition was determined and labelled. None of the inhibitors was found to have a 50 % inhibitory effect on the enzyme (Table 17) and five, namely 3 (77.61 %), 4 (75.31 %), 5 (80.19 %), 8 (65.83 %), and 9 (73.65 %), were shown to have inhibitory effect falling between that of no inhibition

(Full enzyme reaction) and 50 % inhibition. In Figure 54, four inhibitors, namely 1 (90.54 %), 2 (94.54 %), 6 (87.69 %) and 7 (87.30 %) were found to have a lower luminescence reading than the full enzyme reaction. The possibility exists that, due to the *in silico* parameters which were set to bind directly to the catalytic binding site and the ATP binding site, these compounds were acting as false substrates for the enzyme. This will be explored in a subsequent section of this Chapter. None of the inhibitors was further explored to an attempt at an IC_{50} for the reason that a 50 % inhibition was not achieved.

Table 17: ΔLum of enzyme reaction and Inhibitors compared to the ATP control for *E. coli* HPPK. The ΔLum measured the change in luminescence signal between the enzyme reaction and the ATP + substrate control, with the % change shown in brackets.

	ΔLum
Enzyme Reaction	5313.570 (84.69 %)
Inhibitor 1	5680.610 (90.54 %)
Inhibitor 2	5931.604 (94.54 %)
Inhibitor 3	4870.572 (77.61 %)
Inhibitor 4	4725.131 (75.31 %)
Inhibitor 5	5031.568 (80.19 %)
Inhibitor 6	5501.526 (87.69 %)
Inhibitor 7	5477.435 (87.30 %)
Inhibitor 8	4130.592 (65.83 %)
Inhibitor 9	4620.964 (73.65 %)

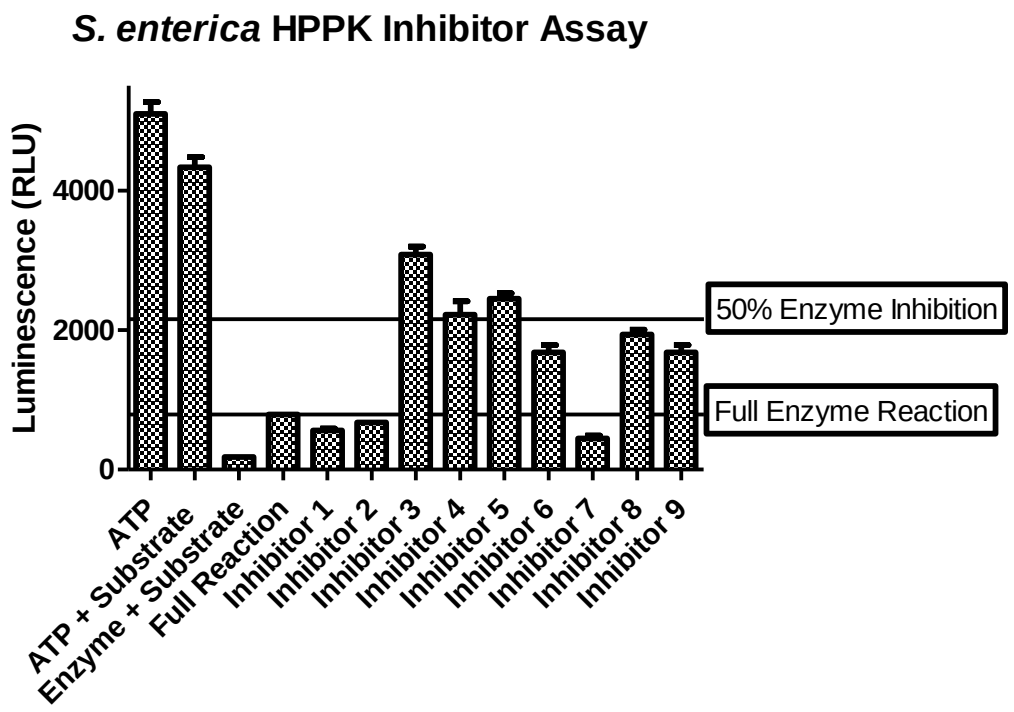


Figure 55: *S. enterica* HPPK Inhibitory study. The assay was performed in a reaction mix containing 0.1 µM HPPK enzyme, 1.4 µM HMDP and 1.5 µM ATP to which 50 µM inhibitors were added. The reaction mix was incubated at 30 °C for 30 min after which 200 µl Kinase-Glo® was added and incubated for a further 10 min at room temperature. The reaction was distributed into a white 96 well assay plat and luminescence was read. ATP, ATP + Substrate and Enzyme + Substrate controls were set up. Results are shown as the mean of technical triplicate readings ± standard deviation.

In Figure 55 the inhibitory effects of the nine inhibitors against *S. enterica* HPPK were determined. The full enzyme reaction was labelled as a standard measure and so was 50 % inhibition. In the graph, it is noted that six inhibitors (Table 18), namely 3 (86.73 %), 4 (56.49 %), 5 (52.00 %), 6 (67.07 %), 8 (62.02 %) and 9 (67.07 %), showed inhibitory effect with three of them showing 50 % or more inhibitory effect showing residual ATP (comparing the ATP + substrate control to the inhibitors, calculations were performed by dividing the difference between the ATP control and the inhibitors reaction, multiplied by 100). It was observed that three inhibitors, namely 1 (88.99 %), 2 (86.73 %) and 7 (91.20 %), showed a decrease in the luminescence reading which could indicate that they are acting as false substrates for the enzyme. Inhibitor 3 was shown to provide the best inhibitory effect in the single concentration assay and therefore an IC_{50} was performed.

Table 18: ΔLum of enzyme reaction and Inhibitors compared to ATP control for *S. enterica* HPPK. The ΔLum measured the change in luminescence signal between the enzyme reaction and the ATP + substrate control, with the % change shown in brackets.

	ΔLum

Enzyme Reaction	4310.340 (84.53 %)
Inhibitor 1	4538.121 (88.99 %)
Inhibitor 2	4422.868 (86.73 %)
Inhibitor 3	2018.028 (39.58 %)
Inhibitor 4	2880.498 (56.49 %)
Inhibitor 5	2654.665 (52.00 %)
Inhibitor 6	3420.028 (67.07 %)
Inhibitor 7	4650.796 (91.20 %)
Inhibitor 8	3162.702 (62.02 %)
Inhibitor 9	3420.028 (67.07 %)

4.3.3.4 *S. enterica* HPPK IC_{50}

For determination of an IC_{50}, the inhibitor showing the highest level of inhibition was selected, namely inhibitor 3. The assay was performed in a reaction mix of 200 µl, containing 0.1 µM HPPK enzyme, 1.4 µM HMDP (substrate) and 1.5 µM ATP to which inhibitor 3 was added in a serial dilution series starting at a concentration of 100 µM to a concentration of 6.25 µM. The reaction was incubated at 30 °C for 30 min after which 200 µl Kinase-Glo® reagent was added and incubated for a further 10 min at room temperature. Once incubation was complete the solution was distributed into a white 96 well assay plate and the luminescence was read. The results are represented graphically in Figure 56.

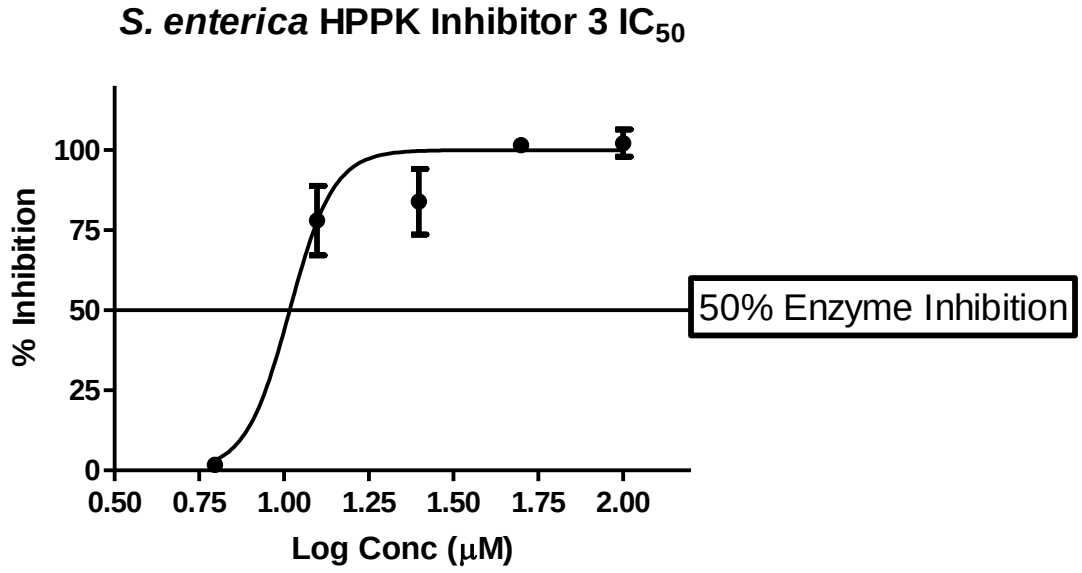


Figure 56: IC_{50} of Inhibitor 3 against *S. enterica* HPPK. The assay was performed in a reaction mix containing 0.1 µM HPPK enzyme, 1.4 µM HMDP and 1.5 µM ATP to which a two-fold serial dilution of Inhibitor 3 was added from 100 µM to 6.25 µM. The reaction mix was incubated at 30 °C for 30 min after which 200 µl Kinase-Glo® was added and incubated for a further 10 min at room temperature. The reaction was distributed into a white 96 well assay plate and luminescence was read. An ATP control was included and used to calculate 0 % inhibition in the preparation of the IC_{50} graph. Results are shown as the mean of technical triplicate readings ± standard deviation.

Figure 56, shows that the IC_{50} of inhibitor 3 was 10.4 µM. This indicates that Inhibitor 3 is an excellent inhibitor of *S. enterica* HPPK when compared to the previously reported HPPK inhibitor 8MG (IC_{50} of 41 µM against *S. aureus* HPPK)(Shi *et al.*, 2001). A more accurate IC50 value may be obtained by repeating the dose-response assay, since the inhibitory activity of the compound may alter from one experiment to the next. An example is that 50 uM of compound 3 showed full inhibition of the enzyme in the dose-response assay, but only 75% inhibition in the initial fixed concentration screen (Figure 55).

4.3.3.5 *E. coli* HPPK and *S. enterica* HPPK False Substrate Assay.

It was noted that the luminescence signal decreased to below that of the full enzyme reaction when inhibitor 2 was added. Thus to determine whether inhibitor 2 was acting as a false substrate for the enzyme the following assay was used. The assay was performed in a reaction mix containing 0.1 µM HPPK enzyme, 1.4 µM HMDP and 1.5 µM ATP and in separate microcentrifuges tubes, two concentrations of inhibitor 2 were selected, being 50 µM and 25 µM. These tubes did not contain any HMDP (the normal HPPK substrate). The reaction was incubated at 30 °C for 30 min after which 200 µl Kinase-Glo® reagent was added and incubated for a further 10 min at room temperature. Once incubation was complete the solution was distributed into a white 96 well assay plate and the luminescence was read. Results are shown in Figure 57.

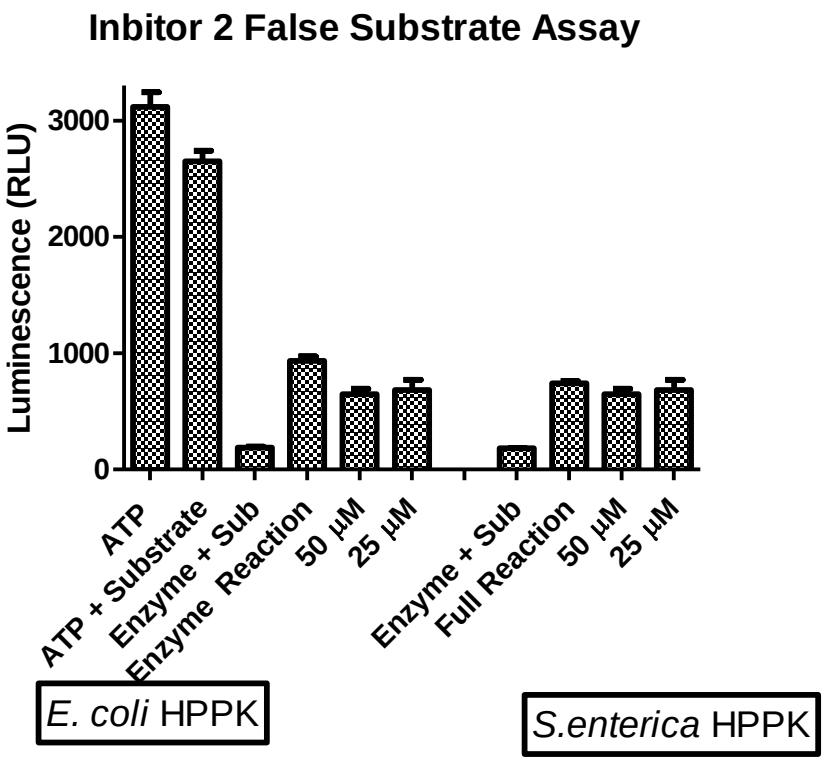


Figure 57: Inhibitor 2 false substrate assay performed with both *E. coli* HPPK and *S. enterica* HPPK. The assay was performed in a reaction mix containing 0.1 µM HPPK enzyme, 1.4 µM HMDP and 1.5 µM ATP with separate reaction mixes lacking HMDP and containing two concentrations of Inhibitor 2 (50 µM and 25 µM). The reaction mix was incubated at 30 °C for 30 min after which 200 µl Kinase-Glo® was added and incubated for a further 10 min at room temperature. The reaction was distributed into a white 96 well assay plat and luminescence was read. ATP, ATP + Substrate and Enzyme + Substrate controls were set up. Results are shown as the mean of technical triplicate readings ± standard deviation.

From Figure 57, both HPPK enzymes showed a significant reduction in ATP when the normal enzyme reaction was analysed. The reduction was increased in presence of inhibitor 2 without HMDP. Thus one can presume that the enzyme is using the inhibitors to act as false substrates. The difference in remaining ATP levels between the two concentrations (50 µM and 25 µM) of inhibitor was very small, indicating that the enzyme acts more efficiently in the presence of the inhibitor than HMDP in its ability to consume ATP. This could be because the inhibitors are potentially activating the enzyme. Further investigation into the inhibitors should occur to determine their ability to act as false substrates, by the use of HPLC and mass spectrometry on the product of the reaction to determine whether the inhibitor structure is altered.

4.4 Discussion

The *T. Brucei* PTR1 NADPH absorbance assay was optimised to yield an effective screening assay using a strategic optimisation strategy. From the initial assay to the final optimised assay there was an increase in Z'-factors indicating a robust assay that can be used for screening (Zhang, Chung and Oldenburg, 1999). The final Z´-factor score of the *T. Brucei* PTR1 NADPH absorbance assay was 0.94, thus providing a verified, robust and sensitive assay for screening purposes. Optimisation strategies employed included NADPH concentration, enzyme concentration and reaction time. Throughout the optimisation process, it was decided that the lowest concentration which yielded an improvement in Z'-factor would be used to minimise the use of reagents and therefore the cost of the assay without compromising the results of the assay. After optimisation was completed screening of compounds that were previously found to dock with PTR1 and have inhibitory effects on the live parasite was done. The decision to screen these compounds came as a method of verification that the enzyme was indeed the target causing parasite death and as this study dealt with the discovery of novel anti-folates it would be a good fit within this study. The compounds were determined by bioinformatics *in silico* screening techniques by Kimuda et al. (2019), who also assisted with the bioinformatics performed in Chapter 2 of this dissertation. Although concentrations between wells of the assay were equal the initial readings at time 0 min varied between wells. This is most likely due to a lag time between plating and plate insertion into the spectrophotometer which could have been overcome had auto-injectors been available. With the addition of inhibitors, the ΔAbs_{340} was expected to decrease throughout the assay as little to no NADPH would have been consumed if the inhibitor binds to the enzyme. For three inhibitors this was the case where a marked change in absorbance was noted, unfortunately, two of the inhibitors produced a significant increase in the absorbance values above that of the NADPH control. The increase in absorbance reading can be explained by the fact that these compounds possibly absorb light at 340 nm. As 340 nm falls within the UV-visible spectrum (200 nm - 800 nm), it states that when a molecule is exposed to a specific wavelength of light, in this case, 340 nm, the light excites the electrons from the ground state to the excited state (Østergaard, 2016). The functional groups of molecules which absorb light in the UV-vis spectrum are called chromophores and are characterised by the presence of delocalised π-electrons. With an increase in the number of π-electrons within the structure of a molecule it allows for these electrons to be spread out over the entire molecule. It is noted that π-electrons occur within double or triple bonds or within fused aromatic ring systems in which tautomerisation will result in the movement of electrons due to the delocalised nature (Mäntele and Deniz, 2017). 1,8-Diazabicyclo[5.4.0]undec-7-ene (DBU), is an example of a fused ring system containing non-bonding electrons as well as π-electrons which will undergo tautomerisation and thus spread the electrons over the entirety of the molecule (Kaljurand *et al.*, 2000). From DBU it is noted that nitrogen atoms,

as well as oxygen atoms, can assist in the delocalisation of electrons throughout a molecule (Cruz-Cruz *et al.*, 2015). The presence of the above-mentioned chemistry can be seen in Inhibitor 1 and Inhibitor 5. All of the inhibitors used contain chemical moieties which can result in the delocalisation of electrons, although only in specific portions of the compound. The structure of Inhibitor 1 and 5 shows that the delocalisation effect is more widespread over the entirety of the molecule, thus increasing the overall absorbance of the molecule. To further investigate the inhibitory effects of Inhibitor 1 and 5, a kinetic assay was performed to determine the ΔAbs_{340} occurring throughout the reaction. From this experiment, the ΔAbs_{340} was significantly lower than that of the enzyme reaction over the reaction period indicating that the two compounds did inhibit the enzyme. NADPH fluorescence measurements could be used as an alternative to measuring absorbance at 340 nm to circumvent test compounds that also potentially absorb at 340 nm (White, Anandraj and Trois, 2014). Alternatively commercially available colorimetric and fluorogenic NADP/NADPH assay kits could be employed such as Sigma-Aldrich MAK312 and MAK038. In this study, the compounds identified by Kimuda et al., (2019) were randomised to create a level of "blindness" for the determination of the efficacy of inhibition. From Table 19, the lowest % ΔAbs_{340}, corresponding to the greatest enzyme inhibition, is Inhibitor 3 (15.61 %), which produced an IC_{50} of 25.4 µM against Trypanosomes. Inhibitor 5 proves to be most effective against Trypanosome with an IC_{50} of 9.6 µM, which corresponds to having a good inhibitory effect against the enzyme (16.83 % ΔAbs_{340}). Inhibitor 4 proves to be the second most effective against Trypanosomes (IC_{50} 12.7µM), has the lowest inhibitory effect against the enzyme (24.47 % ΔAbs_{340}). Inhibitor 1 also shows good inhibitory effects against the parasite (IC_{50} of 14.6 µM) and against the enzyme with a % ΔAbs_{340} of 15.91. Inhibitor 2 had the lowest inhibitory effects with a % ΔAbs_{340} of 31.02 % and likewise had the lowest inhibitory effects against the parasite (IC_{50} 34.9 µM)

Table 19: % ΔAbs_{340} and Trypanosome parasite inhibition assay summary.

	% ΔAbs_{340}	Trypanosome Parasite assay IC_{50} µM
Enzyme reaction	40.24	N/A
Inhibitor 1	15.91	14.6
Inhibitor 2	31.02	34.9
Inhibitor 3	15.61	25.4

Inhibitor 4	24.47	12.7
Inhibitor 5	16.83	9.6

An attempt was made to determine the inhibitory effects of the compounds by using a resazurin assay. Resazurin is a weakly fluorescent phenoxazine dye, which results in a colour change from blue to purple when it is reduced to resorufin. In the presence of NADPH or NADH resazurin is reduced to the highly fluorescent resorufin, which can easily be detected. The general use of resazurin is for cell viability where it is reduced to resorufin by aerobic respiration by a metabolically active cell (Taneja and Tyagi, 2007). The *T. brucei* PTR1 assay uses NADPH as a reagent and thus the decision was made to determine whether resazurin could be used for quantitation of the consumption of NADPH. Unfortunately, the assay was not successful and it was determined that in the presence of the PTR1 enzyme more resazurin is converted to resorufin than in NADPH negative control. Likewise, the resorufin levels were much higher when the full reaction occurred. This is an indication that the enzyme is possibly acting as a reducing agent to resazurin. The assay was not pursued further as a means of quantification. From what was noted within the resazurin assay PTR1 having slight reducing activity, the enzyme could be explored in more detail to determine if it has any commercial application as a reduction catalyst, which is beyond the scope of this dissertation.

Luminescence signal detection of residual ATP proved an effective means for detecting HPPK enzyme activity. The change in luminescence signal from the ATP standard compared to the full enzyme reaction was very high indicating that the enzyme was active and could be used for screening. The luminescence assay works on the principle of light emission. Luciferase is a generic name given to all enzymes that result in bioluminescence, and reactive oxygen is required during the reaction (Cheng *et al.*, 2012). In this study, the luciferase enzyme from fireflies catalysed the reaction between ATP and luciferin which releases a photon of light in the process (Promega corporation, 2015). What was concerning was that there was a large difference between ATP standards from one experiment to the next. It has been documented that any form of detergent or surfactant can result in the formation of bubbles within the assay which will negatively affect the luminescence reading. Furthermore, factors such as the sensitivity of the luminometer, the number of reagents used in the assay and the proximity of high signal wells, will all negatively affect the readings (Iragué *et al.*, 2018). The higher the concentration of assay reagents used, namely ATP, enzyme and substrate, the further from physiological conditions the assay becomes and thus aberrant results will be obtained. Likewise, when a very high luminescence signal well is close to that of a low luminescence signal

well, spillover from one well to the other will result which will cause a false high result (Cheng *et al.*, 2012). This then brings into question the quality of the assay plate, as a low-quality plate could potentially contain areas that are more opaque than others thus increasing the chances of signal spilling into adjacent wells. ATP contamination from sources such as skin cells, bacteria or other organisms will also have a detrimental effect on the quality of the assay, thus gloves were always worn when the assay was performed. The Z′-factor of the HPPK luminescence assay was sufficiently high and the decision was made to proceed. Regardless of the presence of the enzyme ATP was consumed in the reaction, thus giving insight that the chemical reaction will occur in the absence of the enzyme indicating that HPPK dramatically increases the rate at which the reaction occurs, as enzymes function as catalysts in chemical reactions. In further experimentation regarding HPPK, the enzyme kinetic could be explored further. The inhibitor studies again proved challenging but fruitful, summarised in Table 20. The inhibitors were randomly assigned inhibitor numbers from those determined in Chapter 2 to increase "blindness" in the study. It was found that Inhibitors 1, 2 and 7 which corresponded with compounds that show strong *in silico* binding to the HPPK enzymes of both *E. coli* and *S. enterica* produced a lower luminescence signal compared to the full enzyme reaction and the other inhibitors. This indicates a potential false substrate which was explored further and shows a strong indication that they do act as false substrates, which could be confirmed by using bacterial assays to test the theory. If the compounds act as false substrates, the HPPK enzyme will produce a dead-end product that will not be usable by the bacterium for nucleic acid production (which requires folate synthesis) leading to inhibition of bacterial viability. Inhibitor 7 showed a high degree of inhibition (IC_{50} of 3.9 µM) against malaria parasites (*Plasmodium falciparum* [*P. falciparum*]) which then supports the indications of a false substrate. Inhibitor 7 showed a significant decrease in luminescence signal (lower than that observed by the full enzyme reaction) indicating that the enzyme could potentially be using the inhibitor as a false substrate. Potentially an assay could be developed in which the concentration of the compound after the reaction has run to completion is determined using HPLC or other similar quantitative chromatographic methods. Inhibitors 6, 8 and 9 were selective for *E. coli* HPPK *in silico* studies and showed a more pronounced effect on *E. coli* HPPK than on *S. enterica* HPPK. Inhibitor 6 again showed a decrease in luminescence signal compared to the full enzyme reaction. Unfortunately, a 50% inhibitory effect was not noted with Inhibitors 6, 8 or 9, thus none of the inhibitors were selected for the performance of an IC_{50}. Inhibitors 3, 4 and 5 were selective for *S. enterica* HPPK *in silico*, and proved to have a higher inhibitory effect on the enzyme than any other inhibitor. Inhibitor 3 yielded the best inhibitory effect and was selected for an IC_{50} determination. The IC_{50} yielded a positive result, with Inhibitor 3 showing an IC_{50} of 10.4 µM against *S. enterica* HPPK.

Table 20: *In silico* Screening results of *E. coli* and *S. enterica* HPPK with % ΔLum and *P. falciparum* parasite inhibitory assays summary.

Inhibitor	*In silico* Screening organism	% ΔLum *E. coli* HPPK	% ΔLum *S. enterica* HPPK	*P. falciparum* % viability
Enzyme Reaction	N/A	84.69	84.53	N/A
1	*E. coli and S. enterica*	90.54	88.99	68.16
2	*E. coli and S. enterica*	94.54	86.73	70.30
3	*S. enterica*	77.61	39.58 (IC_{50} = 10.4 µM)	76.73
4	*S. enterica*	75.31	56.49	76.14
5	*S. enterica*	80.19	52.00	66.21
6	*E. coli*	87.69	67.07	21.55 (IC_{50} = 5.8 µM)
7	*E. coli and S. enterica*	87.30	91.20	23.11 (IC_{50} = 3.9 µM)
8	*E. coli*	65.83	62.02	66.21
9	*E. coli*	73.65	67.07	63.29

The inhibitors were analysed by Tarryn Swart (Post Doctorate) in the group for activity against *P. falciparum* parasites (Appendix D). All inhibitors resulted in less than 80 % parasite viability at 50 µM. The results showed that Inhibitor 6 and 7 resulted in less than 50 % parasite viability at a concentration of 50 µM and an IC_{50} assay yielded 5.8 µM for inhibitor 6 and 3.9 µM for inhibitor 7 (Appendix E). Inhibitor 7 which provided the best results against *P. falciparum* parasites also had inhibitory effects on both *E. coli* HPPK and *S. enterica* HPPK. The *P. falciparum* parasite has a bifunctional HPPK/DHPS enzyme system and the result of parasite death due to potential HPPK inhibitors is worth exploring using the bifunctional enzyme system (Dennis *et al.*, 2014). The

potential exploration into drug optimisation could result in improved anti-malarial effect via folate enzyme inhibition. The results from these studies are indicative that HPPK is possibly the target of the compounds although to know for certain further experimentation is needed. The compounds were also analysed for their toxicity and were found to be non-toxic (Appendix F) to HeLa (human cervix adenocarcinoma) cells. While this result does indicate low overall toxicity, as cancer cells grow more rapidly than primary cells, it should be noted that primary cell line would have provided a more reliable indication of toxicity as the potential exists that primary cells possess cell receptors and channels which differ from cancer cells. Unfortunately, primary cell lines are not available at the institution currently

To conclude, the assays used proved to provide favourable results in screening compounds with the identification of potential novel inhibitors. Further experimentation with regards to bacterial assays could potentially provide more compelling results to put forward for compound optimisation which could then provide improved results in drug development. From the results in this Chapter along with the appendices, the potential for a broad spectrum HPPK inhibitor exists, and that coupled with the difficulty in developing drugs specifically for Gram-negative organisms is the reason for further investigation and optimisation. The bacterial species used in this study were both Gram-negative.

CHAPTER 5

General Conclusions

Folate enzyme inhibitors have been used extensively in the treatment and modification of diseases, with inhibitors covering bacterial infections, psoriasis, rheumatism and neoplasia. Antimicrobials such as sulphonamides (inhibiting DHPS), trimethoprim and pyrimethamine (inhibiting DHFR) have been deemed essential medicines indicating the important role of these agents. Methotrexate an inhibitor of DHFR has shown both antimicrobial action as well as being used as chemotherapy and a disease-modifying agent in psoriasis and rheumatism (Bertacine Dias *et al.*, 2018). The rationale behind folate enzyme inhibitors is well established, as folates form an integral aspect in the synthesis of nucleic acids. Enzymes such as PTR1, DHPS, DHFR and HPPK, to name a few, have had numerous research studies conducted on them, with ample application resulting from such studies in cases such as DHPS and DHFR and little to none from PTR1 and HPPK (Bermingham and Derrick, 2002)(Cullia *et al.*, 2018). The number of novel drug candidates being produced by phenotypic screening approaches against whole cells or organisms has decreased steadily. One possible reason for this decline in the discovery of new chemical entities may be that all the chemotypes that are available in compound libraries have already been screened. This has led to an increasing emphasis on target-based drug discovery programs, such as *in silico* screening, which provides knowledge of target-compound interaction, facilitating structure-activity relationship (SAR) analysis and thus compound modification and optimisation through synthetic chemistry. Target-based drug discovery additionally allows for more competent prediction of the cross-species activity of inhibitors, as well as possible resistance and how to compensate for it (Radoshitzky *et al.*, 2012). As the use of bioinformatics as a drug discovery technique is increasing, the need for a stringent protocol such as target-based assays and live organism screening should be employed to validate the results and use them as means to determine whether compound optimisation should be performed. This will have a huge impact on the number of drugs that can be used to stem the looming bacterial resistance threat, particularly in gram-negative organisms (Gershell and Atkins, 2003).

The modelling of *S. enterica* HPPK proved to be effective and through the process showed a high degree of similarity with *E. coli* HPPK. From the *in silico* docking studies, nine compounds were identified as interacting very strongly with the HPPK enzymes. This study has been the first to perform homology modelling and *in silico* docking of the *S. enterica* HPPK enzyme, as at the time of experimentation and writing, no others were identified, thus making drawing comparisons impossible. The *E. coli* HPPK enzyme has been studied both *in silico* and *in vitro*, although the emphasis of this study was to identify compounds for both *S. enterica* HPPK and *E. coli* HPPK, as both

are Gram-negative organisms with a high degree of antimicrobial resistance (Reeve *et al.*, 2016). It was shown that the use of bioinformatic techniques as a means to identity compounds for further *in vitro* target assessment is valid and does yield positive results. This was demonstrated by the successful modelling of the *S. enterica* HPPK enzyme, which was shown to have a high degree of similarity to similar-sized proteins and the internal energy configuration of the model was favourable with no high energy configurations being found. The *in silico* docking performed on *E. coli* HPPK and *S. enterica* HPPK yielded 9 interacting compounds all showing low binding energies and small binding distances, which were ideal for further *in vitro* target-based bioassay investigation.

The expression and purification of the *T. brucei* PTR1 and *S. enterica* HPPK and *E. coli* HPPK enzymes could potentially have been optimised further to yield a greater quantity of recombinant protein, although for use in this study the concentrations obtained were sufficient. Screening of larger compound collections may require further optimisation of the expression protocol to improve enzyme yields. The HPPK enzymes require a lower temperature for expression over a longer time to prevent misfolding of the protein. The *T. brucei* PTR1 protein does not require the same low temperature, but it appears that at a moderate temperature the protein expresses well.

The *T. brucei* PTR1 enzyme was tested for confirmation of the target. In a study performed by Kimuda, Laming, Hoppe, & Bishop, (2019), the compounds identified through *in silico* methods were tested against the parasite and shown to cause parasite death. This study intended to determine whether parasite death was indeed due to *T. brucei* PTR1 inhibition. It was found that while there does appear to be a decrease in enzyme activity, not all of the inhibitors showed a 50 % inhibition at 50 µM and were not further investigated. In particular, two of the inhibitors were shown to possess absorbance activity at 340 nm, which made determining their inhibitory effects difficult. The inhibitory effects noted by Kimuda et al., (2019) compared well to those obtained in this study, indicating that parasite death is possibly due to *T. brucei* PTR1 inhibition. However, this conclusion is tenuous without further corroborating experiments. This could include: 1) PTR1 silencing, which should increase the parasites sensitivity to the compounds, 2) overexpression of PTR1, which should decrease parasite sensitivity to the compounds, 3) determining, using mass spectroscopy, whether the concentrations of PTR1 substrate and product increase and decrease respectively in the parasite when treated with the compounds or 4) determining if the PTR1 inhibitors synergise with other antifolates using isobologram analysis.

The use of resazurin for conducting target-based assays, although theoretically plausible, proved of little value in this study. From the results, it was found that the fluorescence signal was greatest in the presence of the enzyme and smallest in the negative NADHP control. This result is the exact

opposite of what is expected, in that resazurin should be reduced to resorufin in the presence of NADPH thus resulting in a high fluorescence signal. In an attempt to investigate the potential effect each individual assay component had on resazurin, an experiment was set up testing each component individually with resazurin. The results remained unchanged but indicated that *T. brucei* PTR1 possibly acts as a reducing enzyme with the ability to reduce compounds other than its native substrate. In this regard further experimentation of the enzymes' capacity as a biological catalyst could be performed, although it would be beyond the scope of this dissertation.

Inhibitors that showed strong interaction with *E. coli* HPPK *in silico* did show *E. coli* HPPK inhibition, but unfortunately, this did not translate into a sufficiently substantial inhibition of > 50 %, to warrant further investigation. In contrast, the compounds which showed strong *in silico* interaction with *S. enterica* HPPK did indeed prove to have a greater than 50 % inhibitory effect, thus one of these compounds were selected for an IC_{50} determination. The IC_{50} of Inhibitor 3 was calculated to be 10.4 µM, making it a promising hit compound to explore further, especially when considering the IC_{50} of a known lead compound identified via *in silico* docking studies to be an HPPK inhibitor, 8MG, which is 41 µM. Inhibitors that showed interaction with both *S. enterica* HPPK and *E. coli* HPPK *in silico*, showed low luminescence readings, lower than that of the enzyme reaction, indicating that they may act as false substrates. The experiment to determine whether they do act as false substrates, tentatively yielded positive results. To better interpret the results in this study, the use of bacterial assays should be employed, for confirmation of inhibitory effects. If the inhibitors do act as false substrates they will produce a dead-end product incapable of being used by the bacterium and bacterial death will result.

The nine inhibitors used against *E. coli* HPPK and *S. enterica* HPPK were tested against live *P. falciparum* parasites and two showed a high degree of parasite death. The *P. falciparum* parasite, although in a bifunctional form, does possess an HPPK enzyme and the inhibitor which had the best IC_{50} against the *P. falciparum* parasite, Inhibitor 7, also had activity against both of the HPPK enzymes in what appears to be a false substrate action.

This study has supported the efficacy of bioinformatic analysis as a tool for use in drug discovery programs, specifically the use of techniques such as homology modelling and *in silico* docking studies. The results supported the hypothesis that PTR1 inhibition is the mechanism by which the compounds identified by Kimuda *et al.*,(2019) caused parasite death. The results suggest that using homology modelling and *in silico* docking, with *S. enterica* HPPK is an effective approach to identify compounds with low micromolar activity against the enzyme. This approach could therefore be

exploited in future to identify more potent inhibitors and use them, in conjunction with bacterial viability and mode of action studies, to validate HPPK as a target for treating *S. enterica* infections.

Bibliography

Aijuka, M. *et al.* (2018) 'Enteroaggregative Escherichia coli is the predominant diarrheagenic E. coli pathotype among irrigation water and food sources in South Africa', *International Journal of Food Microbiology*, 278(2017), pp. 44–51. doi: 10.1016/j.ijfoodmicro.2018.04.018.

Amamuddy, O. S. *et al.* (2020) 'Integrated computational approaches and tools for allosteric drug discovery', *International Journal of Molecular Sciences*. doi: 10.3390/ijms21030847.

Assenberg, R. *et al.* (2013) 'Advances in recombinant protein expression for use in pharmaceutical research', *Current Opinion in Structural Biology*, 23(3), pp. 393–402. doi: 10.1016/j.sbi.2013.03.008.

Atmaramani, R., Pancrazio, J. J. and Black, B. J. (2020) 'Adaptation of robust Z' factor for assay quality assessment in microelectrode array based screening using adult dorsal root ganglion neurons', *Journal of Neuroscience Methods*, 339(January), p. 108699. doi: 10.1016/j.jneumeth.2020.108699.

Ayyildiz, M. *et al.* (2020) 'Identification of Alternative Allosteric Sites in Glycolytic Enzymes for Potential Use as Species-Specific Drug Targets', *Frontiers in Molecular Biosciences*, 7(May), pp. 1–19. doi: 10.3389/fmolb.2020.00088.

Barrett, M. P. *et al.* (2011) 'Drug resistance in human African trypanosomiasis', *Future Microbiology*, 6(9), pp. 1037–1047. doi: 10.2217/fmb.11.88.

Barrett, M. P. (2018) 'The elimination of human African trypanosomiasis is in sight: Report from the third WHO stakeholders meeting on elimination of gambiense human African trypanosomiasis', *PLoS Neglected Tropical Diseases*, 12(12), pp. 10–13. doi: 10.1371/journal.pntd.0006925.

Bates, P. A. *et al.* (2001) 'Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM', *Proteins: Structure, Function and Genetics*, 45(March 2001), pp. 39–46. doi: 10.1002/prot.1168.

Bello, A. R. *et al.* (1994) 'PTR1: A reductase mediating salvage of oxidized pteridines and methotrexate resistance in the protozoan parasite Leishmania major', *Proceedings of the National Academy of Sciences of the United States of America*, 91(24), pp. 11442–11446. doi: 10.1073/pnas.91.24.11442.

Bermingham, A. *et al.* (2000) 'Equilibrium and kinetic studies of substrate binding to 6-hydroxymethyl- 7,8-dihydropterin pyrophosphokinase from Escherichia coli', *Journal of Biological Chemistry*, 275(24), pp. 17962–17967. doi: 10.1074/jbc.M000331200.

Bermingham, A. and Derrick, J. P. (2002) 'The folic acid biosynthesis pathway in bacteria: Evaluation of potential for antibacterial drug discovery', *BioEssays*, 24(7), pp. 637–648. doi: 10.1002/bies.10114.

Bertacine Dias, M. V. *et al.* (2018) 'Folate biosynthesis pathway: Mechanisms and insights into drug design for infectious diseases', *Future Medicinal Chemistry*, 10(8), pp. 935–959. doi: 10.4155/fmc-2017-0168.

Bing, X. *et al.* (1999) 'Crystal structure of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase, a potential target for the development of novel antimicrobial agents', *Structure*, 7(5), pp. 489–496. doi: 10.1016/S0969-2126(99)80065-3.

Blaszczyk, J. *et al.* (2004) 'Reaction trajectory of pyrophosphoryl transfer catalyzed by 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase', *Structure*, 12(3), pp. 467–475. doi: 10.1016/j.str.2004.02.003.

Bourne, C. R. (2014) 'Utility of the biosynthetic folate pathway for targets in antimicrobial discovery', *Antibiotics*, pp. 1–28. doi: 10.3390/antibiotics3010001.

Branchu, P., Bawn, M. and Kingsley, R. A. (2018) 'Genome variation and molecular epidemiology of Salmonella enterica serovar Typhimurium pathovariants', *Infection and Immunity*, 86(8). doi: 10.1128/IAI.00079-18.

Braoudaki, M. and Hilton, A. C. (2004) 'Adaptive Resistance to Biocides in Salmonella enterica and Escherichia coli O157 and Cross-Resistance to Antimicrobial Agents', *Journal of Clinical Microbiology*, 42(1), pp. 73–78. doi: 10.1128/JCM.42.1.73-78.2004.

Brun, R. *et al.* (2010) 'Human African trypanosomiasis', *The Lancet*, 375(9709), pp. 148–159. doi: 10.1016/S0140-6736(09)60829-1.

Bukachi, S. A., Wandibba, S. and Nyamongo, I. K. (2017) 'The socio-economic burden of human African trypanosomiasis and the coping strategies of households in the South Western Kenya foci', *PLoS Neglected Tropical Diseases*, 11(10), pp. 1–21. doi: 10.1371/journal.pntd.0006002.

Büscher, P. *et al.* (2017) 'Human African trypanosomiasis', *The Lancet*, 390(10110), pp. 2397–2409. doi: 10.1016/S0140-6736(17)31510-6.

Cavalli, A. *et al.* (2007) 'Protein structure determination from NMR chemical shifts', *Proceedings of the National Academy of Sciences of the United States of America*, 104(23), pp. 9615–9620. doi: 10.1073/pnas.0610313104.

Cavasotto, C. N. and Phatak, S. S. (2009) 'Homology modeling in drug discovery: current trends and applications', *Drug Discovery Today*, pp. 676–683. doi: 10.1016/j.drudis.2009.04.006.

Cavazzuti, A. *et al.* (2008) 'Discovery of potent pteridine reductase inhibitors to guide antiparasite drug development', *Proceedings of the National Academy of Sciences of the United States of America*, 105(5), pp. 1448–1453. doi: 10.1073/pnas.0704384105.

Chen, R. (2012) 'Bacterial expression systems for recombinant protein production: E. coli and beyond', *Biotechnology Advances*, 30(5), pp. 1102–1107. doi: 10.1016/j.biotechadv.2011.09.013.

Cheng, Z. *et al.* (2012) 'Luciferase Reporter Assay System for Deciphering GPCR Pathways', *Current Chemical Genomics*, 4, pp. 84–91. doi: 10.2174/1875397301004010084.

Chhabra, S. *et al.* (2012) 'Structure of S. aureus HPPK and the discovery of a new substrate site inhibitor', *PLoS ONE*, 7(1). doi: 10.1371/journal.pone.0029444.

Chhabra, S. *et al.* (2013) 'Exploring the Chemical Space around 8-Mercaptoguanine as a Route to New Inhibitors of the Folate Biosynthesis Enzyme HPPK', *PLoS ONE*, 8(4). doi: 10.1371/journal.pone.0059535.

Chin, J. W. *et al.* (2009) 'Analysis of NADPH supply during xylitol production by engineered escherichia coli', *Biotechnology and Bioengineering*, 102(1), pp. 209–220. doi: 10.1002/bit.22060.

Corporation, P. (no date) 'AMP-Glo™ Assay INSTRUCTIONS FOR USE OF PRODUCTS V5011, V5012 AND V5013'. Available at: www.promega.com.

Cruz-Cruz, I. *et al.* (2015) 'Visible Luminescence of Dedoped DBU-Treated PEDOT:PSS Films', *Journal of Physical Chemistry C*, 119(33), pp. 19305–19311. doi: 10.1021/acs.jpcc.5b04016.

Cullen, D. and Mocerino, M. (2017) 'A Brief Review of Drug Discovery Research for Human African Trypanosomiasis', *Current Medicinal Chemistry*, 24(7), pp. 701–717. doi:

10.2174/092986732466617012016003 4.

Cullia, G. *et al.* (2018) 'Folates in Trypanosoma brucei: Achievements and Opportunities', *ChemMedChem*, pp. 2150–2158. doi: 10.1002/cmdc.201800500.

Cutler, P. (2003) *Protein Purification Protocols*. second, *Methods in molecular biology*. second. Edited by P. Cutler. New Jersey: Humana Press. doi: 10.1385/159259655X.

Davies, T. G. *et al.* (2002) 'Structure-based design of a potent purine-based cyclin-dependent kinase inhibitor', *Nature Structural Biology*, 9(10), pp. 745–749. doi: 10.1038/nsb842.

Dawson, A. *et al.* (2006) 'Structure and reactivity of Trypanosoma brucei pteridine reductase: Inhibition by the archetypal antifolate methotrexate', *Molecular Microbiology*, 61(6), pp. 1457–1468. doi: 10.1111/j.1365-2958.2006.05332.x.

De Cesare, A. (2018) 'Salmonella in Foods: A Reemerging Problem', in *Advances in Food and Nutrition Research*. 1st edn. Elsevier Inc., pp. 137–179. doi: 10.1016/bs.afnr.2018.02.007.

de Koning, H. P. (2020) 'The drugs of sleeping sickness: Their mechanisms of action and resistance, and a brief history', *Tropical Medicine and Infectious Disease*, 5(1), pp. 1–23. doi: 10.3390/tropicalmed5010014.

De Lucia, A. *et al.* (2018) 'Role of wild birds and environmental contamination in the epidemiology of Salmonella infection in an outdoor pig farm', *Veterinary Microbiology*, 227, pp. 148–154. doi: 10.1016/j.vetmic.2018.11.003.

Dennis, M. L. *et al.* (2014) 'Structure-based design and development of functionalized mercaptoguanine derivatives as inhibitors of the folate biosynthesis pathway enzyme 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase from staphylococcus aureus', *Journal of Medicinal Chemistry*, 57(22), pp. 9612–9626. doi: 10.1021/jm501417f.

Dennis, M. L. *et al.* (2016) 'Structural Basis for the Selective Binding of Inhibitors to 6-Hydroxymethyl-7,8-dihydropterin Pyrophosphokinase from Staphylococcus aureus and Escherichia coli', *Journal of Medicinal Chemistry*, 59(11), pp. 5248–5263. doi: 10.1021/acs.jmedchem.6b00002.

Dennis, M. L. *et al.* (2018) '8-Mercaptoguanine Derivatives as Inhibitors of Dihydropteroate Synthase', *Chemistry - A European Journal*, 24(8), pp. 1922–1930. doi: 10.1002/chem.201704730.

Dickie, E. A. *et al.* (2020) 'New drugs for human African trypanosomiasis: A twenty first century success story', *Tropical Medicine and Infectious Disease*, 5(1), pp. 1–15. doi: 10.3390/tropicalmed5010029.

Di Pisa, F. *et al.* (2017) 'Chroman-4-one derivatives targeting pteridine reductase 1 and showing anti-parasitic activity', *Molecules*, 22(3), pp. 1–16. doi: 10.3390/molecules22030426.

Doak, B. C. *et al.* (2014) 'Oral druggable space beyond the rule of 5: Insights from drugs and clinical candidates', *Chemistry and Biology*, 21(9), pp. 1115–1142. doi: 10.1016/j.chembiol.2014.08.013.

Eng, S. K. *et al.* (2015) 'Salmonella: A review on pathogenesis, epidemiology and antibiotic resistance', *Frontiers in Life Science*, 8(3), pp. 284–293. doi: 10.1080/21553769.2015.1051243.

Esposito, D. and Chatterjee, D. K. (2006) 'Enhancement of soluble protein expression through the use of fusion tags', *Current Opinion in Biotechnology*, 17(4), pp. 353–358. doi: 10.1016/j.copbio.2006.06.003.

Fairlamb, A. H. and Horn, D. (2018) 'Melarsoprol Resistance in African Trypanosomiasis', *Trends in*

Parasitology, 34(6), pp. 481–492. doi: 10.1016/j.pt.2018.04.002.

Fèvre, E. M. *et al.* (2008) 'The burden of human African Trypanosomiasis', *PLoS Neglected Tropical Diseases*, 2(12). doi: 10.1371/journal.pntd.0000333.

Franco, J., Scarone, L. and Comini, M. A. (2018) *Drugs and Drug Resistance in African and American Trypanosomiasis*. 1st edn, *Annual Reports in Medicinal Chemistry*. 1st edn. Elsevier Inc. doi: 10.1016/bs.armc.2018.08.003.

Gao, K. *et al.* (2015) 'Molecular dynamics simulations of the Escherichia coli HPPK apo-enzyme reveal a network of conformational transitions', *Biochemistry*, 54(44), pp. 6734–6742. doi: 10.1021/acs.biochem.5b01012.

Gershell, L. J. and Atkins, J. H. (2003) 'A brief history of novel drug discovery technologies', *Nature Reviews Drug Discovery*, 2(4), pp. 321–327. doi: 10.1038/nrd1064.

Gräslund, S. *et al.* (2008) 'Protein production and purification', *Nature Methods*, 5(2), pp. 135–146. doi: 10.1038/nmeth.f.202.

Green, D. V. S. (2003) 'Virtual Screening of Virtual Libraries', *Progress in Medicinal Chemistry*, pp. 61–97. doi: 10.1016/S0079-6468(02)41002-8.

Grover, A. K. (2013) 'Use of allosteric targets in the discovery of safer drugs', *Medical Principles and Practice*, pp. 418–426. doi: 10.1159/000350417.

Han, K. Y. *et al.* (2007) 'Enhanced solubility of heterologous proteins by fusion expression using stress-induced Escherichia coli protein, Tsf', *FEMS Microbiology Letters*, 274(1), pp. 132–138. doi: 10.1111/j.1574-6968.2007.00824.x.

Hassan, M. D., Castanha, R. C. G. and Wolfram, D. (2020) 'Scientometric analysis of global trypanosomiasis research: 1988–2017', *Journal of Infection and Public Health*, 13(4), pp. 514–520. doi: 10.1016/j.jiph.2019.10.006.

Herrmann, T., Güntert, P. and Wüthrich, K. (2002) 'Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA', *Journal of Molecular Biology*, 319(1), pp. 209–227. doi: 10.1016/S0022-2836(02)00241-3.

Hillisch, A., Pineda, L. F. and Hilgenfeld, R. (2004) 'Utility of homology models in the drug discovery process', *Drug Discovery Today*, pp. 659–669. doi: 10.1016/S1359-6446(04)03196-4.

Hoffmann, A. and Roeder, R. G. (1991) 'Purification of his-tagged proteins in non-denaturing conditions suggests a convenient method for protein interaction studies', *Nucleic Acids Research*, 19(22), pp. 6337–6338. doi: 10.1093/nar/19.22.6337.

Hughes, R. A., Miklos, A. E. and Ellington, A. D. (2011) *Gene synthesis: Methods and applications*, *Methods in Enzymology*. doi: 10.1016/B978-0-12-385120-8.00012-7.

Hughey, R. and Krogh, A. (1996) 'Hidden markov models for sequence analysis: Extension and analysis of the basic method', *Bioinformatics*, 12(2), pp. 95–107. doi: 10.1093/bioinformatics/12.2.95.

Ingle, D. J. *et al.* (2018) 'Dynamics of antimicrobial resistance in intestinal Escherichia coli from children in community settings in South Asia and sub-Saharan Africa', *Nature Microbiology*, 3(9), pp. 1063–1073. doi: 10.1038/s41564-018-0217-4.

Irague, R. *et al.* (2018) 'A generic HTS assay for kinase screening: Validation for the isolation of an engineered malate kinase', *PLoS ONE*, 13(2), pp. 1–15. doi: 10.1371/journal.pone.0193036.

Irek, E. O. *et al.* (2018) 'A systematic review of healthcare-associated infections in Africa: An antimicrobial resistance perspective', *African Journal of Laboratory Medicine*. doi: 10.4102/ajlm.v7i2.796.

Iversen, P. W. *et al.* (2006) 'A comparison of assay performance measures in screening assays: Signal window, Z' factor, and assay variability ratio', *Journal of Biomolecular Screening*, 11(3), pp. 247–252. doi: 10.1177/1087057105285610.

Jaja, I. F. *et al.* (2019) 'Molecular characterisation of antibiotic-resistant Salmonella enterica isolates recovered from meat in South Africa', *Acta Tropica*, 190, pp. 129–136. doi: 10.1016/j.actatropica.2018.11.003.

Ji, X., Shi, G. and Shaw, G. X. (2019) 'HPPK INHIBITORS USEFUL AS ANTIBACTERIAL AGENTS'. United States of America.

Joe, J. (2005) 'Assessing predictions of protein – protein interaction : The CAPRI experiment', *Protein Science*, 14(2), pp. 278–283. doi: 10.1110/ps.041081905.interacting.

Kaljurand, I. *et al.* (2000) 'Self-consistent spectrophotometric basicity scale in acetonitrile covering the range between pyridine and DBU', *Journal of Organic Chemistry*, 65(19), pp. 6202–6208. doi: 10.1021/jo005521j.

Kazemi, B. *et al.* (2010) 'Molecular cloning, expression and enzymatic assay of pteridine reductase 1 from Iranian lizard Leishmania', *Iranian Biomedical Journal*, 14(3), pp. 97–102.

Keddy, K. H. *et al.* (2017) 'An association between decreasing incidence of invasive non-typhoidal salmonellosis and increased use of antiretroviral therapy, Gauteng Province, South Africa, 2003-2013', *PLoS ONE*, pp. 2003–2013. doi: 10.1371/journal.pone.0173091.

Kennedy, P. G. E. (2004) 'Human African trypanosomiasis of the CNS: current issues and challenges', *Journal of Clinical Investigation*, 113(4), pp. 496–504. doi: 10.1172/jci21052.

Kennedy, P. G. E. (2019) 'Update on human African trypanosomiasis (sleeping sickness)', *Journal of Neurology*, 266(9), pp. 2334–2337. doi: 10.1007/s00415-019-09425-7.

Kim, H. J. and Jang, S. (2018) 'Optimization of a resazurin-based microplate assay for large-scale compound screenings against Klebsiella pneumoniae', *3 Biotech*, 8(1). doi: 10.1007/s13205-017-1034-9.

Kimuda, M. P. *et al.* (2019) 'Identification of Novel Potential Inhibitors of Pteridine Reductase 1 in Trypanosoma brucei via Computational Structure-Based Approaches and in Vitro Inhibition Assays', *Molecules*, 24(1), pp. 1–25. doi: 10.3390/molecules24010142.

Koebel, M. R. *et al.* (2016) 'AutoDock VinaXB: Implementation of XBSF, new empirical halogen bond scoring function, into AutoDock Vina', *Journal of Cheminformatics*, 8(1), pp. 1–8. doi: 10.1186/s13321-016-0139-1.

Kok, D. E., Steegenga, W. T. and McKay, J. A. (2018) 'Folate and epigenetics: Why we should not forget bacterial biosynthesis', *Epigenomics*, 10(9), pp. 1147–1150. doi: 10.2217/epi-2018-0117.

Kumar, K. *et al.* (2018) 'A rapid and simple resazurin assay to detect minimum inhibitory concentrations of first-line drugs for Mycobacterium tuberculosis isolated from cerebrospinal fluid',

Journal of Global Antimicrobial Resistance, 12(2010), pp. 157–161. doi: 10.1016/j.jgar.2017.09.012.

Kurczab, R., Kucwaj-Brysz, K. and Śliwa, P. (2020) 'The significance of halogen bonding in ligand-receptor interactions: The lesson learned from molecular dynamic simulations of the D4 receptor', *Molecules*, 25(1). doi: 10.3390/molecules25010091.

Kurt, N. *et al.* (2003) 'Cooperative fluctuations of unliganded and substrate-bound HIV-1 protease: A structure-based analysis on a variety of conformations from crystallography and molecular dynamics simulations', *Proteins: Structure, Function and Genetics*, 51(3), pp. 409–422. doi: 10.1002/prot.10350.

Landi, G. *et al.* (2020) 'High-resolution crystal structure of Trypanosoma brucei pteridine reductase 1 in complex with an innovative tricyclic-based inhibitor', *Acta Crystallographica Section D: Structural Biology*, 76(2020), pp. 558–564. doi: 10.1107/S2059798320004891.

Langridge, G. C., Wain, J. and Nair, S. (2012) 'Invasive Salmonellosis in Humans', *EcoSal Plus*, 5(1). doi: 10.1128/ecosalplus.8.6.2.2.

Lenk, E. J. *et al.* (2018) 'Socioeconomic benefit to individuals of achieving 2020 targets for four neglected tropical diseases controlled/eliminated by innovative and intensified disease management: Human African trypanosomiasis, leprosy, visceral leishmaniasis, Chagas disease', *PLoS Neglected Tropical Diseases*, 12(3), pp. 1–28. doi: 10.1371/journal.pntd.0006250.

Lesley, S. A. (2001) 'High-throughput proteomics: Protein expression and purification in the postgenomic world', *Protein Expression and Purification*, 22(2), pp. 159–164. doi: 10.1006/prep.2001.1465.

Li, G. *et al.* (2006) 'Mechanism of the conformational transitions in 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase as revealed by NMR spectroscopy', *Biochemistry*, 45(41), pp. 12573–12581. doi: 10.1021/bi061057m.

Linciano, P. *et al.* (2017) 'Exploiting the 2-Amino-1,3,4-thiadiazole scaffold to inhibit trypanosoma brucei pteridine reductase in support of early-stage drug discovery', *ACS Omega*, 2(9), pp. 5666–5683. doi: 10.1021/acsomega.7b00473.

Linciano, P. *et al.* (2019) 'Enhancement of Benzothiazoles as Pteridine Reductase-1 Inhibitors for the Treatment of Trypanosomatidic Infections', *Journal of Medicinal Chemistry*, 62(8), pp. 3989–4012. doi: 10.1021/acs.jmedchem.8b02021.

Linciano, P. *et al.* (2020) 'Identification of a 2,4-diaminopyrimidine scaffold targeting Trypanosoma brucei pteridine reductase 1 from the LIBRA compound library screening campaign', *European Journal of Medicinal Chemistry*, 189. doi: 10.1016/j.ejmech.2020.112047.

Lindner, A. K. *et al.* (2020) 'New WHO guidelines for treatment of gambiense human African trypanosomiasis including fexinidazole: substantial changes for clinical practice', *The Lancet Infectious Diseases*, 20(2), pp. e38–e46. doi: 10.1016/S1473-3099(19)30612-7.

Lipinski, C. A. (2000) 'Drug-like properties and the causes of poor solubility and poor permeability', *Journal of Pharmacological and Toxicological Methods*, 44(1), pp. 235–249. doi: 10.1016/S1056-8719(00)00107-6.

Lupindu, A. M. (2018) 'Epidemiology of Shiga toxin-producing Escherichia coli O157:H7 in Africa in review', *Southern African Journal of Infectious Diseases*, 33(1), pp. 24–30. doi: 10.1080/23120053.2017.1376558.

Luthy, R., Bowei, J. and Einsenberg, D. (1997) 'Verify3D: Assessment of protein models with three-dimensional profiles', in *Methods in enzymology*, pp. 396–404.

Lutumba, P. *et al.* (2007) 'Human African trypanosomiasis in a rural community, Democratic Republic of Congo', *Emerging Infectious Diseases*, 13(2), pp. 248–254. doi: 10.3201/eid1302.060075.

Lynn, M. K. A. G. *et al.* (1998) 'Typhimurium Dt104 Infections in the United States', *The New England Journal of Medicine*, 338(19), pp. 1333–1338.

MacDonald, E. *et al.* (2019) 'The role of domestic reservoirs in domestically acquired Salmonella infections in Norway: Epidemiology of salmonellosis, 2000-2015, and results of a national prospective case-control study, 2010-2012', *Epidemiology and Infection*. doi: 10.1017/S0950268818002911.

Malande, O. O. *et al.* (2019) 'A ten-year review of ESBL and non-ESBL Escherichia coli bloodstream infections among children at a tertiary referral hospital in South Africa', *PLoS ONE*, 14(9), pp. 1–16. doi: 10.1371/journal.pone.0222675.

Mäntele, W. and Deniz, E. (2017) 'UV–VIS absorption spectroscopy: Lambert-Beer reloaded', *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 173, pp. 965–968. doi: 10.1016/j.saa.2016.09.037.

Marimuthu, P., Singaravelu, K. and Namasivayam, V. (2017) 'Probing the binding mechanism of mercaptoguanine derivatives as inhibitors of HPPK by docking and molecular dynamics simulations', *Journal of Biomolecular Structure and Dynamics*, 35(16), pp. 3507–3521. doi: 10.1080/07391102.2016.1260496.

Marsic, D. *et al.* (2008) 'PCR-based gene synthesis to produce recombinant proteins for crystallization', *BMC Biotechnology*, 8, pp. 1–13. doi: 10.1186/1472-6750-8-44.

Mbelle, N. M. *et al.* (2019) 'The Resistome, Mobilome, Virulome and Phylogenomics of Multidrug-Resistant Escherichia coli Clinical Isolates from Pretoria, South Africa', *Scientific Reports*, 9(1), pp. 1–16. doi: 10.1038/s41598-019-52859-2.

Mehlitz, D. and Molyneux, D. H. (2019) 'The elimination of Trypanosoma brucei gambiense? Challenges of reservoir hosts and transmission cycles: Expect the unexpected', *Parasite Epidemiology and Control*, 6(2019), p. e00113. doi: 10.1016/j.parepi.2019.e00113.

Montelione, G. T. *et al.* (2000) 'Protein NMR spectroscopy in structural genomics Protein NMR spectroscopy provides an important complement to X-ray crystallography for structural genomics, both for determining three-dimensional protein structures and in characterizing their biochemical a', (november). Available at: http://structbio.nature.com.

Montso, K. P. *et al.* (2019) 'Antimicrobial Resistance Factors of Extended-Spectrum Beta-Lactamases Producing Escherichia coli and Klebsiella pneumoniae Isolated from Cattle Farms and Raw Beef in North-West Province, South Africa', *BioMed Research International*, 2019. doi: 10.1155/2019/4318306.

Morris, G. M. and Lim-Wilby, M. (2006) 'Molecular Docking', in *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine*, pp. 1149–1153. doi: 10.1007/3-540-29623-9_3820.

Nallamsetty, S. and Waugh, D. S. (2007) 'A generic protocol for the expression and purification of recombinant proteins in Escherichia coli using a combinatorial His6-maltose binding protein fusion tag', *Nature Protocols*, 2(2), pp. 383–391. doi: 10.1038/nprot.2007.50.

Nare, B., Hardy, L. W. and Beverley, S. M. (1997) 'The roles of pteridine reductase 1 and dihydrofolate reductase- thymidylate synthase in pteridine metabolism in the protozoan parasite Leishmania major', *Journal of Biological Chemistry*, 272(21), pp. 13883–13891. doi: 10.1074/jbc.272.21.13883.

Nguyen, H. *et al.* (2004) 'An automated small-scale protein expression and purification screening provides beneficial information for protein production', *Journal of Structural and Functional Genomics*, 5(1–2), pp. 23–27. doi: 10.1023/B:JSFG.0000029195.73810.86.

Nguyen, N. T. *et al.* (2020) 'Autodock Vina Adopts More Accurate Binding Poses but Autodock4 Forms Better Binding Affinity', *Journal of Chemical Information and Modeling*, 60(1), pp. 204–211. doi: 10.1021/acs.jcim.9b00778.

O'Neill, T. E. *et al.* (2014) 'Optimisation of the microplate resazurin assay for screening and bioassay-guided fractionation of phytochemical extracts against mycobacterium tuberculosis', *Phytochemical Analysis*, 25(5), pp. 461–467. doi: 10.1002/pca.2516.

Olobatoke, R. Y. (2017) 'Public health burden of non-typhoidal Salmonella strains in sub-Saharan Africa', *International Research Journal of Public and Environmental Health*, 4(6), pp. 112–119. Available at: https://journalissues.org/wp-content/uploads/2017/07/Olobatoke.pdf.

Omolabi, K. F. *et al.* (2020) 'Could chroman-4-one derivative be a better inhibitor of PTR1? – Reason for the identified disparity in its inhibitory potency in Trypanosoma brucei and Leishmania major', *Computational Biology and Chemistry*, p. 107412. doi: 10.1016/j.compbiolchem.2020.107412.

Østergaard, J. (2016) 'UV/Vis Spectrophotometry and UV Imaging', pp. 3–27. doi: 10.1007/978-1-4939-4029-5_1.

Pagadala, N. S., Syed, K. and Tuszynski, J. (2017) 'Software for molecular docking: a review', *Biophysical Reviews*. Biophysical Reviews, pp. 91–102. doi: 10.1007/s12551-016-0247-1.

Panecka-Hofman, J. *et al.* (2017) 'Comparative mapping of on-targets and off-targets for the discovery of anti-trypanosomatid folate pathway inhibitors', *Biochimica et Biophysica Acta - General Subjects*, 1861(12), pp. 3215–3230. doi: 10.1016/j.bbagen.2017.09.012.

Park, S. E. *et al.* (2019) 'The Severe Typhoid Fever in Africa Program: Study Design and Methodology to Assess Disease Severity, Host Immunity, and Carriage Associated with Invasive Salmonellosis', *Clinical Infectious Diseases*, 69(Suppl 6), pp. S422–S434. doi: 10.1093/cid/ciz715.

Parkhill, J. *et al.* (2001) 'Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18', *Nature*, 413(6858), pp. 848–852. doi: 10.1038/35101607.

Pellizza, L. *et al.* (2018) 'Codon usage clusters correlation: Towards protein solubility prediction in heterologous expression systems in E. coli', *Scientific Reports*, 8(1), pp. 1–12. doi: 10.1038/s41598-018-29035-z.

Pöhner, I. *et al.* (2020) 'Multitarget, selective compound design yields picomolar inhibitors of a kinetoplastid pteridine reductase 1', *ChemRxiv*, pp. 1–34. doi: 10.26434/chemrxiv.13026797.v1.

Poirel, L. *et al.* (2018) 'Antimicrobial Resistance in Escherichia coli', in *Antimicrobial Resistance in Bacteria from Livestock and Companion Animals*, pp. 289–316. doi: 10.1128/microbiolspec.arba-0026-2017.

Ponte-Sucre, A. (2016) 'An overview of trypanosoma brucei infections: An intense host-parasite interaction', *Frontiers in Microbiology*, 7(DEC), pp. 1–12. doi: 10.3389/fmicb.2016.02126.

Pozzan, A. (2006) 'Molecular Descriptors and Methods for Ligand Based Virtual High Throughput Screening in Drug Discovery', *Current Pharmaceutical Design*, 12(17), pp. 2099–2110. doi: 10.2174/138161206777585247.

Präbst, K. *et al.* (2017) 'Chapter 2 of Cell Viability Assays', *Basic Colorimetric Proliferation Assays: MTT, WST, and Resazurin*, 1601, pp. 1–17. doi: 10.1007/978-1-4939-6960-9.

Promega corporation (2015) 'Kinase-Glo® Luminescent Kinase Assay Platform Technical Bulletin #TB372', p. 22.

Radoshitzky, S. R. *et al.* (2012) 'Drug discovery technologies and strategies for Machupo virus and other New World arenaviruses', *Expert Opinion on Drug Discovery*, 7(7), pp. 613–632. doi: 10.1517/17460441.2012.687719.

Reeve, S. M. *et al.* (2016) 'Charged Propargyl-Linked Antifolates Reveal Mechanisms of Antifolate Resistance and Inhibit Trimethoprim-Resistant MRSA Strains Possessing Clinically Relevant Mutations', *Journal of Medicinal Chemistry*, 59(13), pp. 6493–6500. doi: 10.1021/acs.jmedchem.6b00688.

Roussel, C. *et al.* (2017) 'Enterotoxigenic and Enterohemorrhagic *Escherichia coli*: Survival and Modulation of Virulence in the Human Gastrointestinal Tract', Escherichia coli - *Recent Advances on Physiology, Pathogenesis and Biotechnological Applications*, pp. 3–24. doi: 10.5772/intechopen.68309.

Sánchez-Vargas, F. M., Abu-El-Haija, M. A. and Gómez-Duarte, O. G. (2011) 'Salmonella infections: An update on epidemiology, management, and prevention', *Travel Medicine and Infectious Disease*, pp. 263–277. doi: 10.1016/j.tmaid.2011.11.001.

Sanders, W. J. *et al.* (2004) 'Discovery of Potent Inhibitors of Dihydroneopterin Aldolase Using CrystaLEAD High-Throughput X-ray Crystallographic Screening and Structure-Directed Lead Optimization', *Journal of Medicinal Chemistry*, 47(7), pp. 1709–1718. doi: 10.1021/jm030497y.

Seeliger, D. and De Groot, B. L. (2010) 'Ligand docking and binding site analysis with PyMOL and Autodock/Vina', *Journal of Computer-Aided Molecular Design*, 24(5), pp. 417–422. doi: 10.1007/s10822-010-9352-6.

Shanks, E. J. *et al.* (2010) 'Development and validation of a cytochrome c-coupled assay for pteridine reductase 1 and dihydrofolate reductase', *Analytical Biochemistry*, 396(2), pp. 194–203. doi: 10.1016/j.ab.2009.09.003.

Shaw, G. X. *et al.* (2014) 'Structural enzymology and inhibition of the bi-functional folate pathway enzyme HPPK-DHPS from the biowarfare agent Francisella tularensis', *The FEBS journal*, 281(18), pp. 4123–4137. doi: 10.1111/febs.12896.

Shi, G. *et al.* (2001) 'Bisubstrate analogue inhibitors of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase: Synthesis and biochemical and crystallographic studies', *Journal of Medicinal Chemistry*, 44(9), pp. 1364–1371. doi: 10.1021/jm0004493.

Shi, G., Shaw, G., Li, Y., *et al.* (2012) 'Bisubstrate analog inhibitors of 6-hydroxymethyl-7 , 8-dihydropterin pyrophosphokinase : New lead exhibits a distinct binding mode', *Bioorganic & Medicinal Chemistry*, 20(14), pp. 4303–4309. doi: 10.1016/j.bmc.2012.05.060.

Shi, G., Shaw, G., Liang, Y. H., *et al.* (2012) 'Bisubstrate analogue inhibitors of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase: New design with improved properties', *Bioorganic and Medicinal Chemistry*, 20(1), pp. 47–57. doi: 10.1016/j.bmc.2011.11.032.

Sienkiewicz, N., Ong, H. B. and Fairlamb, A. H. (2010) 'Trypanosoma brucei pteridine reductase 1 is essential for survival in vitro and for virulence in mice', *Molecular Microbiology*, 77(3), pp. 658–671. doi: 10.1111/j.1365-2958.2010.07236.x.

Smith, S. I., Seriki, A. and Ajayi, A. (2016) 'Typhoidal and non-typhoidal Salmonella infections in Africa', *European Journal of Clinical Microbiology and Infectious Diseases*. European Journal of Clinical Microbiology & Infectious Diseases, pp. 1913–1922. doi: 10.1007/s10096-016-2760-3.

Sørensen, H. P. and Mortensen, K. K. (2005) 'Advanced genetic strategies for recombinant protein expression in Escherichia coli', *Journal of Biotechnology*, 115(2), pp. 113–128. doi: 10.1016/j.jbiotec.2004.08.004.

Spaulding, A., Gallerstein, M. F. and Ferrins, L. (2019) 'Drug Discovery and Development for Human African Trypanosomiasis', pp. 115–137. doi: 10.1002/9783527808656.ch5.

Spellberg, B. *et al.* (2008) 'The epidemic of antibiotic-resistant infections: A call to action for the medical community from the infectious diseases society of America', *Clinical Infectious Diseases*, 46(2), pp. 155–164. doi: 10.1086/524891.

Spriestersbach, A. *et al.* (2015) 'Purification of His-Tagged Proteins', *Methods in Enzymology*, 559, pp. 1–15. doi: 10.1016/bs.mie.2014.11.003.

Su, L. and Cukier, R. I. (2009) 'An enhanced molecular dynamics study of HPPK-ATP conformation space exploration and ATP binding to HPPK', *Journal of Physical Chemistry A*, 113(10), pp. 2025–2035. doi: 10.1021/jp808664k.

Swarbrick, J. *et al.* (2009) 'Folate biosynthesis - reappraisal of old and novel targets in the search for new antimicrobials', *New Developments in Medicinal Chemistry*, (June), pp. 1–40. doi: 10.2174/1874940200801010012.

Szilagyi, A. and Zhang, Y. (2014) 'Template-based structure modeling of protein-protein interactions', *Current Opinion in Structural Biology*. Elsevier Ltd, pp. 10–23. doi: 10.1016/j.sbi.2013.11.005.

Tack, B. *et al.* (2020) 'Invasive non-typhoidal Salmonella infections in sub-Saharan Africa: A systematic review on antimicrobial resistance and treatment', *BMC Medicine*. BMC Medicine, pp. 1–22. doi: 10.1186/s12916-020-01652-4.

Taneja, N. K. and Tyagi, J. S. (2007) 'Resazurin reduction assays for screening of anti-tubercular compounds against dormant and actively growing Mycobacterium tuberculosis, Mycobacterium bovis BCG and Mycobacterium smegmatis', *Journal of Antimicrobial Chemotherapy*, 60(2), pp. 288–293. doi: 10.1093/jac/dkm207.

Tanner, J. R. and Kingsley, R. A. (2018) 'Evolution of Salmonella within Hosts', *Trends in Microbiology*, 26(12), pp. 986–998. doi: 10.1016/j.tim.2018.06.001.

Tate, C. G. *et al.* (2003) 'Comparison of seven different heterologous protein expression systems for the production of the serotonin transporter', *Biochimica et Biophysica Acta - Biomembranes*, 1610(1), pp. 141–153. doi: 10.1016/S0005-2736(02)00719-8.

Teixeira, B. V. F. *et al.* (2019) 'Dual and selective inhibitors of pteridine reductase 1 (PTR1) and dihydrofolate reductase-thymidylate synthase (DHFR-TS) from Leishmania chagasi', *Journal of Enzyme Inhibition and Medicinal Chemistry*, 34(1), pp. 1439–1450. doi: 10.1080/14756366.2019.1651311.

Temiz, N. A. and Bahar, I. (2002) 'Inhibitor binding alters the directions of domain motions in HIV-1

reverse transcriptase', *Proteins: Structure, Function and Genetics*, 49(1), pp. 61–70. doi: 10.1002/prot.10183.

Terpe, K. (2006) 'Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems', *Applied Microbiology and Biotechnology*, 72(2), pp. 211–222. doi: 10.1007/s00253-006-0465-8.

Tramontano, A. (2017) 'The computational prediction of protein assemblies', *Current Opinion in Structural Biology*. Elsevier Ltd, pp. 170–175. doi: 10.1016/j.sbi.2017.10.006.

Trimpin, S. and Brizzard, B. (2009) 'Analysis of insoluble proteins', *BioTechniques*, 46(5 SPEC. ISSUE), pp. 321–326. doi: 10.2144/000113135.

Tuem, K. B. *et al.* (2018) 'Drug Resistance Patterns of Escherichia coli in Ethiopia: A Meta-Analysis', *BioMed Research International*. doi: 10.1155/2018/4536905.

Tyagi, M. *et al.* (2020) 'Drug Syntheses Beyond the Rule of 5', *Chemistry - A European Journal*, pp. 49–88. doi: 10.1002/chem.201902716.

Valones, M. A. A. *et al.* (2009) 'Principles and applications of polymerase chain reaction in medical diagnostic fields: A review', *Brazilian Journal of Microbiology*, 40(1), pp. 1–11. doi: 10.1590/S1517-83822009000100001.

van den Brulle, J. *et al.* (2008) 'A novel solid phase technology for high-throughput gene synthesis', *BioTechniques*, 45(3), pp. 340–343. doi: 10.2144/000112953.

Ventura, S. (2016) 'Editorial: Protein Solubility and Aggregation in Bacteria', *Frontiers in Microbiology*, 7(JUL), pp. 1–3. doi: 10.3389/fmicb.2016.01178.

Visser, B. J. (2020) 'Human African trypanosomiasis A neglected disease', 33(1), p. 2020.

Wang, Z. *et al.* (2016) 'Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: The prediction accuracy of sampling power and scoring power', *Physical Chemistry Chemical Physics*, 18(18), pp. 12964–12975. doi: 10.1039/c6cp01555g.

Wenthur, C. J. *et al.* (2014) 'Drugs for allosteric sites on receptors', *Annual Review of Pharmacology and Toxicology*, pp. 165–184. doi: 10.1146/annurev-pharmtox-010611-134525.

White, S., Anandraj, A. and Trois, C. (2014) 'NADPH fluorescence as an indicator of hydrogen production in the green algae Chlamydomonas reinhardtii', *International Journal of Hydrogen Energy*, 39(4), pp. 1640–1647. doi: 10.1016/j.ijhydene.2013.11.016.

Withers-Martinez, C. *et al.* (1999) 'PCR-based gene synthesis as an efficient approach for expression of the A+T-rich malaria genome', *Protein Engineering*, 12(12), pp. 1113–1120. doi: 10.1093/protein/12.12.1113.

Yamazaki, T. *et al.* (1998) 'Segmental isotope labeling for protein NMR using peptide splicing', *Journal of the American Chemical Society*, 120(22), pp. 5591–5592. doi: 10.1021/ja980776o.

Yang, R. *et al.* (2005) 'Loop conformation and dynamics of the Escherichia coli HPPK apo-enzyme and its binary complex with MgATP', *Biophysical Journal*, 89(1), pp. 95–106. doi: 10.1529/biophysj.105.061556.

Yun, M. K. *et al.* (2014) 'The identification, analysis and structure-based development of novel inhibitors of 6-hydroxymethyl-7,8-dihydropterin pyrophosphokinase', *Bioorganic and Medicinal Chemistry*, 22(7), pp. 2157–2165. doi: 10.1016/j.bmc.2014.02.022.

Zhang, J. H., Chung, T. D. Y. and Oldenburg, K. R. (1999) 'A simple statistical parameter for use in evaluation and validation of high throughput screening assays', *Journal of Biomolecular Screening*, 4(2), pp. 67–73. doi: 10.1177/108705719900400206.

Zoltner, M. *et al.* (2020) 'Suramin exposure alters cellular metabolism and mitochondrial energy production in African trypanosomes', *Journal of Biological Chemistry*, 295(24), pp. 8331–8347. doi: 10.1074/jbc.RA120.012355.

Appendices

Appendix A: Amino acid sequence alignment of *S. enterica* and *E. coli* HPPK

Chain identity between *S. enterica* HPPK and *E. coli* HPPK was performed using ClustalW2 sequence alignment software.

```
[Salmonella     MTIAYIALGSNLASPLEQVNAALKAIADIPDSRIVAVSSFYRTPPLGPQDQPDYLNAAVA   60
[Escherichia    MTLVYIALGSNLASPLTQVNAAIAALGTLPDSQVVAVSGFYRTPPLGPQDQPDYLNA     60

[Salmonella     LDTALTPEELLNHTQRIERQQGRERKAERWGPRTLDLDIMLFGDEVISTERLTVPHYDMK  120
[Escherichia    LETTLAPEALLDHTQRIELDQGRVRKAERWGPRTLDLDIMLFGDETLATSRLTVPHYDMK  120

[Salmonella     NRGFMLWPLFEIAPELIFPDDGVALAQRVHQFTPVPARW                       159
[Escherichia    NRGFMLWPLFEIAPELRFPDGETLVSVLARLGAAKPASW                       159
```

74 % Similarity between sequences

Appendix B: Protein sequences and Genscript optimisation

***Trypanosoma brucei brucei* Pteridine reductase** (UniProt access number O76290)

Amino acid sequence:

MEAPAAVVTGAAKRIGRAIAVKLHQTGYRVVIHYHNSAEAAVSLADELNKERSNTAVVCQ
ADLTNSNVLPASCEEIINSCFRAFGRCDVLVNNASAFYPTPLVQGDHEDNSNGKTVETQV
AELIGTNAIAPFLLTMSFAQRQKGTNPNCTSSNLSIVNLCDAMVDQPCMAFSLYNMGKHA
LVGLTQSAALELAPYGIRVNGVAPGVSLLPVAMGEEEKDKWRRKVPLGRREASAEQIADA
VIFLVSGSAQYITGSIIKVDGGLSLVHA

Genscript codon optimised (for expression in *E. coli*) coding sequence:

ATGGAAGCGCCGGCGGCGGTGGTTACCGGTGCGGCGAAGCGTATTGGTCGTGCGATTGCGGTGAAACTGCACCAG
ACCGGTTACCGTGTGGTTATCCACTATCACAACAGCGCGGAGGCGGCGGTTAGCCTGGCGGACGAACTGAACAAG
GAGCGTAGCAACACCGCGGTGGTTTGCCAAGCGGATCTGACCAACAGCAACGTGCTGCCGGCGAGCTGCGAGGAA
ATCATTAACAGCTGCTTCCGTGCGTTTGGCCGTTGCGACGTGCTGGTTAACAACGCGAGCGCGTTCTACCCGACC
CCGCTGGTTCAGGGTGACCACGAAGATAACAGCAACGGCAAGACCGTGGAAACCCAAGTTGCGGAGCTGATCGGT
ACCAACGCGATTGCGCCGTTCCTGCTGACCATGAGCTTTGCGCAGCGTCAAAAAGGCACCAACCCGAACTGCACC
AGCAGCAACCTGAGCATCGTGAACCTGTGCGACGCGATGGTTGATCAGCCGTGCATGGCGTTCAGCCTGTACAAC
ATGGGTAAACACGCGCTGGTTGGTCTGACCCAGAGCGCGGCGCTGGAGCTGGCGCCGTATGGTATTCGTGTGAAC
GGTGTTGCGCCGGGCGTGAGCCTGCTGCCGGTTGCGATGGGCGAGGAAGAGAAGGACAAATGGCGTCGTAAGGTT
CCGCTGGGTCGTCGTGAAGCGAGCGCGGAGCAGATTGCGGATGCGGTGATTTTTCTGGTTAGCGGTAGCGCGCAA
TATATTACCGGCAGCATCATTAAAGTGGATGGTGGCCTGAGCCTGGTTCACGCG

TAA stop codon added

Cloned between *Nhe*I and *Xho*I sites of pET28a (+)

***Salmonella enterica* HPPK**

Amino acid sequence:

MTIAYIALGSNLASPLEQVNAALKAIADIPDSRIVAVSSFYRTPPLGPQDQPDYLNAAVALDTALAPEELLNHTQ
RIELQQGRVRKAERWGPRTLDLDIMLFGDEVINTDRLTVPHYDMKNRGFMLWPLFEIAPDLIFPDGISLHQHLTH
LGAAKPAHW

Genscript codon optimised (for expression in *E. coli*) coding sequence:

ATGACCATCGCGTACATTGCGCTGGGTAGCAACCTGGCGA
GCCCGCTGGAGCAGGTGAACGCGGCGCTGAAGGCGATCGCGGACA
TTCCGGATAGCCGTATTGTGGCGGTTAGCAGCTTCTACCGTACCC
CGCCGCTGGGTCCGCAGGACCAACCGGATTATCTGAACGCGGCGG
TGGCGCTGGACACCGCGCTGGCGCCGGAGGAACTGCTGAACCACA
CCCAACGTATCGAGCTGCAGCAAGGTCGTGTTCGTAAAGCGGAAC
GTTGGGGTCCGCGTACCCTGGACCTGGATATTATGCTGTTCGGTG
ACGAGGTGATCAACACCGATCGTCTGACCGTTCCGCACTATGATA
TGAAAAACCGTGGCTTCATGCTGTGGCCGCTGTTTGAAATCGCGC
CGGACCTGATTTTTCCGGATGGTATCAGCCTGCATCAGCACCTGA
CCCACCTGGGTGCGGCGAAACCGGCGCACTGG

TAA stop codon added

Cloned between *Nhe*I and *Xho*I sites of pET28a (+)

***E. coli* HPPK**

Amino acid sequence:

MTVAYIAIGSNLASPLEQVNAALKALGDIPESHILTVSSFYRTPPLGPQDQPDYLNAAVALETSLAPEEL
LNHTQRIELQQGRVRKAERWGPRTLDLDIMLFGNEVINTERLTVPHYDMKNRGFMLWPLFEIAPELVFPD
GEMLRQILHTRAFDKLNKW

Genscript codon optimised (for expression in *E. coli*) coding sequence

ATGACAGTGGCGTATATTGCCATAGGCAGCAATCTGGCCTCTCCGCTGGAGCAGGTCAATGCTGCCCTGAAAGCA
TTAGGCGATATCCCTGAAAGCCACATTCTTACCGTTTCTTCGTTTTACCGCACCCCACCGCTGGGGCCGCAAGAT
CAACCCGATTACTTAAACGCAGCCGTGGCGCTGGAAACCTCTCTTGCACCTGAAGAGCTACTCAATCACACACAG
CGTATTGAATTGCAGCAAGGTCGCGTCCGCAAAGCTGAACGCTGGGGACCACGCACGCTGGATCTCGACATCATG
CTGTTTGGTAATGAAGTGATAAATACTGAACGCCTGACCGTTCCGCACTACGATATGAAGAATCGTGGATTTATG
CTGTGGCCGCTGTTTGAAATCGCGCCGGAGTTGGTGTTTCCTGATGGGGAGATGTTGCGTCAAATCTTACATACA
AGAGCATTTGACAAATTAAACAAATGG

TAA stop codon added

Cloned between *Nhe*I and *Xho*I sites of pET28a (+)

Appendix C: pET 28a (+) plasmid map

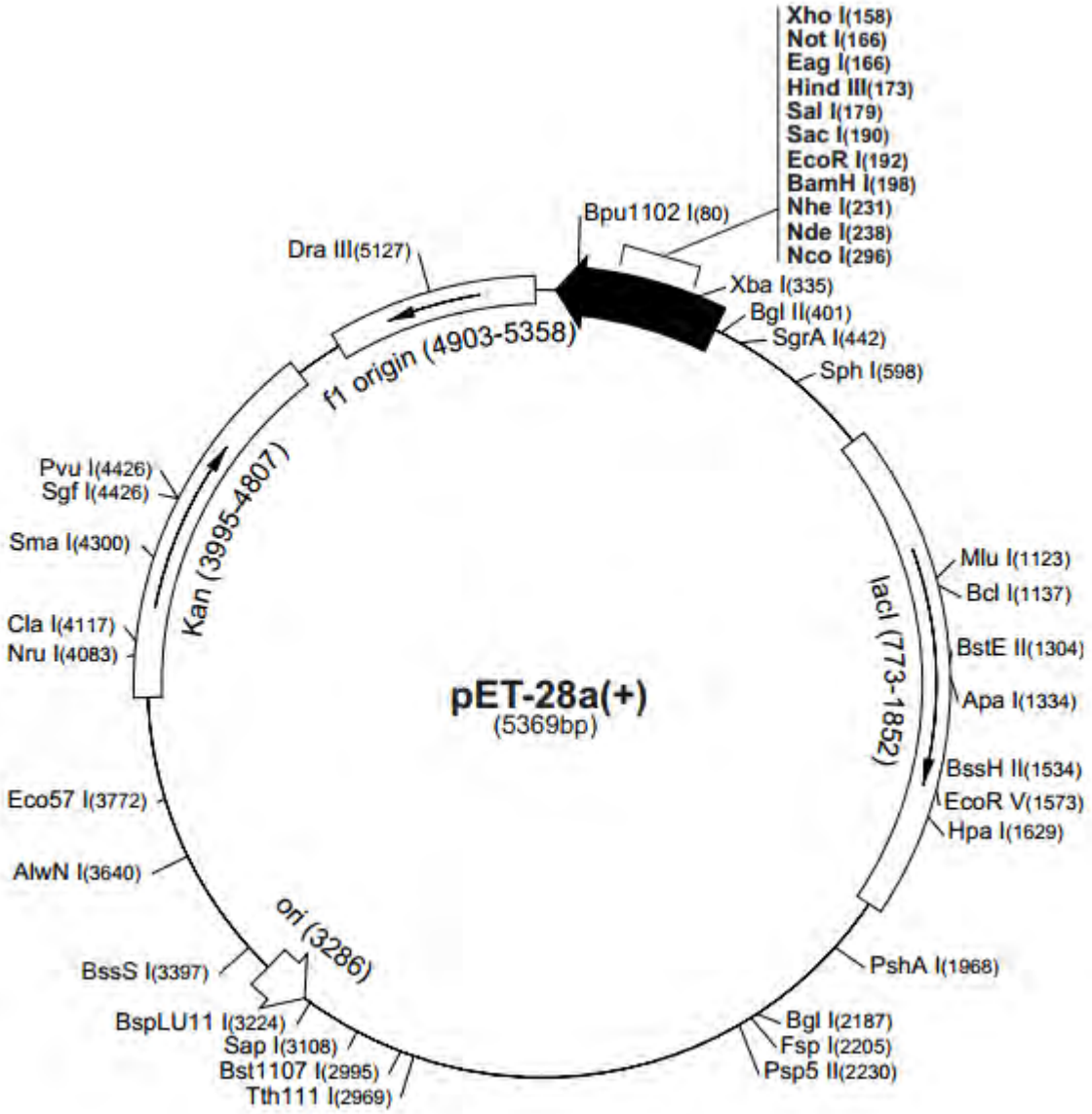


Figure 58: pET 28a (+) plasmid map (obtained from Novagen).

Appendix D: *P. falciparum* Assay

Rhodes University

Centre for Chemico- and Biomedicinal Research

---

pLDH (Malaria) Assay – Single Concentration Screen

Background

Malaria parasites (*Plasmodium falciparum* strain 3D7) were maintained in RPMI 1640 medium containing 2mM L-glutamine and 25mM Hepes (Lonza). The medium is further supplemented with 5% Albumax II, 20 mM glucose, 0.65 mM hypoxanthine, 60 µg/mL gentamycin and 2-4% hematocrit human red blood cells. The parasites were cultured at 37^{o}C under an atmosphere of 5% CO_2, 5% O_2, 90% N_2 in a sealed T75 culture flask.

Single concentration screening was conducted using compounds at 20 µM or 50 µg/ml for natural extracts (unless otherwise stated) added to parasite cultures in 96-well, clear plates and incubated for 48 hours in a 37^{o}C CO_2 incubator. After 48 hours, 20 µL of culture was removed from each well and combined with 125 µL of a mixture of Malstat and NBT/PES solutions in a fresh 96-well plate. These solutions measure the activity of the parasite lactate dehydrogenase (pLDH) enzyme in the cultures. A purple product is formed when pLDH is present, and this product is quantified in a Spectramax M3 microplate reader (Abs_{620}). The Abs_{620} reading in each well is thus an indication of the pLDH activity and hence number of parasites present.

For each compound concentration, % parasite viability – the pLDH activity in compound-treated wells relative to untreated controls – was calculated. Compounds were tested in duplicate and standard deviations (SD) derived. For comparative purposes, Chloroquine (a standard anti-malarial drug) was used as a standard (IC_{50} values range from 0.01-0.05 µM).

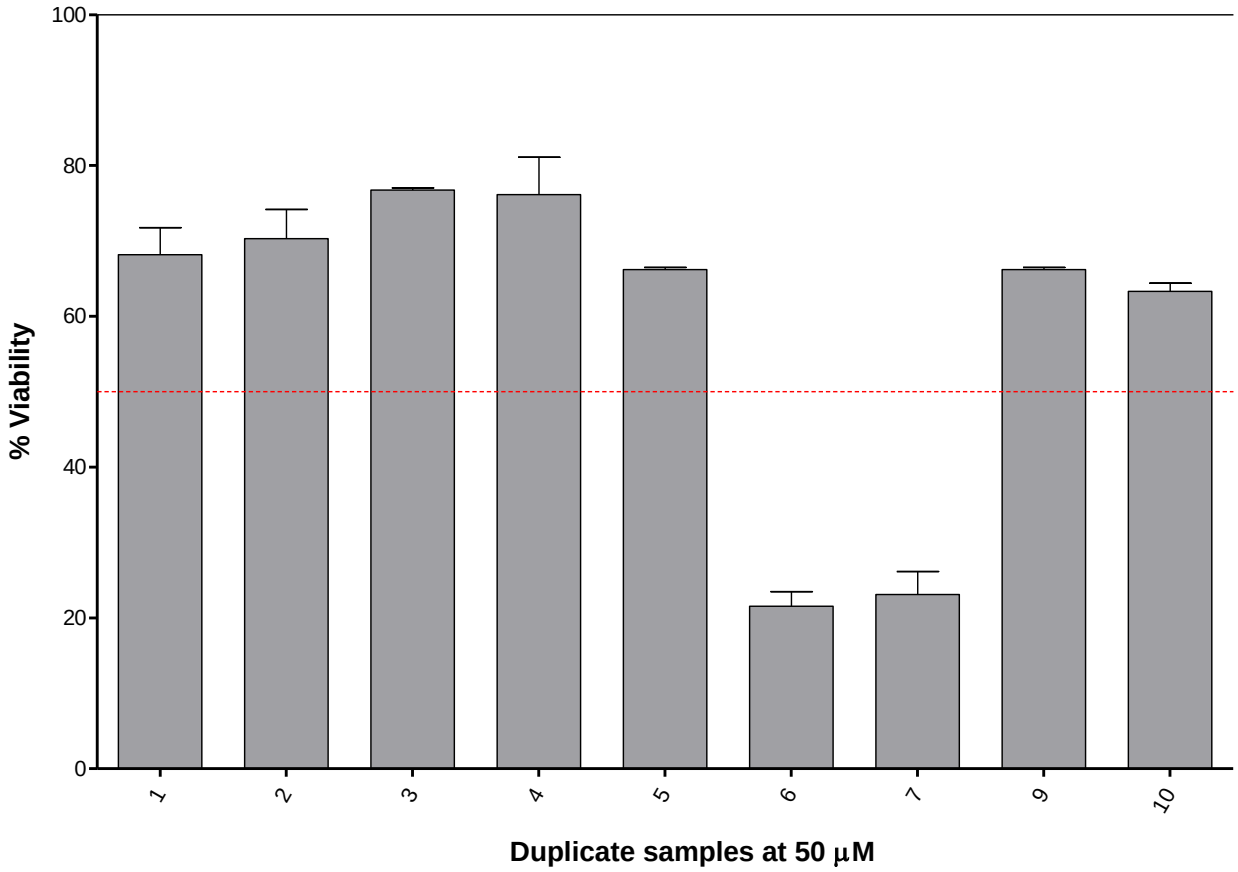
(*Note: for publication purposes, a detailed description of the method is available on request*).

Assay details

Date: 04 November 2019
Concentration used: 50 µM

Results

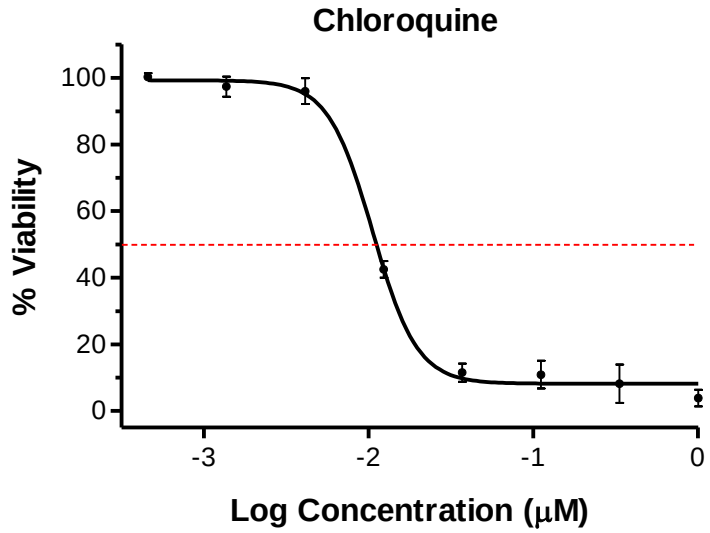
The bar graphs and tables below show the % parasite viability ±SD obtained for the individual compounds.



	50 µM	
Compound	**Viability %**	**SD**
1	68.16	3.58
2	70.30	3.86
3	76.73	0.28
4	76.14	4.96
5	66.21	0.28
6	21.55	1.93
7	23.11	3.03

9	66.21	0.28
10	63.29	1.10

Chloroquine standard @ IC50 – 0.01 µM



Conclusion

Samples 6 and 7 (highlighted in red) produced a significant decrease in pLDH at a concentration of 50 µM (a reduction in pLDH activity of at least 50%) would need to be put forward for IC_{50} screening.

Appendix E: *P. falciparum* IC50 Assay

Rhodes University

Centre for Chemico- and Biomedicinal Research

PLDH (Malaria) Assay – IC_{50} Screen

Background

Malaria parasites (*Plasmodium falciparum* strain 3D7) are maintained in RPMI 1640 medium containing 2mM L-glutamine and 25mM Hepes (Lonza). The medium is further supplemented with 5% Albumax II, 20 mM glucose, 0.65 mM hypoxanthine, 60 µg/mL gentamycin and 2-4% hematocrit human red blood cells. The parasites are cultured at 37°C under an atmosphere of 5% CO_2, 5% O_2, 90% N_2 in sealed T25 or T75 culture flasks.

For screening compounds against malaria parasites, compounds are added to parasite cultures in 96-well plates and incubated for 48h in a 37°C CO_2 incubator. After 48h the plates are removed from the incubator. Twenty µL of culture is removed from each well and mixed with 125 µL of a mixture of Malstat solution and NBT/PES solution in a fresh 96-well plate. These solutions measure the activity of the parasite lactate dehydrogenase (pLDH) enzyme in the cultures. A purple product is formed when pLDH is present, and this product can be quantified in a 96-well plate reader by absorbance at 620nm (Abs_{620}). The Abs_{620} reading in each well is thus an indication of the pLDH activity in that well and the number of parasites in that well.

Compounds are tested in a range extending from 100 µM to 0.0457 µM (3 fold-dilution series). For each compound concentration, **% parasite viability** – the pLDH activity in compound-treated wells relative to untreated controls – is calculated. Compounds are tested in **triplicate** wells, and a standard deviation (SD) is derived. For each compound, percentage viability is then plotted against Log (compound concentration) and the **IC_{50}** (50% inhibitory concentration) obtained from the resulting dose-response curve by non-linear regression. For comparative purposes, Chloroquine (an anti-malarial drug) is used as a drug standard (duplicate) and yields IC_{50} values in the range 0.01-0.05 µM.

(*Note: for publication purposes. a detailed description of the method is available on request*)

Assay details

Date: 14 November 2019

Concentration used: 100 µM down in 3-fold dilutions (100 µM – 0.0457 µM)

Assay layout:

Plate 1: Chloroquine standard

Sample 6

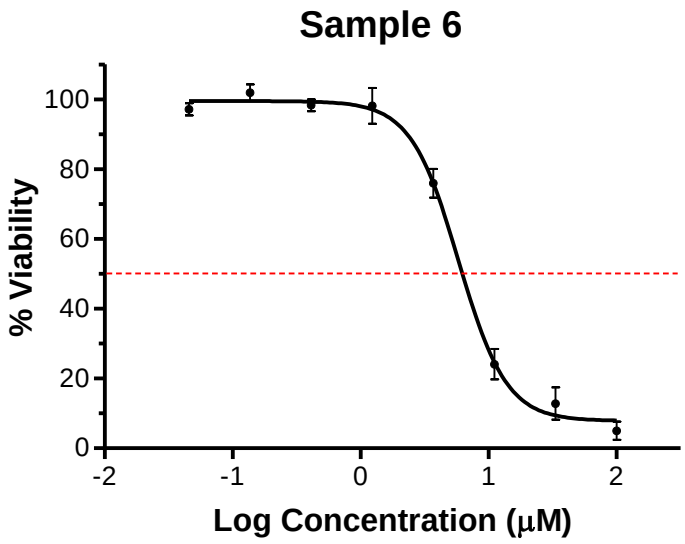
Sample 7

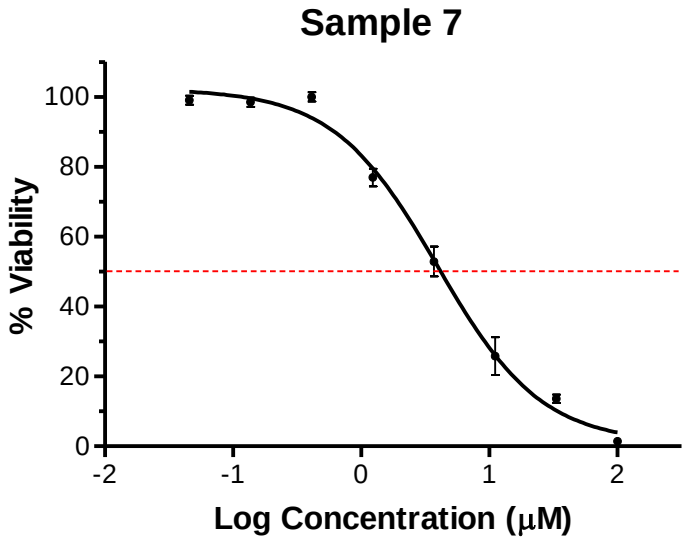
Results

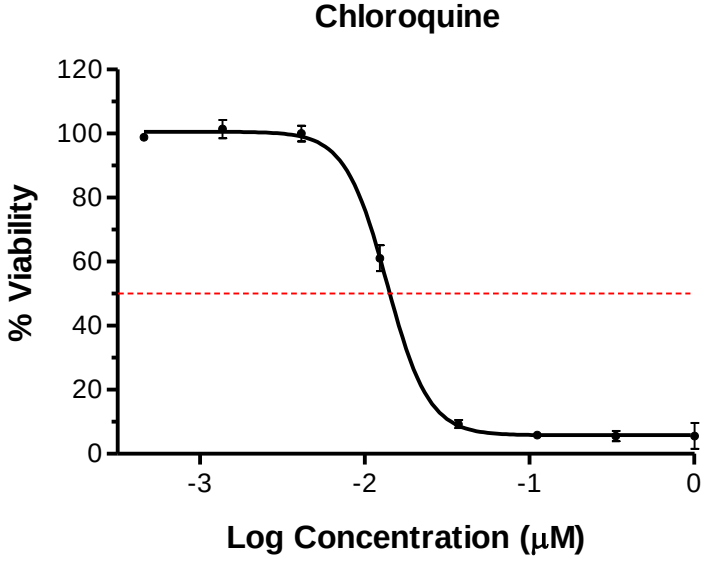
The table below shows the IC_{50} values obtained for individual compounds, followed by the dose-response plots and % viability ±SD data used to prepare the graphs.

Compound	IC_{50} (µM)	SD
Sample 6	5.8	0.18
Sample 7	3.9	0.76
Chloroquine standard	0.014	-

Plate 1


Sample 6
% Viability
100
80
60
40
20
0
-2
-1
0
1
2
Log Concentration (μM)


Sample 7
% Viability
100
80
60
40
20
0
-2
-1
0
1
2
Log Concentration (μM)


Chloroquine
% Viability
120
100
80
60
40
20
0
-3
-2
-1
0
Log Concentration (μM)

		Sample 6		Sample 7	
Conc (µM)	Log (Conc)	% Viab	SD	% Viab	SD
100	2	5.01	2.62	1.38	0.99
33.33	1.523	12.83	4.63	13.59	1.19
11.11	1.046	24.08	4.37	25.80	5.44
3.704	0.568	75.97	4.13	52.89	4.30
1.235	0.091	98.09	5.13	76.92	2.49
0.412	-0.385	98.28	1.72	100.00	1.32
0.137	-0.863	101.91	2.38	98.47	1.32
0.0457	-1.34	97.14	1.72	99.05	1.32

Conclusion

The pLDH IC_{50}'s are listed in the report above.

Appendix F: HeLa Cytotoxicity Assay

Rhodes University

Centre for Chemico- and Biomedicinal Research

Cytotoxicity Assay – Single Concentration Screen

Background

To assess the overt cytotoxicity of the compounds, they are incubated at a fixed concentration of 20 µM for pure compounds and 50 µg/mL for extracts (unless otherwise stated) in 96-well plates containing HeLa (human cervix adenocarcinoma) cells for 48 hours. The numbers of cells surviving drug exposure are also determined by using the resazurin based reagent and reading resorufin fluorescence in a Spectramax M3 microplate reader.

Results are expressed as **% viability** – the resorufin fluorescence in compound-treated wells relative to untreated controls. Compounds were tested in duplicate and standard deviations (SD) derived. Emetine (which induces cell apoptosis) is used as a control drug standard.

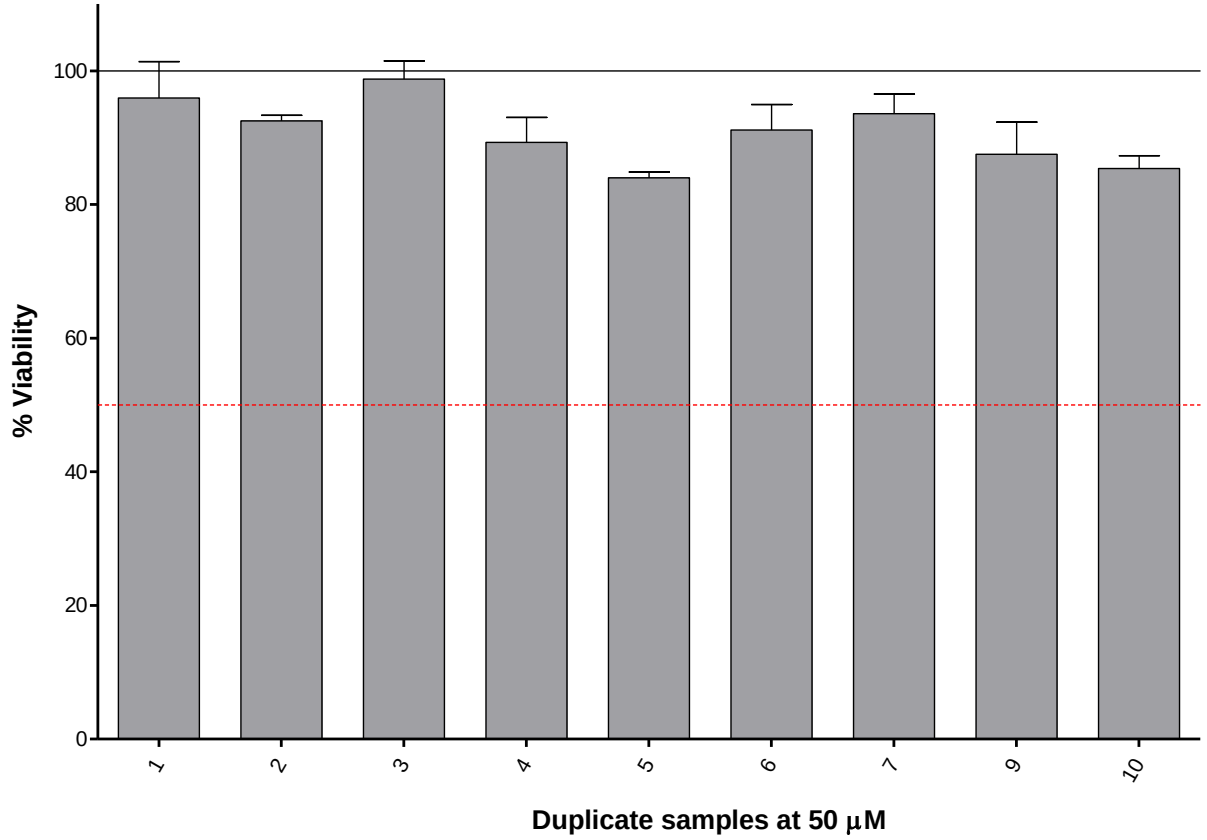
(*Note: for publication purposes, a detailed description of the method is available on request*)

Assay details

Date: 04 November 2019
Concentration used: 50 µM

Results

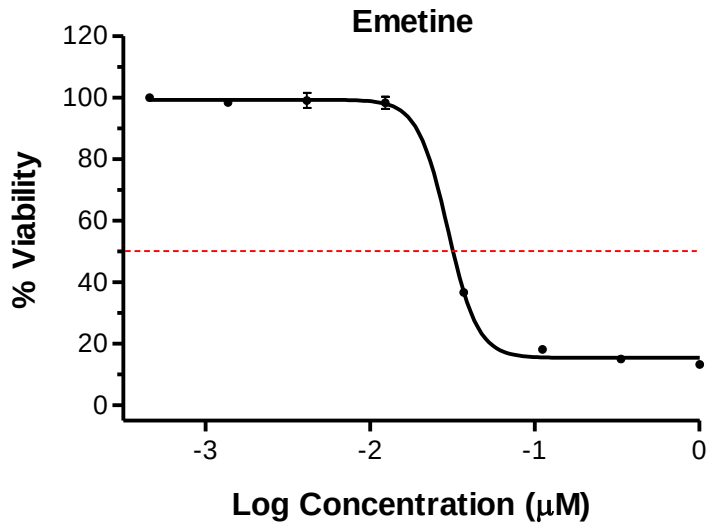
The bar graph and table below show the % HeLa cell viability ±SD obtained for the individual compounds.



	50 µM	
Compound	**Viability %**	**SD**
1	95.94	5.45
2	92.51	0.84
3	98.76	2.72
4	89.30	3.73
5	83.99	0.89
6	91.15	3.82
7	93.61	2.94

9	87.52	4.82
10	85.39	1.93

Emetine standard @ IC50 – 0.03 µM



Conclusion

None of the samples produced a significant cytotoxic effect at a concentration of 50 µM (reduce viability of HeLa cells to below 50%).