

Structure-based inhibitor design and validation:
Application to *Plasmodium falciparum* glutathione
S-transferase

Maria Magdalena Botha

Submitted in partial fulfillment of the requirements for the degree *Magister Scientiae*
in the Faculty of Natural and Agricultural Sciences

Department of Biochemistry

University of Pretoria

Pretoria

June 2006

Acknowledgments

- My faithful companions Magrathea, Eddie, Liewe Heksie as well as Eskarina for all their undaunted support in all our failures and successes.
- Without wind, grass does not move. Without Software, Hardware is useless.
- My supervisors Professors F Joubert, CP Kenyon and AI Louw for support and guidance.
- Dr. LM Birkholtz for all the time, support, interest and understanding.
- Prof. E Liebau for kindly providing me with the PfgST gene cloned into a high expression vector as well as the other technical and experimental detail.
- The CSIR and University of Pretoria for awarding me bursaries to enable me to complete my studies.
- My fellow CSIR scientists for helping this Bioinformaticist to survive in the wet lab, especially Lyndon, Robyn, Anjo and Ilze.
- My fellow students at the University of Pretoria for all their valued advice, insight and useful discussions.
- My parents as well as my Pretoria parents for their love, patience and support.
- Religion is for those who don't want to go to hell. Spirituality is for those who have been there.
- I declare that the thesis/dissertation, which I hereby submit for the degree MSc. Bioinformatics at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE: DATE:

Contents

Acknowledgments	i
List of Figures	vi
List of Tables	viii
List of Abbreviations	ix
Chapter 1. Literature Review	1
1.1. The malaria parasite <i>Plasmodium falciparum</i>	1
1.2. Malaria in Africa	2
1.2.1. Malaria: The Disease	2
1.2.2. The <i>Plasmodium</i> parasite life cycle	2
1.3. Overview of current antimalarials	3
1.4. The problems and the prospects of drug resistance	5
1.5. Drug Targets	6
1.6. Antioxidant defence in <i>P. falciparum</i>	7
1.6.1. Glutathione S-transferase (GST)	9
1.6.2. PfGST as a drug target	9
1.7. Properties of Glutathione S-transferase (GST)	10
1.7.1. The enzymatic mechanism of GST	10
1.7.2. Structure of GST	11
1.7.3. Crystal structure of PfGST (PDB 1Q4J)	12
1.7.3.1. Active site characteristics	13
1.7.3.2. The μ -loop	14
1.7.3.3. The biological dimer	15
1.7.3.4. Known inhibitors to PfGST	16
1.8. Research Aims	16
Chapter 2. <i>In silico</i> ligand design and validation of ligands	18
2.1. Introduction	18
2.1.1. Structure-based ligand design	20
2.1.2. Docking algorithms	23

2.1.3.	Scoring functions	26
2.2.	Goals	29
2.3.	Materials and Methods	30
2.3.1.	Software validation	30
	Docking studies using FlexX	30
	Docking studies using DOCK	30
	Docking studies using AutoDock	31
2.3.2.	Ligand design	31
2.3.2.1.	Modifications of S-hexyl glutathione using LUDI	32
2.3.2.2.	Knowledge-based drug design using NEWLEAD	33
2.3.3.	Estimation of protein binding affinities using consensus scoring based on results from AutoDock , LUDI and XScore	35
2.3.4.	Hierarchical clustering of molecules based on molecular fingerprints	36
2.3.5.	Investigation of the <i>in silico</i> -determined properties of the four commercially available compounds	36
2.4.	Results	37
2.4.1.	Validation of software	37
2.4.2.	Ligand design	39
2.4.3.	<i>In silico</i> screening: Estimation of protein binding affinities using consensus scoring based on results from AutoDock , LUDI and XScore	44
2.4.4.	Hierarchical clustering of database molecules based on molecular fingerprints	45
2.4.5.	<i>In silico</i> -determined properties of the four commercially available compounds	46
2.5.	Discussion	52
2.5.1.	Program validation	52
2.5.2.	Ligand design	53
2.5.3.	EDP, LAP and NDA as inhibitors of PfGST	53
2.5.4.	Clustering and biological testing criteria	54
2.5.5.	<i>In silico</i> binding affinity determination of EDP, LAP and NDA	55
Chapter 3.	<i>In vitro</i> evaluation of inhibitors	59
3.1.	Introduction	59
3.1.1.	Correlation between <i>in silico</i> and biological data	59
3.1.2.	Previous cloning and recombinant expression of PfGST	61
3.1.3.	Testing the activity of PfGST using a colourimetric assay	62
3.1.4.	Additional kinetic data available for PfGST	63
3.1.5.	Criteria for validation of methodology	63
3.2.	Goals	64

3.3.	Materials and Methods	66
3.3.1.	Restriction enzyme digestion	66
3.3.2.	Agarose gel electrophoresis	66
3.3.3.	Preparation of electrocompetent cells	66
3.3.4.	Transformation of BL21(DE3) using electrocompetent cells	67
3.3.5.	Conventional miniprep plasmid isolation	67
3.3.6.	Protein expression	67
3.3.7.	Protein extraction	68
3.3.8.	Protein purification	69
3.3.8.1.	GSTrap column specifications	69
3.3.8.2.	Affinity purification methodology	69
3.3.8.3.	SDS-PAGE analysis	69
3.3.8.4.	Sample preparation for enzyme assay	70
3.3.8.5.	Protein concentration determination	70
3.3.9.	Enzyme assays	71
3.3.9.1.	DMSO solvent controls	71
3.3.9.2.	Standard GTX inhibition assay	71
3.3.10.	Inhibition assays	72
3.3.10.1.	Inhibition of PfGST by L-3-aminomethylphenylalanine (LAP)	72
3.3.10.2.	Inhibition of PfGST by ethyl 4-amino-2-[(4-ethoxy-2, 4-dioxobutyl) thio]-5-pyrimidine carboxylate (EDP)	72
3.3.10.3.	Inhibition of PfGST by 3-(2-Naphthyl)-D alanine (NDA)	72
3.3.11.	Lineweaver-Burk double reciprocal plots	73
3.3.12.	<i>In silico</i> comparison between the binding of NDA and Praziquantel (PZQ) to PfGST	74
3.4.	Results	75
3.4.1.	Restriction enzyme digestion	75
3.4.2.	Protein expression	75
3.4.3.	Protein purification	76
3.4.4.	Protein concentration determination	77
3.4.5.	Enzyme assay	78
3.4.6.	Inhibition assays	78
3.4.7.	Enzyme kinetics	79
3.4.8.	<i>In silico</i> comparison between the binding of NDA and Praziquantel (PZQ) to PfGST	81
3.5.	Discussion	82
3.5.1.	Overexpression and purification of recombinant PfGST	82

3.5.2. Enzyme assays	82
3.5.3. Comparison between <i>in silico</i> predictions and biological results	84
Chapter 4. Concluding Discussion	87
Summary	92

List of Figures

1.1.	The <i>Plasmodium</i> parasite life cycle.	3
1.2.	The global status of multidrug resistance.	6
1.3.	The antioxidant defense system of the <i>Plasmodium</i> parasite	8
1.4.	The chemical and 3D structures of glutathione (GSH).	9
1.5.	The addition of GSH to CDNB, catalyzed by GST.	11
1.6.	The basic GST dimeric fold is built out of a thioredoxin part on the N-terminal and helices on the C-terminal part.	11
1.7.	Secondary structure of the PfGST monomer.	12
1.8.	Stereoview of a ball and stick presentation showing the binding of S-hexyl GSH at the G-site of PfGST.	14
1.9.	Structural comparison of glutathione S-transferase enzymes.	15
1.10.	The biological dimer shown along the 2-fold axis.	16
2.1.	Summary of computer-aided structure-based ligand design (CASBLD)	19
2.2.	The difference between the design methodology of LUDI and NEWLEAD.	23
2.3.	Structure of GTX with the atoms which were replaced or where addition occurred indicated by a blue circle.	33
2.4.	Fragments of GSH used by NEWLEAD	34
2.5.	Ligands designed by LUDI	41
2.6.	The 46 ligands designed by NEWLEAD.	43
2.7.	The scaffold from each cluster produced by Ward's hierarchical clustering.	47
2.8.	The twelve ligands that were commercially available.	48
2.9.	AutoDock docking of GTX, EDP, LAP and NDA.	49
2.10.	Representation of the bonds that the inhibitors GTX, EDP, LAP and NDA could form with PfGST.	51
3.1.	Addition of GSH to CDNB, catalyzed by GST.	62
3.2.	The four commercially available inhibitors GTX, EDP, LAP and NDA.	65
3.3.	BamHI and HindIII restriction analyses of the pJC20-PfGST plasmid.	75
3.4.	SDS-PAGE analysis of protein expression at different temperatures.	76
3.5.	SDS-PAGE analysis of protein expression of induced and uninduced protein expression.	77

3.6. SDS-PAGE gel of the protein purification.	77
3.7. Standard curve for protein concentration determination constructed by the Folin-Lowry assay.	78
3.8. Effect the inhibitors GTX, EDP, LAP and NDA had on PfGST activity.	79
3.9. Lineweaver-Burke Plots for GTX, LAP, EDP and NDA.	80
3.10. Lineweaver-Burke Plots for GTX, LAP, EDP and NDA.	81
3.11. PfGST and SjGST dimer with PZQ and NDA located at the non-substrate binding site.	83
3.12. Chemical structures of Praziquantel and NDA.	84

List of Tables

1.1.	Current antimalarials and their commercial names.	4
2.1.	Comparison of AutoDock, FlexX and DOCK with regards to the energy values and RMSD values of the top 5 ligands.	38
2.2.	The cluster distribution of the ligands in the library as well as the commercially available compounds.	46
2.3.	The <i>in silico</i> derived screening data of the four commercially available compounds. . . .	47
2.4.	Interactions formed between PfGST and the commercially available inhibitors (GTX, LAP, EDP and NDA).	52
2.5.	Comparison of GTX binding between PfGST and the human μ -class glutathione S-transferase (HmGST) enzyme. Hydrophobic interactions are represented by the plain text and hydrogen bonds are represented by bold text.	57
3.1.	Comparison of cloning and expression between the methods used by Perbandt <i>et al.</i> (2002) and Harwaldt <i>et al.</i> (2002).	61
3.2.	Inhibition of PfGST by selected compounds	63
3.3.	The volumes of DMSO that were added to the reaction corresponding to each inhibitor concentration.	71
3.4.	Comparison between <i>in silico</i> and biological assay data of GTX, NDA, EDP and LAP. . .	85

List of Abbreviations

3D	Three dimensional
ACD	Available chemicals directory
ADMET	Absorption distribution metabolism excretion toxicity
BSA	Bovine serum albumin
CDNB	1-chloro-2,4-dinitrobenzene
CQ	Chloroquine
DFP	Davidson-Fletcher-Powell
DHFR	Dihydrofolate Reductase
DMSO	Dimethyl Sulphoxide
DNA	Deoxyribonucleic acid
EDP	Ethyl 4-amino-2- [(4-ethoxy-2, 4-dioxobutyl) thio]-5-pyrimidine carboxylate
EDTA	Ethylenediaminetetraacetic acid
EtBr	Ethidium Bromide
FPIX	Ferriprotoporhyrin IX
GSH	Glutathione
GST	Glutathione S-transferase
GTX	S-hexyl glutathione
HCL	Hydrochloric acid
HEPES	4-2-hydroxymethyl-1-piperazineethanesulfonic acid
HTS	High Throughput System
HZ	Hemozoin
IPTG	Isopropyl-beta-D-thiogalactopyranoside
Ki	Inhibition constant
LAP	L-3-aminomethylphenylalanine
LB	Luria-Bertani
NDA	3-(2-Naphthyl)-D alanine
NCI	National Cancer Institute
NMR	Nuclear Magnetic Resonance

PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein database
RMSD	Root mean square deviation
ROS	Reactive Oxygen Species
PfGST	<i>Plasmodium falciparum</i> glutathione S-transferase
PMSF	Phenylmethylsulphonyl fluoride
RNA	Ribonucleic acid
SDS	Sodium dodecyl sulphate
SH	Thiol
TAE	Tris Acetate Electrophoresis Buffer
TEMED	N,N,N',N'-Tetramethylethylenediamine
TSB	Transformation and Storage Buffer

Chapter 1

Literature Review

“A man’s mind stretched to a new idea never goes back to its original dimensions”
- Oliver Wendell Holmes

1.1. The malaria parasite *Plasmodium falciparum*

Malaria is caused by the protozoan pathogen *Plasmodium*, which infects the red blood cells of many birds, reptiles and mammals. Four of these *Plasmodium* species infect humans, namely *P. vivax*; *P. ovale*; *P. malariae* and *P. falciparum* (Winstanley, 2000). Amongst these four species, *P. falciparum* stands out as the most malignant form, causing severe complications such as cerebral malaria, severe anemia, renal failure and pulmonary affliction. Transmission of the parasite takes place *via* certain female *Anopheles* mosquitoes during the ingestion of a blood meal from its human host. The parasite has to undergo a crucial development process in the mosquito, and this can only take place in specific mosquito species including *Anopheles gambiae*. *P. falciparum* is dominant in tropical Africa, where most of the people are infected during their childhood; morbidity and mortality occur mainly in children under the age of 5 years (Delfino *et al.*, 2002). Inadequate knowledge is severely limiting our ability to bring forth new strategies to win the war against malaria. It is therefore critical to gain insights to the parasite’s vulnerability and the basic mechanisms through which antimalarials function and resistance is conferred, in order to improve existing drugs and recruit new drugs with novel mechanisms (Olliaro, 2001).

1.2. Malaria in Africa

1.2.1. Malaria: The Disease

Malaria occurs in many locations of the tropical world and in some locations of the subtropics. Ninety percent of the world's malaria cases occur in Africa, especially in Sub Saharan Africa. Malaria outbreaks are being reported in some locations of Africa that had been previously thought to be at elevations too high for malaria transmission, such as the highlands of Kenya. Also, malaria has resurged in certain locations of Africa that previously had effective control programs, such as Madagascar, South Africa and Zanzibar (Hay *et al.*, 2004).

1.2.2. The *Plasmodium* parasite life cycle

The severity of the disease is influenced by the cellular and molecular events that take place during the parasitic life cycle. The life cycle of the malaria parasite begins with a mosquito having a blood meal where it ingests erythrocytes infected with malaria gametocytes, shown in Figure 1.1.

Inside the host the gametocytes are released from erythrocytes and differentiate into male and female gametes, microgametes and macrogametes. A zygote is formed from the fertilisation of the macrogametes by the microgametes. These zygotes advance into ookinetes and penetrate the midgut wall of the mosquito. Oocysts (differentiated ookinetes) rupture releasing sporozoites that travel to the mosquito's salivary glands. These sporozoites are then transferred to the intermediate human host through the bite of an infected mosquito. Sporozoites migrate towards the liver and invade the hepatocytes. Most sporozoites grow into schizonts containing thousands of merozoites. During the asexual blood stage, merozoites are released into the blood stream to infect circulating red blood cells. Inside the erythrocyte the parasite first develops into the ring stage, followed by the trophozoite stage, and finally a schizont is produced. Merozoites are released, with the destruction of the erythrocytes to begin the cycle again. Alternatively, some merozoites differentiate into gametocytes (Miller *et al.*, 2002; Moore *et al.*, 2002).

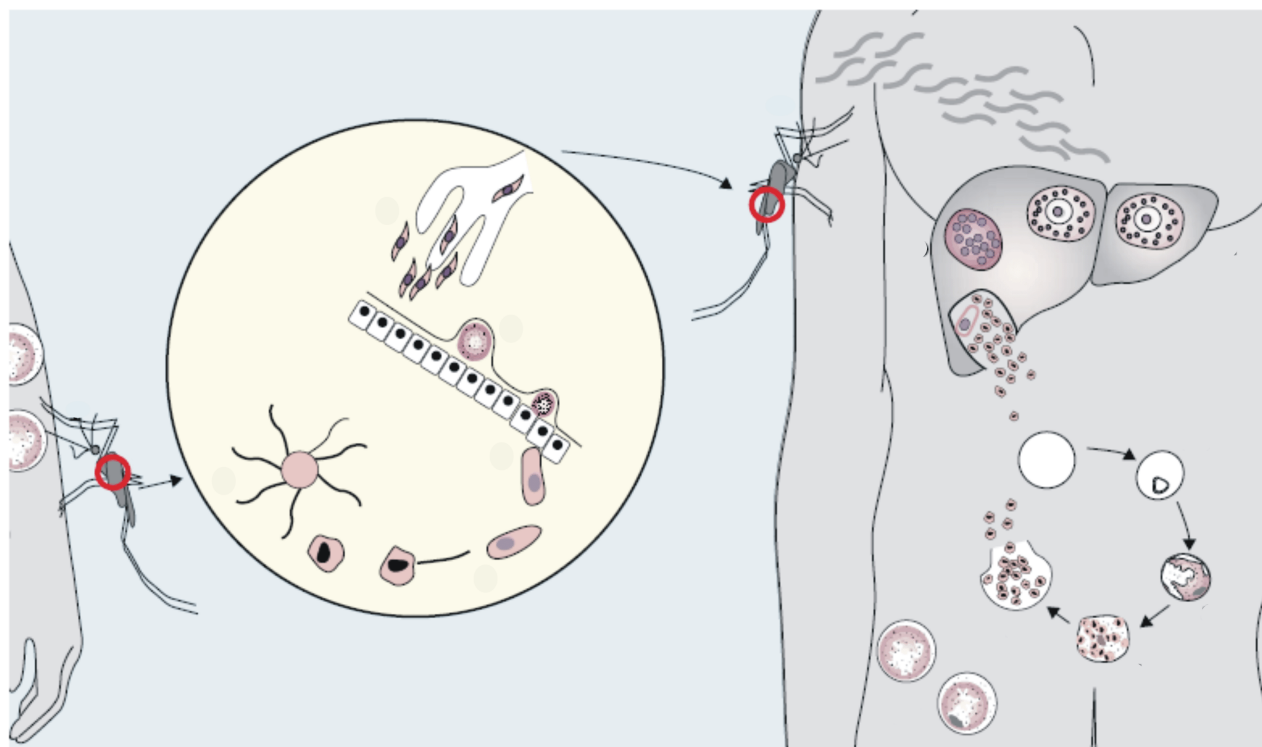


Figure 1.1: The *Plasmodium* parasite life cycle (adapted from Moore *et al.*, 2002).

Humans are the hosts for the asexual phase of the parasite life cycle who contract the illness from the bite of an infected female *Anopheles* mosquito, host for the sexual development phase.

1.3. Overview of current antimalarials

Killing the mosquitoes that transmit malaria has been an established method for controlling and preventing it but it is difficult to eradicate a complete species, therefore one can try to prevent individual infections by methods such as insecticide-treated bed nets. However, the most effective way of controlling malaria is the usage of antimalarials of which a brief overview is given below. A summary of the commercial names of these antimalarials is given in Table 1.1

Quinolines

The quinoline antimalarials and related aryls were derived from quinine, which is found in the bark of the *Cinchona* tree. The limitations of finding enough raw material to do the extractions from led to the development of synthetic 4-aminoquinoline antimalarials, chloroquine, amodiaquine, mefloquine and the structurally similar halofantrine. The suc-

cess of chloroquine as a first line defence drug against malaria, can be attributed to its clinical efficacy, limited host toxicity, easy use and simple cost-effective synthesis. Unfortunately, the mode of action of the quinoline compounds is not completely understood, hence the difficulty in circumventing the parasite's resistance (Biagini *et al.*, 2003).

Artemisinin

This class of antimalarial drugs contain semisynthetic derivatives of artemisinin (originally found in the herb *Artemisia annua*), like artemether, arteether and artesunate. Inside the human body these drugs are metabolized into a more active compound, namely dihydroartemisinin. These drugs are known for their ability to quickly reduce the rates of parasitaemia (White, 1997) and can be used for treatment of severe and uncomplicated types of malaria (Ollario, 2001). There has been no reported case of clinically-relevant resistance against artemisinin or its derivatives in Africa (Meshnick, 2002). Resistance may be delayed due to the compound's short half life and elimination time (White, 2004).

Table 1.1: Current antimalarials and their commercial names.

Current Antimalarials	Commercial names
Chloroquine	Avloclor & Nivaquine
Mefloquine	Lariam & Mefliam
Pyrimethamine-Sulfadoxine	Fansidar
Pyrimethamine-Dapsone	Maloprim
Proguanil-Dapsone	Lapdap & Paludrine
Atovaquone-Proguanil	Malarone
Artemisinin	Coartem & Riamet

Antifolates

The most successfully used antifolate drug is the combination of 2,4-diaminopyrimidine pyrimethamine and sulphadoxine, which inhibits the enzymes dihydrofolate reductase (DHFR) and dihydroopteroate synthetase (DHPS), respectively. One dose of this combination is usually highly effective due to the slow elimination from the body. Unfortunately, this combination is subjected to a high emergence of drug resistance (Winstanley, 2000).

Atovaquone/Proguanil

Atovaquone was used in monotherapy as a selective inhibitor of parasite mitochondrial electron transport leading to the disruption of membrane potential, but a high frequency of parasite resistance occurred. Atovaquone was then combined with proguanil (a DHFR

inhibitor) in a synergistic approach for therapy and prophylaxis of malaria. The biggest disadvantage for using atovaquone and proguanil is the high production cost (Kremsner, 2004).

1.4. The problems and the prospects of drug resistance

In the light of the overview of antimalarial drugs it is important to remember that there is no such thing as a perfect drug. Each and every drug known to man has its benefits but also its liabilities. Drug resistance is an additional liability when performing research on an infectious disease, this however should be a primary motivation for continued innovation.

Plasmodium falciparum resistance against the following antimalarial drugs: chloroquine, mefloquine, halofantrine, lumefantrine, pyrimethamine, cycloguanil, chlorocycloguanil, atovaquone, sulfonamides and sulfones has been well documented (White, 2004). *P. falciparum* resistance to chloroquine was first reported in southern Asia and South America in 1957, and in 1978 drug resistant malaria parasites reached Africa. In 1989 chloroquine resistance became a major health threat in Africa. Consequently, the first line defence switched to a combination therapy of sulfadoxine and pyrimethamine. In Africa resistance against sulfadoxine-pyrimethamine occurred in the late 1980's. This combination drug acts synergistically to inhibit enzymes in the folate pathway. The molecular basis for this resistance is specific mutations in their target enzymes, and has been studied widely. Clinical resistance to quinoline occurs only sporadically in southern Asia and western Oceania. Therefore, quinolines have established themselves as a second line or third line drug against severe malaria, it is used in combination with antibiotics such as tetracycline and doxycycline to increase effectiveness. Studies suggest the presence *P. falciparum* with low mefloquine sensitivity in Africa but clinical mefloquine resistance is rare. Unfortunately the molecular basis of resistance to other antimalarials like mefloquine, amodiaquine, halofantrine and artemisinin derivatives is incompletely understood (Wongsrichanalai *et al.*, 2002; May and Meyer, 2003).

Classically, multidrug resistance was defined as resistance to more than two operational antimalarial drugs belonging to different chemical classes. In Figure 1.2 the global status of multidrug resistance is shown. For *P. falciparum* this usually means a parasite that shows resistance against 4-aminoquinolines (like chloroquine) and the antifolates (like the sulfadoxine-pyrimethamine combination). Future prospects to combat multidrug resistance

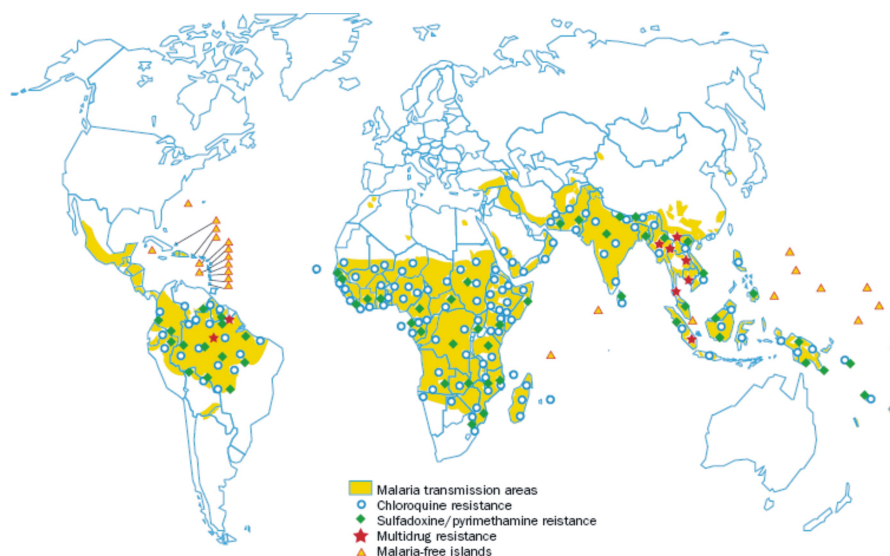


Figure 1.2: **The global status of multidrug resistance (adapted from Wongsrichanalai *et al.*, 2002).** The yellow shaded parts are those areas where malaria transmission occurs. Multidrug resistance is indicated with a red star, sulphadoxine/pyrimethamine resistance with a green diamond and chloroquine resistance with a blue circle. The few malaria-free islands are indicated by a yellow triangle.

include using drugs that are used to treat other diseases like antibacterial and antifungal drugs, increased usage of combination therapies and the discovery and exploitation of novel molecular targets (Arav-Boger and Shapiro, 2004).

1.5. Drug Targets

The completion of the genome sequence of *P. falciparum* has already aided the identification of potential drug targets and associated drugs (Gardner *et al.*, 2002). However, it is not viable to discuss all these novel molecular targets for malarial drug discovery but some of these areas deserve to be highlighted.

A major focus for antimetabolic therapy against *P. falciparum* has been the *de novo* pyrimidine biosynthetic pathway. The human host can utilise and synthesise pyrimidine nucleosides but the malarial parasite is dependent solely on *de novo* synthesis. A few of these enzymes have been identified as possible drug targets: carbamoyl phosphate synthetase II, aspartate transcarbamylase, dihydroopterose, CTP synthetase and phosphoribosyl pyrophosphate synthase (Macreadie *et al.*, 2000; Tamta and Mukhopadhyay, 2003). One of the most promising drug targets that has recently emerged is the apicoplast. The genes found

on the apicoplast genome code for enzymes that play a role in the fatty acid biosynthesis, non-mevalonate isoprenoid biosynthesis and heme synthesis pathways. It is hypothesized that some of the enzymes needed for Fe-S cluster biosynthesis are also located on the apicoplast. These enzymes make attractive drug targets for two reasons, firstly, most of these enzymes are vital for the parasites survival. Secondly, the apicoplast is a relict chloroplast, consequently the anabolic pathways coded by the apicoplast will be fundamentally different from the human host (Becker and Kirk, 2004). It is known that after the infection of a red blood cell by a malaria parasite there is an increase in the permeability of the host erythrocyte membrane. The reason for this can be attributed to the emergence of new permeability pathways (Staines *et al.*, 2005). This pathway comprises of channels and transporters (integral proteins that mediate the movement of ions and molecules across biological membranes). In principle, the inhibition of channels and transporters will deprive the parasite of nutrients needed for growth and proliferation, but also poison the parasite because of the accumulation of hazardous metabolic waste (Kirk, 2004).

1.6. Antioxidant defence in *P. falciparum*

Antioxidant defence in *P. falciparum* comprises of a functional thioredoxin and glutathione system - these two pathways are therefore very appealing drug targets. Parasites as well as other rapidly dividing cells are highly dependent on a functional antioxidant defence system. For most parasites the sources of reactive oxygen species is mainly their high metabolic rate as well as oxidative stress imposed by the host's immune system. Additionally, the *P. falciparum* parasite performs hemoglobin degradation - an important process necessary for intraerythrocytic parasite development. Hemoglobin degradation is also a source of oxidative stress and releases free radicals (Muller *et al.*, 2004). Oxidative stress may play a vital role in malaria infection, therefore the antioxidant defence system in the malarial parasites is a very promising drug target. The antioxidant defence system of *P. falciparum* is mediated by an ensemble of antioxidants like glutathione as well as antioxidant enzymes as shown in Figure 1.3 (Becker *et al.*, 2004).

GSH is a tripeptide (γ -Glu-Cys-Gly) found in most eukaryotic cells. GSH is implicated in many cellular functions which protect the living cell against toxicity and stress, induced by environmental and endogenous chemical agents (Dixon *et al.*, 2002). Digestible hemoglobin from the red blood cell provides the amino acids for the formation of malarial GSH. The

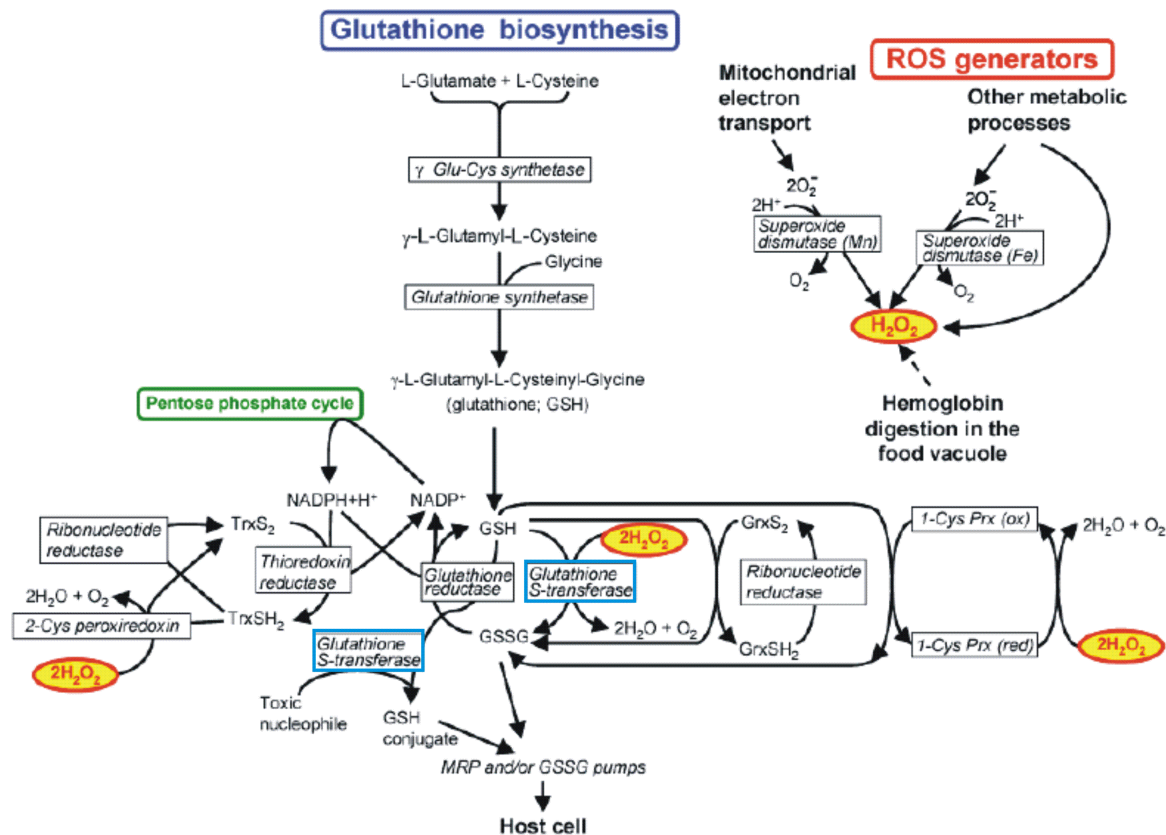


Figure 1.3: The antioxidant defense system of the *Plasmodium* parasite (from Bozdech and Ginsburg, 2004). A generalization of the biochemical processes involved in the malaria parasite's antioxidant defence system. The PfGST enzyme is indicated by a blue frame where it conjugates glutathione (GSH) to toxic nucleophiles to inactivate the harmful species. PfGST also shows peroxidase activity.

chemical structure (as well as the 3D structure) of GSH is shown in Figure 1.4.

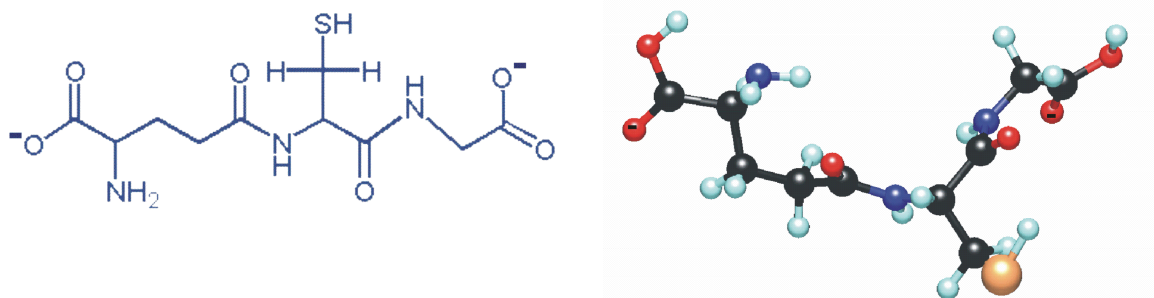


Figure 1.4: **The chemical and 3D structures of glutathione (GSH).** Glutathione (GSH) is a tripeptide composed of gamma - glutamyl - cysteinyl - glycine. In the ball-and-stick representation of GSH the carbons were coloured black, hydrogens white, nitrogens blue, oxygen red and the sulphur was coloured yellow.

1.6.1. Glutathione S-transferase (GST)

GSTs (EC 2.5.1.18) are ubiquitous multi-functional proteins involved in the detoxification of carcinogens, mutagens and other corrosive chemical substances (endogenous and exogenous xenobiotic compounds) (Fritz-Wolf *et al.*, 2003). Figure ?? shows a general scheme of the various biochemical processes of antioxidant defense (Bozcheg and Ginsburg, 2004). The primary function of GST lies in the protection of cellular macromolecules. GST deactivates harmful chemicals *via* the nucleophilic addition of the thiol (SH) group from glutathione (GSH), to the hydrophilic moiety of the toxic agent, thus rendering the electrophilic compounds harmless and enabling the removal of the substance. Because of the inactivation of potentially hazardous substances, GST activity is beneficial to an organism's health and survival (Riganese *et al.*, 2000; Sheenan *et al.*, 2001).

1.6.2. PfGST as a drug target

In chloroquine-resistant parasites GST activity is directly and positively related to drug pressure (Dubois *et al.*, 1995; Srivastava *et al.*, 1999). The emergence of multidrug resistant parasites as well as the absence of a vaccine make the targeting of antioxidant defense pathways attractive. GST activity has been detected in simian (*Plasmodium knowlesi*), rodent (*Plasmodium yoelli*, *Plasmodium berghei*), and human (*Plasmodium falciparum*) malaria parasites. Consequently, the malaria parasite is dependant on a functional GST (Perbandt *et al.*, 2003). Inhibition of GST will impair the general detoxification processes and, because the

enzyme has peroxidase activity, reduce the antioxidant capacity of the parasite (Fritz-Wolf *et al.*, 2003; Muller *et al.*, 2003). The reduction of hemozoin is a well-known feature of chloroquine resistant *P. berghei*. There exists an inverse relationship between the hemozoin content, versus GSH and *P. berghei* glutathione S-transferase (pbgST) levels. GSH can detoxify hemozoin within the food vacuole, thus precluding its polymerization and preventing the activity of chloroquine and other quinoline containing drugs. In chloroquine-resistant parasites GST activity significantly increases compared to the sensitive strain. In theory, if GST can be inhibited, it will impair GSH's detoxification capability and improve the effectiveness of drugs like chloroquine (Platel *et al.*, 1999). In addition PfGST also reduces hydroperoxides, binds and sequesters structural non-substrate ligands. Another function of PfGST is to dissociate non-polymerized hemozoin that diffuses from the food vacuole into the cytosol consequently preventing the toxic effects within the parasite. PfGST is highly abundant, it occupies between 0.1 and 1% of the total parasite protein fraction (Srivastava *et al.*, 1999).

1.7. Properties of Glutathione S-transferase (GST)

1.7.1. The enzymatic mechanism of GST

Different isoforms can accommodate a wide variety of substrate molecules. GST can further detoxify lipid peroxidation products and serve as carrier protein (ligandins) of various organic molecules, which leads to the inactivation and immobilization of these compounds. The cytosolic GST isoforms have been grouped in seven species-independent classes: Alpha, Mu, Pi, Theta, Sigma, Kappa and Phi, based on sequence similarity, immunological cross reactivity and substrate specificity (Strange *et al.*, 2000; Fritz-Wolf *et al.*, 2003). One of the most extensively used assays for GST is the addition of GSH to 1-chloro-2,4-dinitrobenzene (CDNB), a nucleophilic aromatic substitution reaction that occurs *via* an addition-elimination sequence involving a short-lived intermediate (Figure 1.5).

This is the only reaction type for which a high-resolution crystal structure of a transition state analogue is available. The reaction was crystallized for a μ -class rat GST (Armstrong, 1997). *Schistosoma japonicum* glutathione S-transferase (SjGST) exhibits conjugation activity towards alkenals, established secondary products of membrane lipid peroxidation and ethacrynic acid (Cardoso *et al.*, 2003).

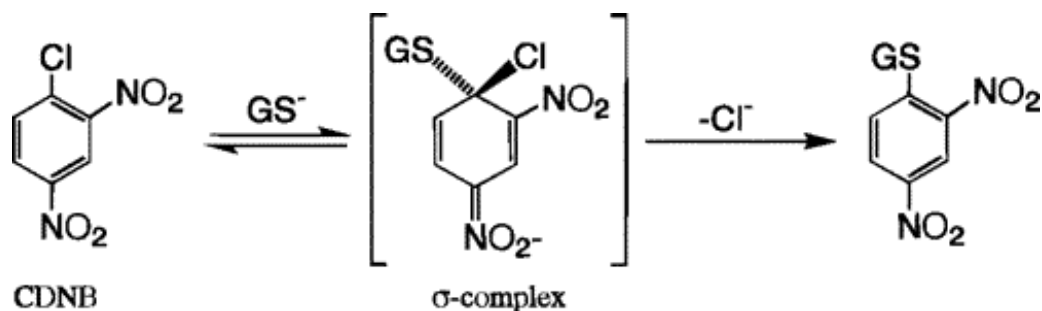


Figure 1.5: The addition of glutathione to 1-chloro-2,4-dinitrobenzene, catalyzed by glutathione S-transferases (Armstrong, 1997). The addition of glutathione to 1-chloro-2,4-dinitrobenzene results in the formation of dinitrophenyl thioester that can be detected with a spectrophotometer at 340nm, this reaction can be used to follow GST activity.

1.7.2. Structure of GST

The overall fold of the N-terminal domain is classified as the thioredoxin superfamily fold, which also includes glutaredoxin and glutathione peroxidase. This N-terminal domain constitutes roughly one-third of the protein and consists of a $\beta\alpha\beta\alpha\beta\alpha$ topology. The N-terminal part is therefore responsible for the ability to bind glutathione. To the C-terminal of the protein a bundle of α -helices was added, this influences secondary substrate specificity of the H site. The gradual build up from the fundamental thioredoxin fold to the principal glutathione S-transferase fold is shown in Figure 1.6.

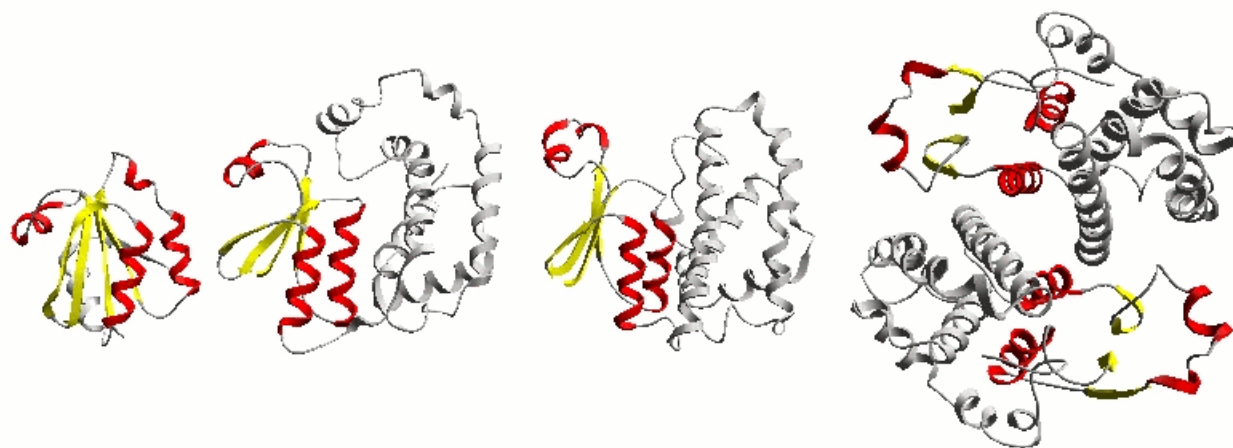


Figure 1.6: The basic GST dimeric fold is built out of a thioredoxin part on the N-terminal and helices on the C-terminal part. The PfGST structure is the basic thioredoxin fold (yellow and red), the gray helices are added to the fold to depict the basic dimeric GST fold. The N-terminal domain is thus the yellow and red thioredoxin fold and the C-terminal domain the gray helices (Dirr, 2001).

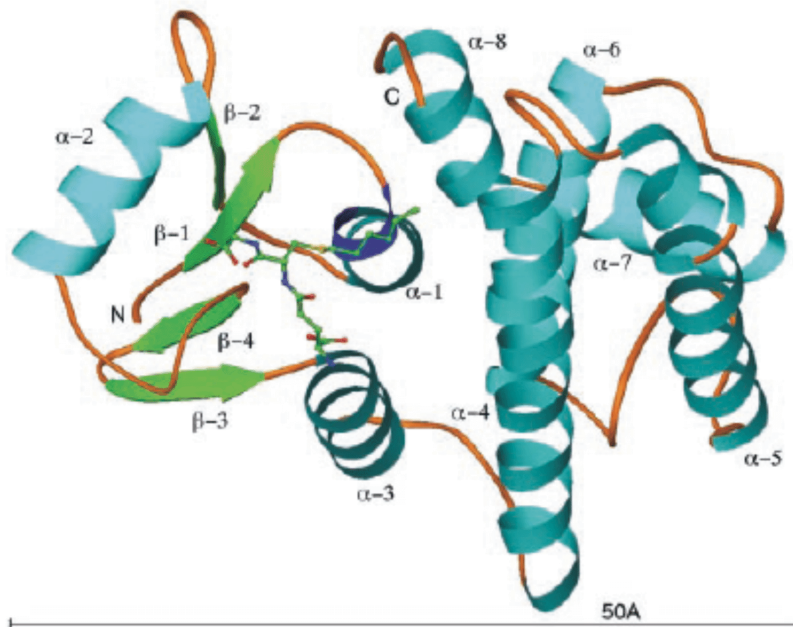


Figure 1.7: **The PfGST monomer with emphasis on secondary structure elements.** The α -helices are cyan, β -sheets are green and the 3_{10} helix are coloured dark blue. The S-hexyl glutathione (GTX) molecule bound to the active site is shown in a ball-and-stick representation (Perbandt *et al.*, 2003).

The C-terminal of the protein is an α -helical domain with a unique fold, the core consists of a bundle of four α -helices. The xenobiotic moiety lies in the crevice between the two domains and makes a number of contacts with the C-terminal domain. The C-terminal domain provides the structural elements for recognition of the nucleophilic compounds and assists substrate selectivity of the various isozymes. The glutathione moiety is bound in an extended conformation (Figure 1.7), with the glutamyl residue pointing down towards the dimer interface, the cysteinyl sulphur pointing to the subunit to which it is bound, and the glycyl residue residing near the surface of the protein (Armstrong, 1997). In particular, the tyrosine residue near the N-terminal end of the enzyme plays a crucial role in the deprotonation of glutathione (Riganese *et al.*, 2000).

1.7.3. Crystal structure of PfGST (PDB 1Q4J)

PfGST (EC 2.5.1.18) is a multi-functional protein consisting of two monomers. A single monomer is shown in Figure 1.7. In accordance with other GST enzymes each monomer of PfGST contains an N-terminal α/β domain and C-terminal α -helical domain. The active site is defined by two binding sites: the G site, which binds GSH, and the more flexible H

site, which can bind various other substrates (Perbandt *et al.*, 2003). The PfGST monomer consists of two domains linked by a loop comprising residues Tyr83 to Glu88. Domain 1 (Met 1 to Lys 82) contains a four-stranded mixed β -sheet surrounded by three α -helices. Domain 2 (Ser 89 to Tyr 211) is mostly α -helical.

Domain 1 is highly conserved and forms part of the glutathione-binding G site. The G site is relatively rigid and not greatly affected by inhibitor binding, with the exception of the C-terminal tail and the loop connecting the α -4 and α -5 helices. This region is very specific for its natural substrate (GSH). The recognition and binding occur via a network of polar interactions between PfGST and GSH.

1.7.3.1. Active site characteristics

Like most members of the GST superfamily, the PfGST active site is situated in the crevice that forms between domain 1 and 2 in each monomer. The active site comprises of two binding sites: the G site, which binds GSH, and the more accommodating H site (Perbandt *et al.*, 2003). Harwaldt *et al.* (2002) identified alternative substrates for PfGST, that could bind in the H site. 1-chloro-2,4-dinitrobenzene (CDNB) is a known secondary substrate for PfGST as well as ethacrynic acid and *o*-nitrophenyl acetate (Harwaldt *et al.*, 2002).

The thiol group on the GSH needs to be activated before the conjugation of GSH to the electrophile can take place, The OH group from Tyr 9 is believed to assist in this activation. The alignment of the Tyr 9 and the sulphur on GSH is shown in Figure 1.8.

The specific interactions between amino acids in the G site and GTX are shown in Figure 1.8 in green (Perbandt *et al.*, 2003; Fritz-Wolf *et al.*, 2003). The most conserved part of the G site is very typical for the μ -class GST: Tyr 9, Gln 71 and Asp 105. The conservation of these amino acids in the active site was used to comparatively prove that PfGST belongs to the μ -class GSTs. The assignment of PfGST to a specific class is still a matter of controversy. In addition Ser 72, Val 59, Lys 15 and Glu 58 are also involved in GSH binding. Lys 15 is in close proximity of substrate and will consequently play an important part in the additional stabilization of GSH (or any other electrophilic substrate). The hydrophobic binding pocket (H site) is considerably more variable than the G site, due to the nature of second substrates. The substrate specificity of different isozymes in the GST superfamily can be attributed to the variation of amino acids present in the H site consequently leading to different interactions a ligand can form with amino acids in the H site of the enzymes

(Koehler, 1997). The H site of PfGST consists of the top part of the α -4 helix (residues 100 to 112), the loop connecting α -4 and α -5 (residues 113 to 116), the loop connecting α -1 and β -1 and the C-terminal of the protein. The C-terminal shields the H site from solvents. The degree of shielding is dependent on the class of enzyme. This cavity binds the hydrophobic moiety of the electrophilic substrate, the auxiliary stabilization is conferred by the side chains of the following amino acids : Phe 10, Phe 42, Phe 45, His 107, Phe 100, Val 103, His 107, Tyr 108, Phe 110, Val 111, Phe 116, Leu 115, Val 210 and Tyr 211.

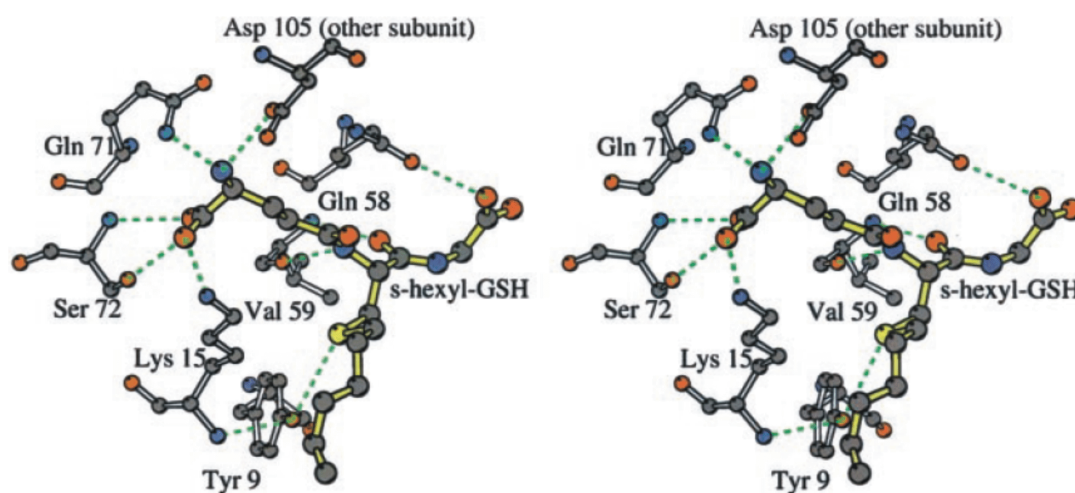


Figure 1.8: **Stereo view of a ball-and-stick presentation showing the binding of S-hexyl GSH at the G-site of PfGST (Perbandt *et al.*, 2003).** Hydrogen bonding interactions with the most concerned amino acids of the G site is indicated by the green dotted lines. Bonds between atoms that form part of the enzyme are shown in grey and the bonds between atoms that form part of GTX is highlighted in yellow. Carbon atoms were coloured grey, nitrogen atoms blue, oxygen red and the sulphur of GTX was coloured yellow.

1.7.3.2. The μ -loop

PfGST also possesses a short μ -loop. In contrast to other μ -class GST enzymes, PfGST has only 5 residues after α -8, which is too short to form a wall or an α -helix. From Figure 1.9 it can be seen that the active site is composed of two binding sites: the G site, which binds reduced glutathione and the more variable H site. The μ -class enzyme contains the glutathione conjugate 1-hydroxy-2-S-glutathionyl-3-para-nitrophenoxy-propane, and the π -class enzyme contains the glutathione conjugate of ethacrynic acid. All three enzymes have a similar binding mode for glutathione, but the secondary substrate can be bound in different conformations in the H site. A typical feature of μ - and π -class enzymes is the C-terminal loop that lines the H site. This feature is lacking in PfGST, resulting in a

more solvent-accessible H site. The result is that the H site is less shielded from solvents (Fritz-Wolf *et al.*, 2003; Srivastava *et al.*, 1999; Liebau *et al.*, 2002; Perbandt *et al.*, 2003; Becker *et al.*, 2004).

1.7.3.3. The biological dimer

PfGST is active as a homodimeric protein of 25 kDa subunit size each, as shown in Figure 1.10 (Liebau *et al.*, 2002). Despite the low overall sequence identity across the different classes of enzymes, all the structures follow a similar canonical fold, with each subunit consisting of two distinct domains (shown in Figure 1.10) The 2-fold axis is perpendicular to the plane, and helices α -3 and α -4 are arranged around the 2-fold axis. The molecular interactions between the two monomers of PfGST are class-specific. The monomers are predominantly held together by hydrophobic effects, but the four salt bridges and four hydrogen-bonded pairs of residues also contribute to the dimerization (Fritz-Wolf *et al.*, 2003; Perbandt *et al.*, 2003; Sheenan *et al.*, 2001).

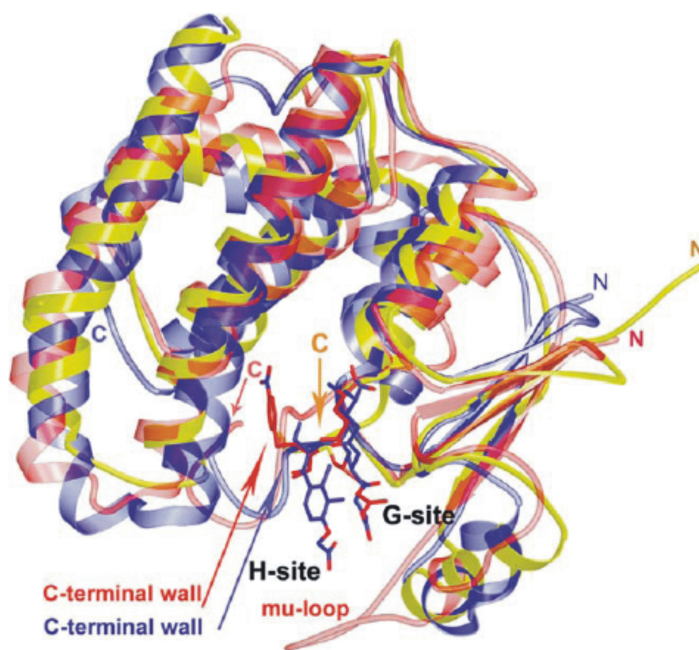


Figure 1.9: **Structural comparison of glutathione S-transferase enzymes.** Superimposed structures of *Plasmodium falciparum* glutathione S-transferase (PfGST) (yellow and green), μ -class GST (red, 1c72) and π -class GST (blue, 11gs). The μ -class enzyme contains the glutathione conjugate 1-hydroxy-2-S-glutathionyl-3-para-nitrophenoxy-propane (red lines), and the π -class enzyme contains the glutathione conjugate of ethacrynic acid (blue lines) (Becker *et al.*, 2004).

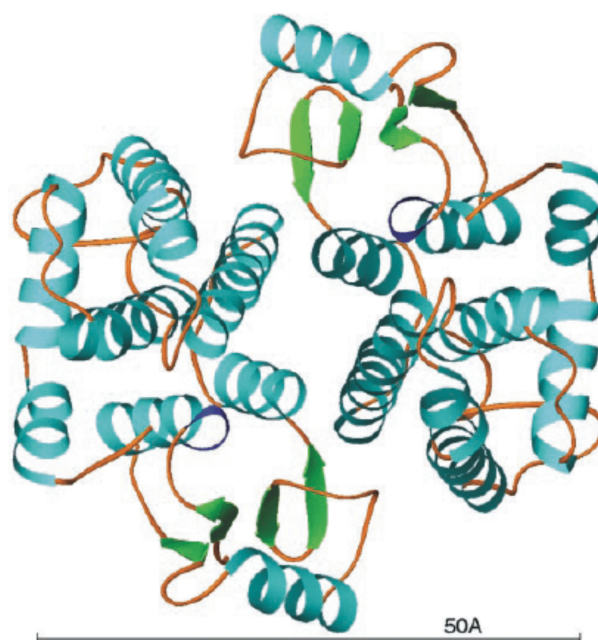


Figure 1.10: **The biological dimer shown along the 2-fold axis (Perbandt *et al.*, 2003).** The β -sheets are indicated by green arrows and α -helices by cyan coils.

1.7.3.4. Known inhibitors to PfGST

Harwaldt *et al.* (2002) did inhibition assays on PfGST using established GST inhibitors as well as compounds related to heme metabolism. As a positive control all these compounds were tested on the human placenta GST (hmGST). Cibacron blue, ethacrynic acid, hemin, S-hexyl glutathione and protoporphyrin IX inhibited both PfGST and the hmGST. Fritz-Wolf *et al.* (2003) explored inhibition of PfGST by using polypeptides. A heptapeptide Asn-Asn-Thr-Asn-Leu-Phe-Lys and an undecapeptide Asn-Asn-Thr-Asn-Leu-Phe-Lys--Asn-Asn-Ala-Thr were tested but it was concluded that inhibition by peptides is strongly dependent on pH and buffer content (Harwaldt *et al.*, 2002; Fritz-Wolf *et al.*, 2003).

1.8. Research Aims

The primary aim of this study was to use a computational structure-based ligand design strategy in finding novel ligands that can act as inhibitors to PfGST and form the basis of future antimalarial drug development. Since there is only one PfGST isoenzyme present in the parasite and the architecture of the binding site differs significantly from its human counterpart, PfGST is considered a highly attractive drug target. Inhibition of PfGST is expected to interfere in more than one metabolic site in synergy: it is likely to disrupt the

GSH-dependent detoxification process which will lead to an increase in the cytotoxic peroxide concentration and most likely lead to an increase in the levels of ferriprotoporphyrin IX and hemin as well. S-hexyl glutathione is co-crystallized with PfGST, consequently it is seen as one of the most important lead compounds in developing PfGST inhibitors. Hence the first step in the rational drug design strategy was to modify S-hexyl glutathione - concentrating on its ability to bind competitively to the G site. *In silico* validation was done by means of molecular docking and empirical scoring functions of the binding interactions between the inhibitors and the receptor. Biological validation was done by using recombinantly-expressed PfGST in a colourimetric assay.

- *Chapter 2* describes the *in silico* assessment and implementation of different ligand docking, scoring and design software used to build a focussed library of hypothetical inhibitory ligands of *P. falciparum* glutathione S-transferase.
- *Chapter 3* focuses on the biological evaluation of the designed inhibitors by using the recombinantly-expressed *P. falciparum* glutathione S-transferase in a colourimetric assay.
- In *Chapter 4* the relationship between the above chapters and their relevance to drug discovery and future prospects are discussed.

Chapter 2

In silico ligand design and validation of ligands

2.1. Introduction

The earliest structure-based drug design strategies date back to the early 1970s with ideas like the modification of insulin to increase its half life, design of serine protease inhibitors to regulate blood clotting as well as lead optimization of aspartic protease inhibitors (Congrewe *et al.*, 2005). In the last decade the number of protein structures published in the RCSB Protein Data Bank have risen from 3468 (December 1995) to 31400 (December 2005) (<http://www.pdb.org/>). This increase in available 3D protein structures marked the beginning of the structure-based inhibitor design era. Computer-aided ligand design is concerned with the prediction of molecules that are expected to bind with high affinity to key regions of pharmacologically relevant enzymes, to inhibit or alter their function. There are two major problems that the *in silico* design process needs to deal with. Firstly, the prediction of the optimum binding orientation a ligand will have in the active site of the receptor, in this case PfGST. Secondly, the estimation of binding affinity of the ligand-receptor complex. Various docking programs can be used to perform conformational sampling of specific inhibitors, addressing the binding mode prediction problem. Scoring functions can then be used to estimate the binding affinity between the target protein and different conformations of the ligand or different ligands (Wolf and Dormeyer, 2003; Ferrara *et al.*, 2004). Thirdly, the biological results have to support the *in silico* predictions to validate the strategy for future use in drug design.

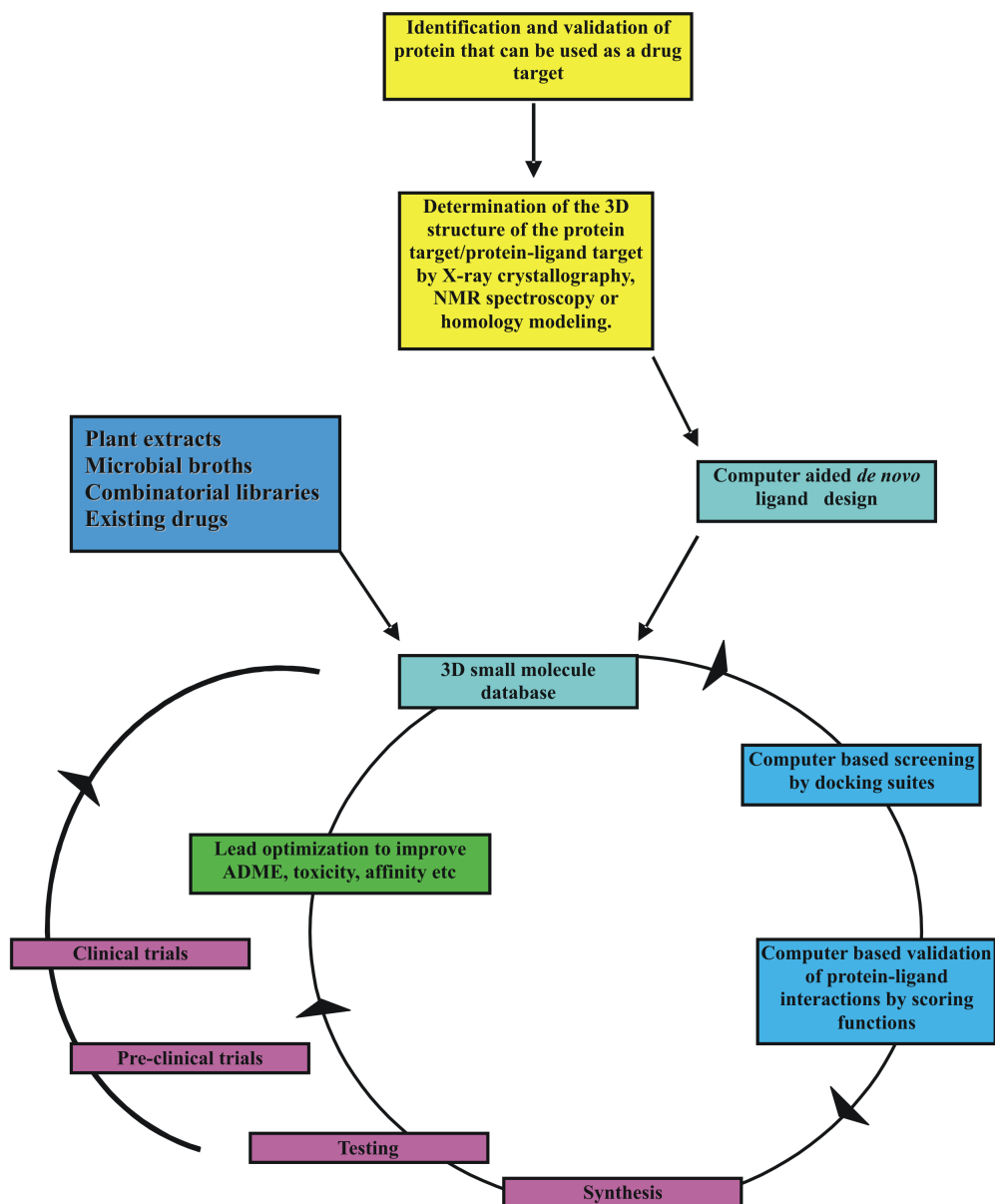


Figure 2.1: Summary of the basic methodology of computer-aided structure-based ligand design adapted from Anderson *et al.* (2003). The infinite circle of computer-aided structure based ligand design is depicted, starting with a validated target and ending with a drug like molecule for the pharmaceutical industry.

2.1.1. Structure-based ligand design

The basic methodology of computer-aided structure-based drug design is illustrated in Figure 2.1. All leading rational drug design strategies should be founded on a validated target protein, with a proven relationship between its activity and the disease.

Another requirement is that the three dimensional (3D) structural information of the target protein should be available; X-ray or nuclear magnetic resonance (NMR) diffraction usually produces these structures, preferably with a bound inhibitor. Homology modeling can be employed to overcome a lack of experimental structural information (Wolf and Dormeyer, 2003; Hillisch *et al.*, 2004).

Glutathione S-transferase of *Plasmodium falciparum* was validated as a drug target by Harwaldt *et al.* (2002). Two separate groups crystallized PfGST. In November 2003 the first crystal structure of the dimeric PfGST at a resolution of 1.9Å (PDB ID code 1OKT) was published (Fritz-Wolf *et al.*, 2003). In January 2004, Perbandt *et al.* (2004) published the second crystal structure with a resolution of 2.2Å (PDB ID code 1Q4J). The last structure was used for the structure-based drug design because the inhibitor S-hexyl glutathione (PDB ID code GTX) was co-crystallized in the active site of this PfGST enzyme.

The chemical nature of the active site determines the interaction energy between the protein and the ligand, which needs to be sufficiently strong for a complex to form. Once the active site has been explored the next step is to create a site map to define the key amino acids for ligand binding, which includes hydrogen bonding and hydrophobic patches (Brown and Van der Jagt, 2004). Protein active sites can contain various interaction points, although only a small number of points are utilized for ligand binding. This is an important consequence for drug design because it creates a combinatorial choice of points to which a novel inhibitor can bind. Selection of specific site points for protein-ligand interactions is crucial and should be done in such a manner that it differs vastly from other members in the same protein family. Designing the ligands to bind these site points is a notorious task, and can be accomplished with various strategies. Two main approaches applied are the modification of substrates and cofactors, and *de novo* lead generation. Database screening can be used as an alternative to ligand design (Zanders *et al.*, 2002). The *in silico* design and docking of ligands mimics a medicinal chemist and binding affinity predictions performs a function analogous to virtual assay (Schneider and Flechner, 2005).

Since GTX was co-crystallized with PfGST, the modification of GTX was used as a starting point of designing parasite-specific inhibitors. Many different approaches have been used to do ligand modifications, the two main approaches used in this study will be described. Although GTX is a universal inhibitor, in the usage of the PfGST structure during the design cycle, parasite specific intermolecular interactions were exploited.

The comparison of the difference in the basic methodology of how LUDI and NEWLEAD work is explained in Figure 2.2. LUDI uses the interaction points indicated by R as points for attachment of new fragments. In Figure 2.2 these fragments are indicated by the blue and red squares. The result is molecules that still have the same scaffold but with novel groups attached to the scaffold that might increase binding affinity and specificity. NEWLEAD maintains the fragments that are encircled by the orange, green and yellow circles. The core of the molecule is then redesigned to link these pharmacophore fragments by novel linkers. The result is molecules that still have the pharmacophore pattern used for binding but the core of the molecule is different.

The LUDI package has the capability to quickly generate a series of potential ligands for a molecular receptor. LUDI uses potential interaction points in the receptor binding pocket to construct novel ligands by assembling atoms and molecular fragments that are complementary to these interaction points (www.accelrys.com; Schneider and Bohm, 2002). LUDI is capable of identifying small fragments from a library containing over a 1000 random molecular fragments, and fitting them into the active site based on each fragment's ability to form favourable hydrogen bonds or hydrophobic interactions with the receptor. These fragments can then be ligated to the already existing ligand (Lew and Chamberlin, 1999). In the case of PfGST, the modification module of LUDI was used with GTX as existing ligand. The modification of GTX was done to increase binding affinity as well as to increase its specificity to PfGST, consequently manipulating GTX not to bind to all the enzymes in the GST family but only to PfGST.

Another approach that can be followed was implemented by the use of the NEWLEAD package. NEWLEAD was developed for the automatic generation of candidate structures conforming to the requirement of a given pharmacophoric pattern (Tschinke and Cohen, 1993). This technique entails the connection of given pharmacophoric fragments with spacers assembled from a library of small chemical entities (atoms, chains and ring moieties). NEWLEAD can generate novel ligands that are chemically unrelated to the reference molecule but still retain

the binding mode of the reference ligand. This provides an unbiased starting point for the design of new generations of ligands. The output is a set of molecules containing the original pharmacophore pattern now connected by novel spacers (Tschinke and Cohen, 1993). When applying NEWLEAD to the PfGST ligand design problem the pharmacophore fragments were taken from glutathione. After designing the ligands, the computer-aided structure-based ligand design study reaches the level where molecular docking and binding affinity scoring can be integrated. These approximations of how strong the ligand will bind to the receptor as well as the proposed orientation of binding, can be used to rank these putative inhibitor libraries for *in vivo* and *in vitro* testing. Molecular docking and affinity scoring play various roles in the field of drug design: providing a fast and reliable filter for high throughput virtual screening. Affirmation of the binding mode of a class of novel structurally-different inhibitors can be used in database screening but most of the time it is used to discriminate between lead-like and non-lead like drug molecules. Docking and scoring can also be used to reduce the cost of drug design by producing a means of filtering libraries before chemical synthesis and testing starts.

The LUDI approach has been successfully used to design inhibitors for dihydrofolate reductase (DHFR) and trypsin. LUDI generated fragments that were similar to known inhibitors like methotrexate and benzamidine that inhibit DHFR and trypsin, respectively (Bohm, 1992). Lew and Chamberlin, (1999) also had success in designing blockers of human T-cell potassium channels by using LUDI. The NEWLEAD program has been used to design inhibitors for DHFR as well, by using the pharmacophore pattern of methotrexate (Tschinke and Cohen, 1993). Tschinke and Cohen, (1993) also used the prostaglandin inhibitor indometacin as well as the HIV-I protease inhibitor A74704, to validate the functionality of NEWLEAD.

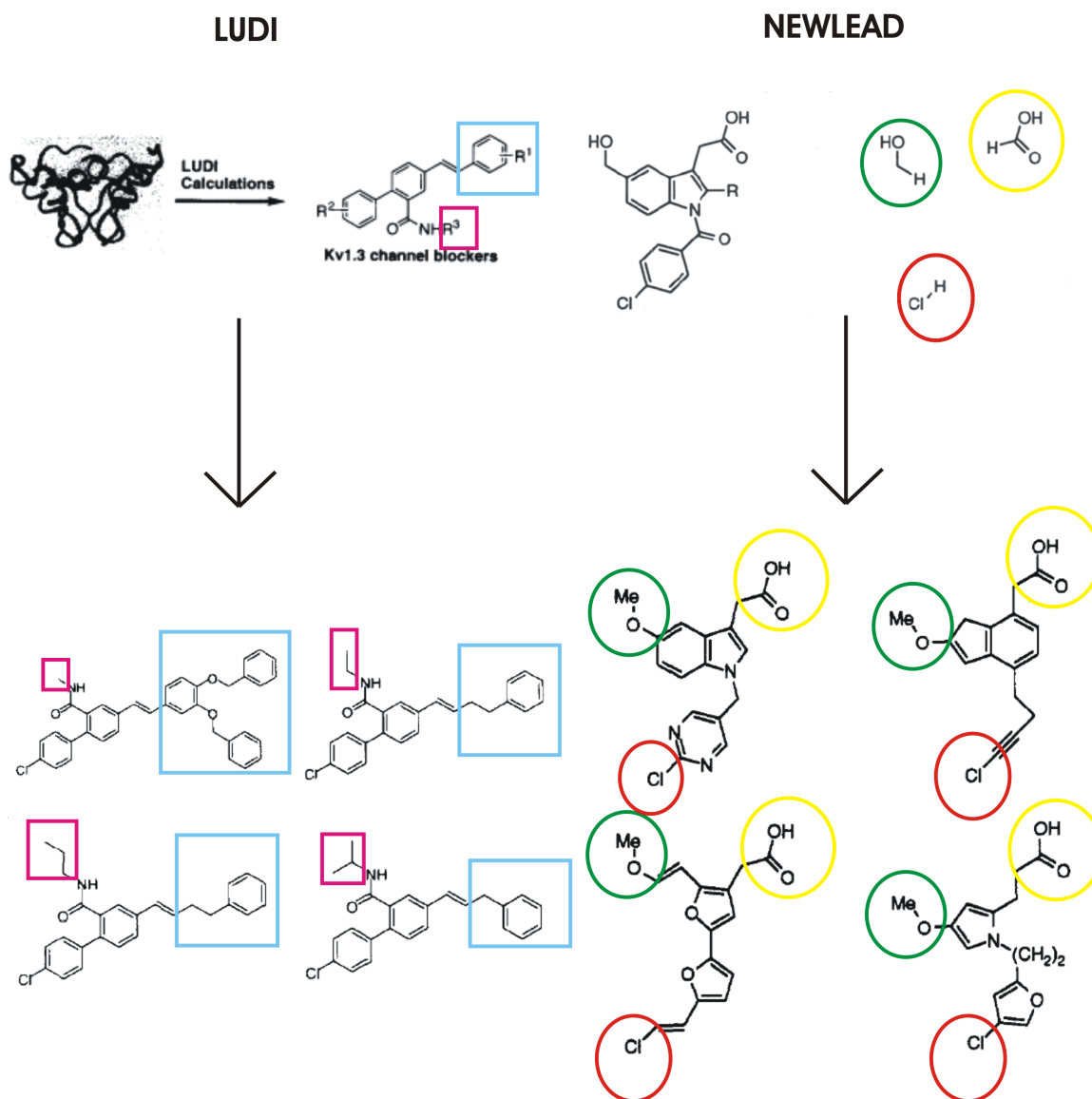


Figure 2.2: The difference between the design methodology of LUDI (Lew and Chamberlin, 1999) and NEWLEAD (Tschinke and Cohen, 1993). This general scheme visualizes the differences in the principles on which LUDI and NEWLEAD function. The variable sites on the molecules built by LUDI are framed in blue and pink and the pharmacophore fragment that NEWLEAD used to build new molecules is encircled in green, yellow and red.

2.1.2. Docking algorithms

Molecular docking designates the following: provided the atomic coordinates of a receptor and a ligand molecule, the docking algorithm should be able to predict their correct bound association. Molecular docking can be subdivided into bound and unbound docking. Bound

docking relates to the methodology of docking a ligand structure into a receptor site where the coordinates of the complex is known. The screening of inhibitors against PfGST is a good example of bound docking since the X-ray crystal structure of PfGST is available. Unbound docking implies the docking of ligands into unbound receptor structures. Unbound receptor structures can consist of native structures, pseudo-native structures or modeled structures (Abagyan *et al.*, 2001; Dym *et al.*, 2002; Stoichet *et al.*, 2002). During a rational drug design strategy the docking step is incorporated to predict whether a ligand will bind to the receptor *in vivo* and provide the 3D structure of the ligand-receptor complex. From an algorithmic perspective a configuration is sought that will minimize a given energy function (Schneidman-Duhovny *et al.*, 2004).

The docking accuracy of different docking programs can be validated by removing the ligand from the active center of a crystal structure, followed by an attempt to correctly dock the ligand back into the active center of the receptor protein. Docking programs can be rated according to the root mean square deviation (RMSD) of the docked ligand against that of the ligand in the crystal structure. The predicted binding affinities from the experimental docking will be correlated with the binding energy from the crystal structure, as well as with data reported from experimental methods. The rationale of this approach is to propose that if a docking program can correctly re-position a known inhibitor to the target protein, this indicates the program's ability to successfully dock novel compounds into the enzyme active site. In this capacity, docking accuracy refers to finding the binding mode of the ligand matching closely the experimental conformation with the lowest energy among a large sampling of potential docking solutions (Hu *et al.* 2004). From the literature the acceptable value for a good docked complex is a RMSD value smaller than 2.5Å (Hu *et al.*, 2004). The following docking suites have been predominantly used in drug design strategies: GOLD (Jones *et al.*, 1997), DOCK (Stoichet *et al.*, 1993), FlexX (Rarey *et al.*, 1996; Rarey *et al.*, 1997), AutoDock (Morris *et al.*, 1998), CDocker, GLIDE, ICM and LigandFit (Abagyan and Totrov, 2001; Brooijmans and Kuntz, 2003; Schneidman-Duhovny *et al.*, 2004).

Ligand flexibility plays a significant role in molecular docking since ligands can adopt various configurations within different proteins. Currently, there are four ways to deal with ligand flexibility, namely conformer generation, incremental construction, Monte Carlo implementations and genetic algorithm implementation. Only those approaches validated in this study to dock ligands into PfGST will be discussed.

Conformer generating algorithms generate an ensemble of pre-calculated conformers based on the original ligand, these algorithms are used in DOCK, EUDOC and FLOG (Anderson and Wright, 2005). This ensemble is then docked into the target protein, each conformer is considered a novel ligand (Schneider and Böhm, 2002). DOCK was validated in the structure-based ligand design of inhibitors against PfGST.

The initial step in the *incremental construction* algorithm consists of breaking down the ligand into smaller chemical fragments (Wolf and Dormeyer, 2003). This approach is implemented within the FlexX docking suite and the algorithm as well as the application were described in Rarey *et al.* (1996a), Rarey *et al.* (1999) and Kramer *et al.* (1999). A selected base fragment is then placed in a previously-defined active center where the rules for steric and electronic interactions are enforced, leaving an ensemble of well-placed base fragments (Rarey *et al.*, 1996b). The construction of the ligand takes place by adding fragments stepwise to the partially constructed ligand as directed by tolerable steric and electronic interactions (Kramer *et al.*, 1999; Wolf and Dormeyer, 2003). Ligand flexibility is measured by individual torsion angles formed by the rotatable bonds. A clustering step is incorporated to reduce the ensemble of molecules, to a set of molecules with maximal binding energy and optimized binding geometry. Hence, an empirical scoring function is used to assess the interactions between the ligands and the receptor molecule (Kramer *et al.*, 1999). From this group of docking programs that uses incremental construction, FlexX was validated in the structure-based ligand design of inhibitors against PfGST.

The *Monte Carlo* approach entails random alteration in the translation, orientation and torsion angles of the ligand in a protein-ligand complex. After each alteration the structure is minimized and the new configuration is evaluated, the most energetically favoured configuration is maintained. AutoDock, MCDock and Prodock conduct Monte Carlo simulations for finding the best spatial orientation of the ligand in the target protein active site (Morris *et al.*, 1998).

The preparation of the protein as well as the ligand prior to the use of docking and scoring applications are equally important for attaining accurate docking results. Therefore, protonation state settings and hydrogen bonding patterns should be considered a serious part of the design strategy (Brooijmans and Kuntz, 2003). Consequently, docking programs cannot be judged based on their overall performance on a mixed data set. A docking program's optimum performance is dependent on the following parameters: crystallographic resolution,

protein family, metal presence, active site topology, occurrence of water molecules as well as ligand and protein flexibility (Erickson *et al.*, 2004; Ferrare *et al.*, 2004; Hu *et al.*, 2004; Krovat *et al.*, 2005). Visual inspection should also play an intricate part of the final round of selection of a docking-based virtual screening study (Golke *et al.*, 2000). This may remove some of the shortcomings of the scoring and docking suites. Rational post-processing steps should be used to enhance docking and scoring studies.

2.1.3. Scoring functions

To avoid confusion, a docking function will be defined as an algorithm used for evaluating different poses of a ligand. Therefore, a scoring function can be defined as an algorithm used to rank the docked poses or compare the affinity of different ligands for the same receptor (Perola *et al.*, 2004). Providing an estimate of the relative binding affinity is the fundamental requirement of a scoring function, i.e. different ligands or different poses of a ligand should be ranked according to the ligand's resemblance to the experimental structure. One should bear in mind that the crystal structure is also affected by experimental errors, therefore, a RMSD value smaller than 2.5Å is acceptable for defining a well-docked complex (Gohlke *et al.*, 2000a). Since these functions are mostly applied in conjunction with a docking strategy it should be sufficiently fast and computationally inexpensive to accomplish high throughput screening (HTS). The ability of a scoring function to determine the complementarity between the protein and the ligand is usually evaluated by its ability to separate a known active compound from a random pool of ligands (Wang *et al.*, 2003; Ferrarra *et al.*, 2004).

In protein-ligand binding it is accepted that electrostatic interactions determine non-covalent interactions that comprise of salt bridges, hydrogen bonds, dipole-dipole interactions, van der Waals interactions as well as interactions with metal ions. Protein-ligand interactions that can be calculated include the non-covalent interactions; internal energies (bond, angle and torsion terms) and ligand strain (internal energy of the bound ligand minus the energy of the unbound ligand) (Schneidman-Duhovny *et al.*, 2004). These descriptors as described by Wang *et al.* (1998), Wang *et al.* (2002) and Schneidman-Duhovny *et al.* (2004) can be used to score a given complex or be used as an exclusion filter. The primary exclusion parameter used in scoring functions is shape complementarity, which leads to a significant decrease in the population size (Mitchell *et al.*, 2001; Mitchell *et al.*, 2004). This exclusion filter is mostly used in docking programs as a first line exclusion filter. Scoring descriptors can be sub-classified into those parameters that define a property for the whole molecule and those that characterize a single atom or specific residue (Ajay and Murcko,

1995; Joseph-McCarthy, 1999). Scoring functions can be classified into three categories: empirical, knowledge-based and force field-based scoring functions.

In principle, the functionality of *empirical scoring functions* is not confined to either a congeneric set of ligands like force field-based scoring functions or specific type of receptor molecule (Wang *et al.*, 2002). Each of the respective components of the scoring function has a physical implication, which can be determined at an accuracy level of ~ 2 kcal/mol (Wang *et al.*, 2002). There are many examples in the literature of empirical scoring function applied to rational drug design such as Bohm's scoring function, ChemScore and XSCORE (Wang *et al.*, 1998; Wang *et al.*, 2002; Wang *et al.*, 2003; Ferrarra *et al.*, 2004). XScore and LUDI are also examples of this type of scoring function and were used to score the interactions that form between the designed ligands and PfGST in this study.

Knowledge-based scoring functions are heuristic in composition; consequently, they are used to solve problems by enforcing the rules that were learned from a sufficient large training set. Several knowledge based scoring functions have been proposed to predict binding affinity e.g. Smog (De Witte *et al.*, 1996), PFM-Score (Muegge *et al.*, 1999), BLEEP (Mitchel *et al.*, 2001; Mitchel *et al.*, 2004), and DrugScore (Gohlke *et al.*, 2000a). These knowledge-based scoring functions differ with regards to the size of the databases used in training, definitions of reference states and the type of molecular interaction descriptors employed (Giordanetto *et al.*, 2004). Due to the absence of a big enough training set and an open source knowledge-based scoring function, this type of scoring function was not incorporated in this study.

Force field-based scoring methods make use of molecular mechanical force fields, which define intramolecular forces (bond, angle and dihedral terms) between bonded atoms. This may also include van der Waals and Coulomb terms that portray forces that exist between non-bonded atoms (Krumrine *et al.*, 2003). Initially, these force field-based scores were calculated by using enthalpies rather than free energies and disregarded the solvent effects and solute entropies. Stoichet *et al.* (1999) improved the above-mentioned algorithms by adding the effect of protein-ligand interactions with the solvent utilizing the implicit solvent model. Minimization is usually needed because the energy landscapes incorporated by the algorithm are too robust. Another deficiency is the propensity of force field-based scoring functions to overestimate the contribution of polar interactions, although desolvation energies are exploited to counteract this (Majeux *et al.*, 1999). Some force field scoring

functions like AMBER-based algorithms are widely used for virtual library screening because these scoring functions maintain a good equilibrium between lipophilic and hydrogen bonding terms, consequently this type of scoring function can be applied on most receptor proteins (<http://www.amber.ucsf.edu>). The AutoDock package implements a force field-based scoring function that is used to determine the binding energy of each ligand to PfGST, this value was also incorporated in ranking the compounds for acquisition and experimental testing for activity against PfGST.

The following scoring functions have been evaluated in the literature as good scoring functions: LigFit module in Cerius2 (LigScore, PLP, PFM and LUDI), CScore module in SYBYL (Gold, F-Score, G-Score, D-Score, ChemScore), the scoring function implemented in AutoDock as well as DOCK and four stand alone scoring functions: BLEEP, DFIRE, DrugScore and XScore (Wang *et al.*, 2003; Marsden *et al.*, 2004; Zang *et al.*, 2005). The scoring functions implemented in AutoDock and LUDI were used in this study as well as the XScore stand-alone function. The choice of scoring function to be used was based on availability. XScore was the only open source scoring function available, and the AutoDock and LUDI scoring functions were implemented in applications used elsewhere in this study.

In order to improve the general performance of scoring functions, it is generally a good idea to use different scoring functions that all estimate a property independently. The property should relate to binding energy or RMSD for all the configurations. Although these scores might have different scales, the average values or summing of ranks should increase the discrimination ability of the scoring function. Since most of these scoring functions are used to rank the output from docking studies, consensus scoring has a lot to offer (Clark *et al.*, 2002). Like any other scientific methodology, *de novo* drug design also has some weak points. Firstly, synthetic feasibility is seldom considered during the design process, which will complicate the testing of compounds later on in the development process. Secondly, due to the novelty of the *de novo* designed structure classes, binding affinity predictions can be less accurate and this can influence the potency of the drug (Homna, 2003).

2.2. Goals

From Figure 2.1, one can deduce that this process forms a perpetual cycle of designing, testing, improving absorption, distribution, metabolism, excretion and toxicity (ADMET) properties followed by testing again. Structure-based ligand design ultimately aims to identify hit and lead compound that can be used to broaden the chemical horizon. Virtual screening of molecules was used to reduce the number of molecules to be tested. For *Chapter 2* the followings outcomes were set to design ligands that will comply with these characteristics.

- Use prior structural information of the protein structure to design compounds with 3D complementarity to the target protein binding site (Terstappen and Reggiani, 2001) by using LUDI and NEWLEAD to design ligands. LUDI identified suitable space for hydrogen bonds and hydrophobic pockets, then matched fragments to these sites (Sottriffer and Klebe, 2002). NEWLEAD added linker molecules to preplace building blocks to satisfy all key interaction sites, linkers were selected with the objective to have favourable interactions with the protein (Schneider and Fechner, 2005).
- Non-covalent bonds with *key amino acids* play a major role in reducing ligand numbers for testing because they define strong and specific requirements for protein-ligand binding (Schneider and Fechner, 2005). Generate a functional map of the interactions with *key amino acids* by analyzing these ligands in the active site of PfGST. Use AutoDock to guide the placement of these ligand in the binding sites. Molecular docking evaluates the binding potential between the designed ligands to reduce the library size and enhance the hit rate (Chen *et al.*, 2004).
- Use LIGPLOT to visualize the non-covalent intermolecular contacts between the ligand and the protein.
- Leadlike molecules bind non-covalently with a high affinity to the protein target (Rishton, 2003). A consensus score, as predicted by AutoDock, LUDI and XScore, provides an indication of the binding affinity the ligands will have for PfGST. Prioritize ligands for testing by ranking according to affinity scores.
- Used *in silico* filters to remove ligands that are unsuitable for drug development.
- Other important characteristics that should be considered in designing a leadlike molecule were metabolic stability, selectivity, toxicity, half-life and minor addictive potential, unfortunately none of these characteristics could be modeled *in silico*.
- Once all the compounds had been ranked by *in silico* methods, screen them for acquisition or synthetical accessibility.

2.3. Materials and Methods

The 3D crystal structure used in all the design, molecular docking and scoring studies was the *P. falciparum* glutathione S-transferase (PDB ID code 1Q4J) with S-hexyl glutathione co-crystallized in the active site (PDB ID code GTX). Although 1Q4J (2.2Å) has a lower molecular resolution than 1OKT (1.9Å), 1Q4J was preferable because there was a ligand present in the active site (GTX). Hence, there was an inhibited as well as an uninhibited 3D structure of PfGST available. The coordinates used for GTX were also extracted from the Protein Data Bank. The forcefield used through the duration of this study was, the CFF91 forcefield of *InsightII* (Accelrys), although the force-field used internally by *AutoDock* was derived from an *AMBER* type forcefield.

2.3.1. Software validation

Docking studies using *FlexX*

The ligand, GTX, was built and converted into *SYBYL mol2* file format using *Insight II* (Accelrys). The formal atomic charges on GTX were assigned using the *ASSIGN_FORMAL_CHARGES* flag of *FlexX*. None of the torsion angles were fixed, therefore the ligand was marked as flexible. The preparation of the receptor was composed of two steps. Firstly, defining the binding site and secondly, creating the receptor description file. The PfGST active site was defined by a placed ligand in the active site, the GTX present in the 1Q4J.pdb receptor file. The receptor description file needed to contain the following information: protein coordinates in PDB format, active site definition, presence of all the HIS amino acids in the active site, heteroatoms and hydrogen torsions. Given the ligand and receptor information, *FlexX* selects a base fragment from the ligand and places the base fragment inside the active site of the enzyme. During an iterative process the ligand is rebuilt in the most energetically efficient orientation. The scoring function used in *FlexX* is a Davidson-Fletcher-Powell (DFP) algorithm. The output of each *FlexX* run was a table with each ligand's name with the RMSD score and the rank of the specific ligand based on RMSD. There was also a PDB file containing 3D coordinates of each ligand created during the docking process.

Docking studies using *DOCK*

InsightII was used for hydrogen addition and charge assignment and the receptor was consequently converted into *SYBYL mol2* format. A grid box was defined by using the *GRID* module. The grid defined the steric and electrostatic properties at each grid point to increase the scoring rate of ligand orientation during the *DOCK* run. The molecular surface was created

by the **MidasPlus** package (Computer Graphics Laboratory, University of California, San Francisco), based on the already-placed GTX molecule. This binding surface definition was used by the **SPHGEN** module to place spheres that could fill the active site of the receptor. The sphere placement was optimized manually by visual inspection in **InsightII**. The dock run was performed with **DOCK 5.1.1** on a Silicon Graphics Octane system.

Docking studies using AutoDock

All the compounds from **LUDI** and **NEWLEAD** were built and geometrically optimized using **InsightII** (Accelrys). Polar hydrogen atoms were added to the protein and the ligand using **InsightII** and the partial charges were assigned using **PDB2PDBQ** (from J.L. Pellequer, The Scripps Research Institute La Jolla, CA) using the **CFF91** forcefield. The energies were calculated using **AutoGrid**. The grid consisted of 60x60x60 grid points with a spacing of 0.375Å centered at the active site of the protein, which was the sulphur atom of the co-crystallized inhibitor, GTX. For each ligand docked into PfGST, **AutoDock** was run using 50 Lamarckian genetic algorithm cycles with 150 000 energy evaluations. The remaining parameters used by the genetic algorithm were set to their default values. As an output **AutoDock** produced a cluster of the 10 top-ranked ligand conformations each with two energy values (docked energy and binding free energy).

2.3.2. Ligand design

When investigating the G site of the PfGST enzyme the following amino acids were identified as important for ligand binding: Tyr 9, Gln 71, Ser 72, Val 59, Lys 15, Glu 58 and Asp 105' (from the other subunit) (Perbandt *et al.*, 2003). Therefore, it would be advantageous for binding affinity if the novel inhibitors could interact with these amino acids as well. Consequently, there was a need for an existing inhibitor that utilized these amino acids to bind to PfGST. Unfortunately, the availability of PfGST inhibitors was very limited and the only structural information available for a PfGST inhibitor was the crystal structure of PfGST with GTX. Therefore, it was decided that GTX would be used for the modification. Secondly, GTX was a derivative of glutathione that was the primary substrate to all GST and would definitely bind in the G site. Modifications made to the GSH moiety could enhance binding affinity of the inhibitor. The modifications made on the hexyl part of GTX or any other fragment added to GSH to replace the hexyl moiety could increase the specificity of the inhibitor towards PfGST because this part would putatively bind to the variable, more solvent-exposed H site.

2.3.2.1. Modifications of S-hexyl glutathione using LUDI

During the binding of GTX to PfGST many of the above-mentioned amino acids are involved in binding GTX, consequently the modification of GTX seemed to be a good starting point. The purpose of using LUDI was to identify modifications that may enhance activity and synthetic accessibility or create new candidate lead molecules. Variations on GTX with regards to the GSH moiety were needed to create a molecule that will compete competitively with glutathione but will bind specifically to the PfGST and not to any of the human GST enzymes. The dimeric 1Q4J.pdb was loaded into *InsightII* (Accelrys), the waters as well as both the inhibitor molecules GTX were removed. A single GTX was placed into the A-subunit of the enzyme. Since the active sites are identical only the A-subunit was utilized in this study. Hydrogens were added to the protein receptor and the inhibitor by *InsightII*. The partial charges as well as the potentials were assigned within *InsightII*. Minimization was performed in *InsightII* with a CFF91 forcefield. For LUDI to function optimally the enzyme and the inhibitor should not be merged to form one object. Hence, PfGST enzyme and the GTX inhibitor molecules were separate objects in *InsightII*.

The LUDI command in *InsightII* was selected to activate the LUDI module. LUDI was used in linkage mode, meaning that fragments from the BIOSYM library that would bind to the receptor would be chosen. Accelrys distributes this library of about 1100 small molecules derived from Bohm's original fragments for both receptor and link modes. Typical fragments are benzene, naphthalene, acetic acid, benzoic acid, indole, piperidine, etc. The topologies of these fragments were stored in MDL Mol format, with definitions of link sites for each fragment. Another requirement of this mode was that the newly selected fragment must be able to form the necessary hydrogen bonds (in 3D space) with the existing inhibitor molecule or substitute an existing group on the inhibitor. The parameters in the LUDI module were set to default, specifying how a fragment must fit into the receptor site. The targeted mode allowed LUDI to assume that any of the hydrogen atoms positioned on GTX could be defined as link sites. The electrostatic filter was activated to ensure that no unacceptable electrostatic repulsions were found. Fragments were treated as flexible, this was controlled by the amount of rotatable bonds that were specified. The degree of rotation depended on the periodicity of the dihedral angle of the bonds. The linkage parameter that regulates the amount of fragments that can be added or substituted on GTX, was varied between zero and two fragments per novel molecule. In order to prevent logarithmic growth of the library of putative inhibitors the invert parameter was switched off to remove unlikely enantiomers from the library. The center of search was filled with the coordinates of the sulphur atom

of GTX and a radius of 8Å was used. This determined the atoms/fragments that could be replaced as well as the size of the active site in which LUDI could find fragments that would bind the receptor. Link sites needed to be specified for the replacement or addition of the newly defined fragments are indicated in Figure 2.3. The resulting fragment(s) of each run were linked to GTX. The result was a library of molecules that had different combinations of fragments added and/or substituted on GTX. During the assembly of the LUDI library all the duplicate molecules were removed.

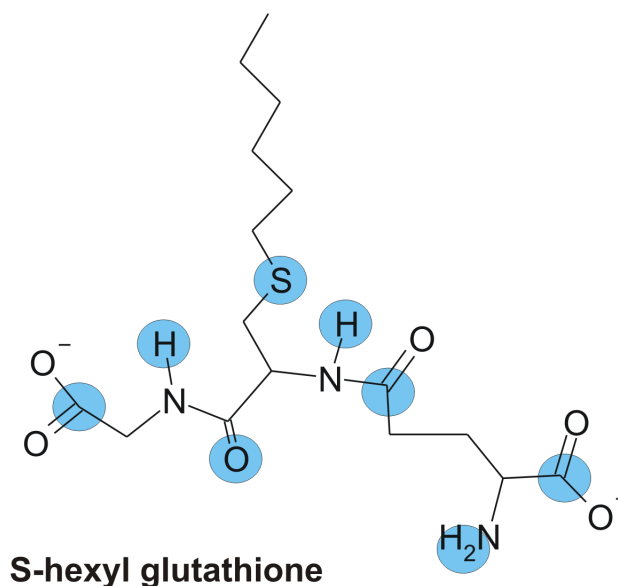


Figure 2.3: The structure of GTX with the atoms which were replaced or where addition occurred (indicated by a blue circle). Allowed addition or replacement sites have to be supplied for the LUDI program.

2.3.2.2. Knowledge-based drug design using NEWLEAD

NEWLEAD was used to generate ligands that will conform to a given pharmacophoric pattern. A pharmacophoric pattern was defined as an ensemble of interactive functional groups with a defined geometry that carries the essential features responsible for an inhibitor's biological activity. Typical pharmacophore features are where a molecule has a hydrophobic patch or an aromatic patch or a hydrogen bond acceptor, a hydrogen bond donor, a cation or an anion.

The primary aim was to combine the known pharmacophore features with new linkers to get novel inhibitors, as described by Tschinke and Cohen (1993) (Figure 2.2). Using all the pharmacophore fragments that GTX possesses would limit the number of designed ligands. Therefore the amount of pharmacophore fragments as well as the combination in which the pharmacophore fragments were used, were varied. The program was used on several sets of input fragments in PDB format, each comprising selected pharmacophore fragments obtained from the bio-active conformations of the GTX inhibitor molecule and GSH substrate molecule. During this process the indicated parts on Figure 2.4, in different combinations, were defined as the pharmacophoric fragments. NEWLEAD linked these fragments, consequently ending up with novel ligands. Default parameters were used.

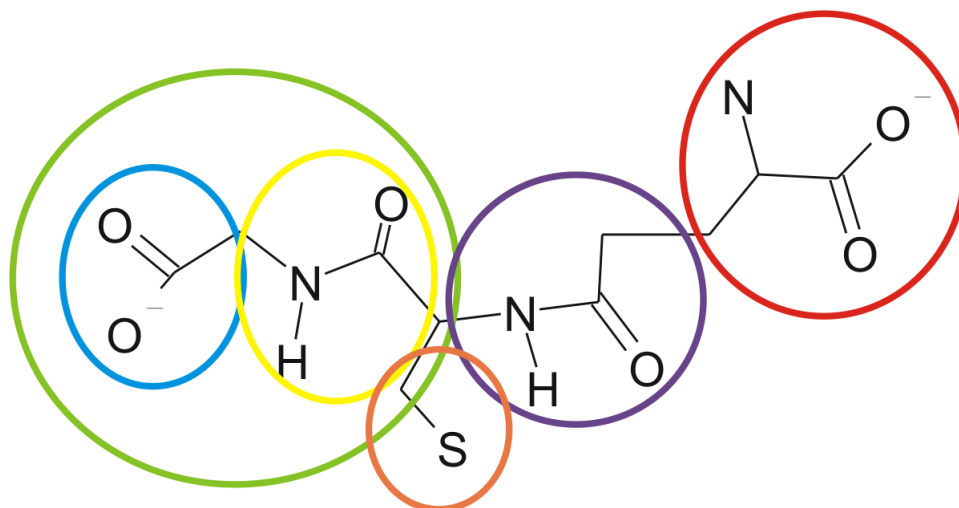


Figure 2.4: **The structure of glutathione with fragments encircled that were used by NEWLEAD as building blocks to form new ligands.** For NEWLEAD to function pharmacophore fragments have to be defined, the pharmacophore fragments defined for glutathione are the atoms encircled in various colours.

NEWLEAD then generated new structures that were chemically unrelated to the reference molecules. This provided an unbiased starting point for the design of new generations of lead structures. The run time was very fast, as only a few bonds were created between pharmacophore fragments since they already had ideal geometries, hence no minimization steps were required. The library of spacers (assembled from small chemical entities like atoms, chains or ring moieties) to connect the fragments was provided with the software. The same pharmacophore pattern can produce various different output molecules due to the number of possible connections of different spacers. The output results were clustered based

on the similarity of the connecting spacers.

2.3.3. Estimation of protein binding affinities using consensus scoring based on results from AutoDock, LUDI and XScore

From the AutoDock output the PDB file of the highest ranked ligand orientations were extracted and used for the determination of binding affinities. The consensus score for the binding affinity between the ligand and the protein was determined using 3 different scoring functions. AutoDock uses a force field-based scoring function where different conformations were ranked according to their predicted free energy of binding given in kcal/mol. The conformation with the lowest free binding energy was used in the consensus score. XScore and LUDI use empirical scoring functions. XScore assumes that the overall free energy changes in a protein-ligand binding process is the summation of the following energy terms: Van der Waals interactions, hydrogen bonding, deformation effect, hydrophobic effect and a regression constant that accounts for the entropy loss due to translational and rotational effects upon ligand binding. Deformation effect was defined by Wang *et al*, 2002, as the energy needed to change the conformations of the ligand and the receptor upon complex formation (binding).

XScore uses three different models to calculate the hydrophobic effect: 1) solvent accessible surface model 2) hydrophobic contact algorithm 3) hydrophobic matching algorithm. Consequently, three different algorithms were used for a single protein-ligand binding affinity prediction. Therefore, the mathematical mean of the binding affinities from the three algorithms was used in the consensus score. The XScore package only accepts ligand files that are in mol2 format, therefore, all the ligands in the library were converted to mol2 files by using OPENBABEL 1.6 (<http://openbabel.sourceforge.net/>). The XScore FIXMOL2 program was used to ensure that the format was correct. The receptor file had to be in PDB format, hence the XScore FIXPDB program was used to ensure that the format of the receptor file was correct. The output of the XScore package was a file with all the ligands ranked according to the mathematical mean of the binding affinity determined. XScore also provided an estimation of solubility in the form of a LogP value for each ligand.

The LUDI scoring function entails a two-term scoring function. The first term computes the quantity (magnitude) and quality (strength) of hydrogen bonds between the protein and the ligand. The second term accounts for the hydrophobic contact area between the protein

and the ligand. This score is correlated with the dissociation constant K_i for the ligand receptor complex (a score of 100 can be correlated with a K_i of 100mM) (LUDI user manual, March 2000). The assembly of ligands in PDB format were loaded into **InsightII** together with the dimeric PfGST enzyme. All the water molecules were removed and hydrogens were added by using the **Modify/Hydrogens** command. All the ligands, GTX inhibitor and the PfGST enzyme were separate objects in **InsightII**. The LUDI module was activated and the return command value set to return the output as a table of scores. The **LUDI/SCORE** module was executed as a back ground job. The output table contained information on hydrogen bond contacts, ionic bonds, rotational score, aromatic score and a final binding affinity score, the latter was used in the consensus score for each ligand.

2.3.4. Hierarchical clustering of molecules based on molecular fingerprints

After the design of the library, a representative subset of ligands was needed for biological testing. Ward's hierarchical clustering (Ward, 1963) as implemented in the **Chemaxon** package (Budapest, Hungary) was used to cluster these molecules into conformationally related subfamilies based on molecular fingerprints. All the ligands in MOL2 format were parsed into a single structural data file (**SDF**) containing structural information. **GENERATEMD** (Pearlman and Smith, 1998) is a module from **Chemaxon** that was used to characterize the ligands in the **SDF** file based on their chemical, structural as well as pharmacological properties, known as molecular fingerprints. Molecular fingerprints are a set of binary values associated with two and three dimensional molecular structural data. Ward's clustering module uses Murtagh's reciprocal nearest neighbour algorithm to form hierarchical clusters of ligands on the basis of similarity with respect to their molecular fingerprints (Kelly *et al.*, 1996). The representing scaffold from each cluster was extracted.

2.3.5. Investigation of the *in silico*-determined properties of the four commercially available compounds

The next step in this structure-based design strategy would be to screen the best-scoring inhibitors experimentally, hence to test their activity against PfGST. Unfortunately during the design process of ligands that would inhibit PfGST, synthetic accessibility was underestimated and that led to an increase in the complexity of the *in vivo* and *in vitro* screening problem. Due to the time limitations of a masters degree none of the compounds could be synthesized as required. The package **SciFinder** (www.cas.org) was used to find out which of these designed compounds were commercially available. Of the 96 compounds

designed only twelve were commercially available. Due to financial constraints only four compounds could be obtained, S-hexyl glutathione (GTX), L-3-aminomethylphenylalanine (LAP), ethyl 4-amino-2- [(4-ethoxy-2, 4-dioxobutyl) thio]-5-pyrimidine carboxylate (EDP) and 3-(2-Naphthyl)-D alanine (NDA). These four compounds were chosen to represent four different clusters of molecules in order to explore the inhibitory capacity of these scaffold structures. Visualization of their docked orientation (provided by `AutoDock`) as well as complete scores from `LUDI` and `XSCORE` will be given in the following section.

The binding interactions were visualized in `LIGPLOT`. `LIGPLOT` automatically generates schematic diagrams of protein-ligand interactions for a given PDB file that contains both the ligand and receptor. `HBPLUS` is the module that calculates the hydrogen bonds and hydrophobic contacts for `LIGPLOT`. For the complete interaction map the first map was drawn of the ligands in the active site as proposed by `AutoDock`. The receptor-ligand structures were subjected to 1000 steps of simple minimization in `InsightII` using the Polak-Ribiere algorithm and another map was drawn. This iterative process of minimization and map drawing continued until 1,000,000 steps were done. A global binding map was constructed based on all the results from the various maps.

2.4. Results

2.4.1. Validation of software

As a first step, the validation of different docking programs took place by removing GTX from the active site of the enzyme, and attempting to return it correctly by docking. `AutoDock`, `FlexX` and `DOCK` were used. This was done to establish which docking program could correctly reposition a known inhibitor to the target protein, thus indicating the program's ability to successfully dock novel compounds into the enzyme active site. Docking programs were rated according to the RMSD of the docked ligand against that of the ligand in the crystal structure.

A RMSD value is a statistical measure computed to quantify the magnitude of variance between the atomic coordinate set of docked ligand and the ligand present in the crystal structure. A RMSD value of less than 2Å as well as catalytic site prediction accuracy defined a successful docking event. When comparing the RMSD values of the three docking suites, `AutoDock` and `FlexX` were able to reproduce a spatial orientation of GTX that was

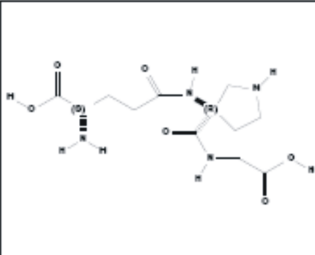
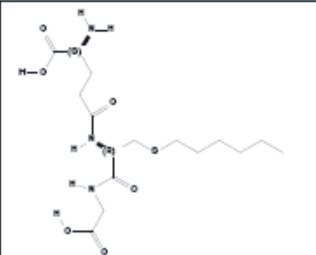
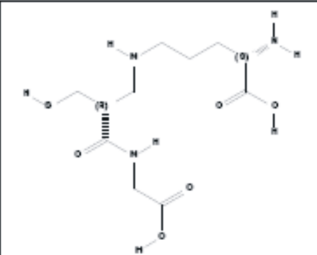
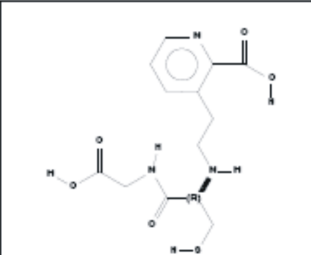
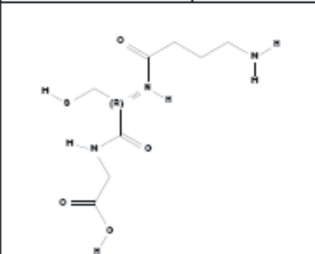
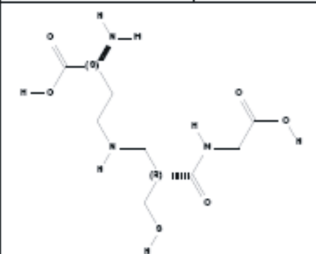
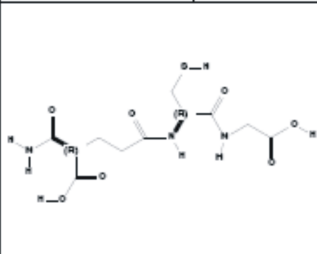
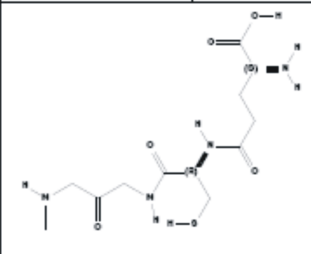
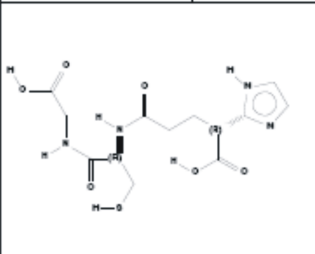
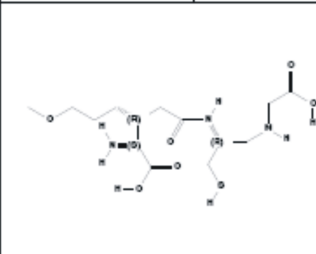
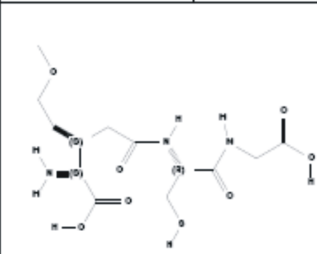
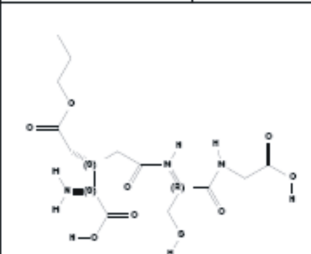
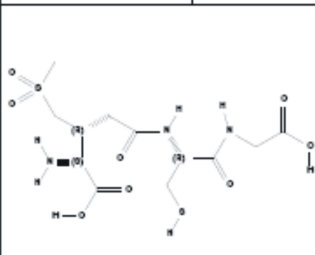
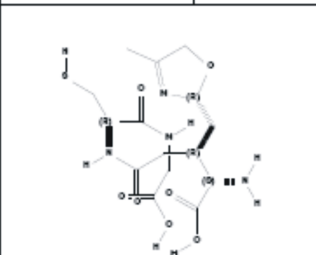
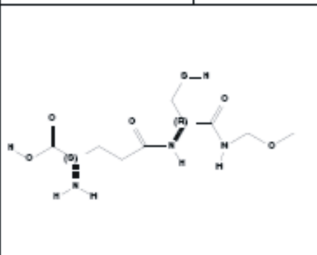
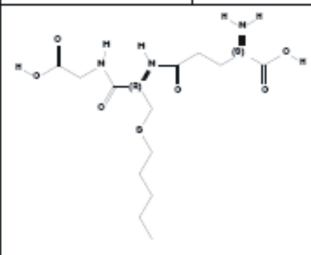
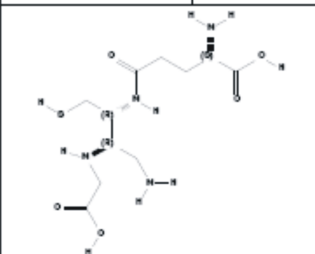
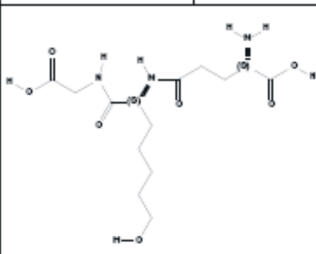
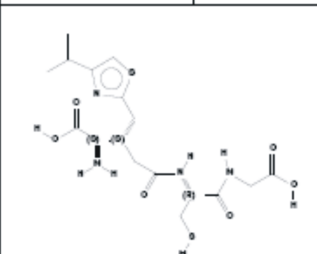
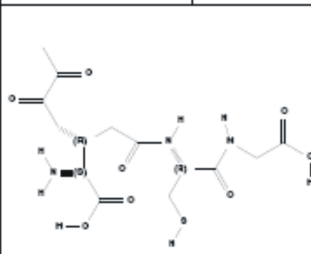
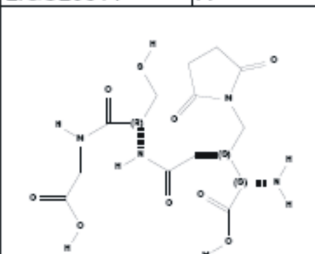
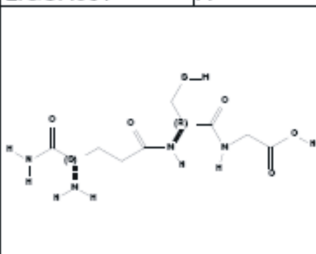
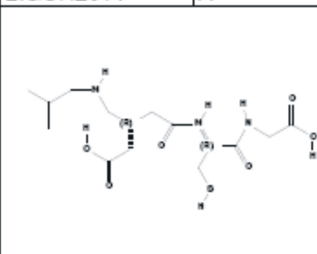
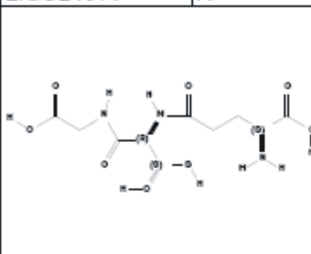
in good agreement with the crystallographic data. A comparison between the top 5 ranked ligands of **AutoDock**, **FlexX** and **DOCK** is shown in Table 2.1. **DOCK** was highly unsuccessful in docking the ligand into the correct active site, as evidenced by the high RMSD values given. The energy values provided by the docking suites could not be used as a measure to compare the docking suites, since all three suites used a different type of scoring function. These energy scores can be used to rank the different orientations of the different ligands in further studies, therefore, the first ranked ligand often has a higher energy value than lower ranked ligands. Based on the high RMSD values and the inability to place the ligand into the accurate catalytic site, **DOCK** was rendered unsuitable to be used as a docking suite in the rest of this study. When comparing the RMSD values of **AutoDock** and **FlexX**, **AutoDock** scored better on average than **FlexX**. Not only based on RMSD values but also on computational time did the incremental construction algorithm performs poorer than the genetic algorithm when applied to PfgST.

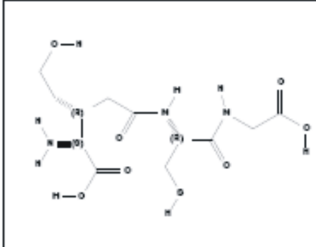
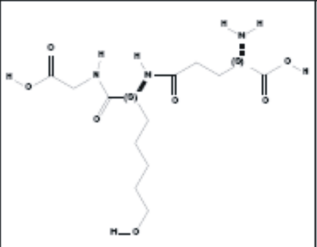
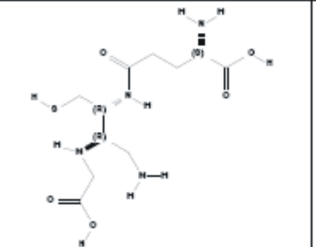
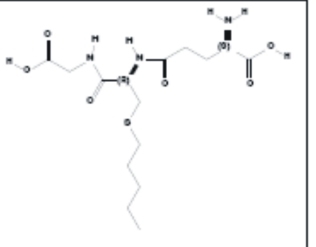
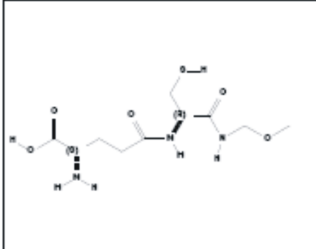
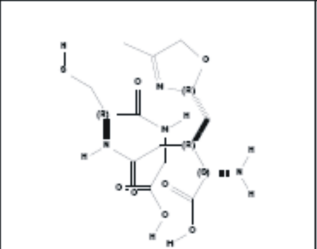
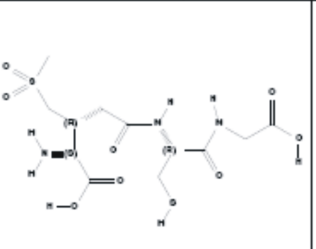
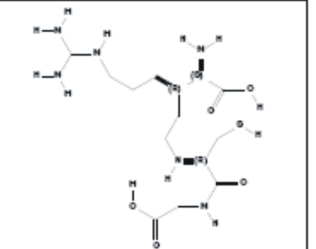
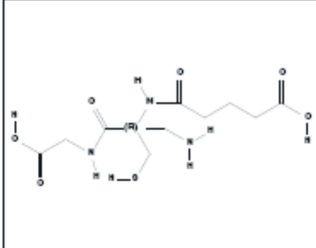
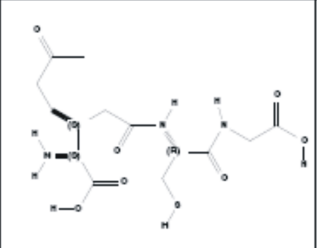
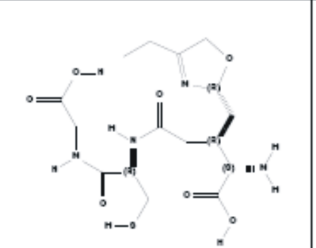
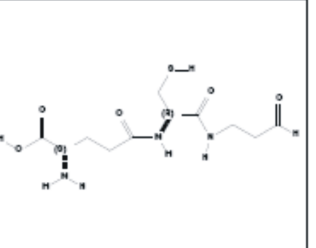
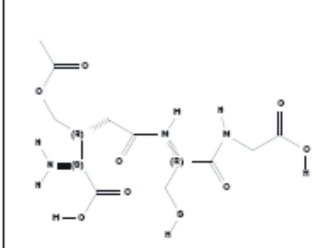
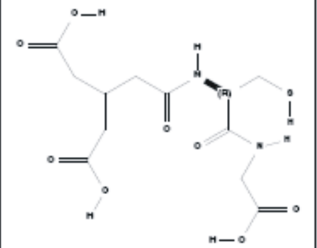
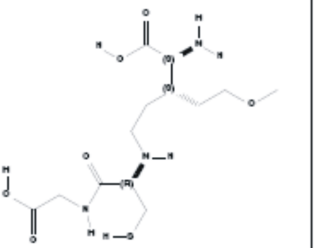
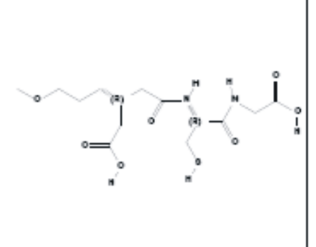
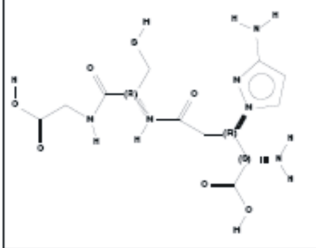
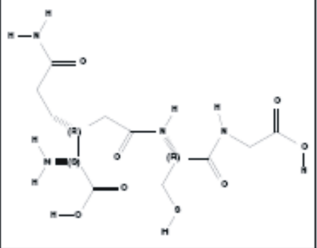
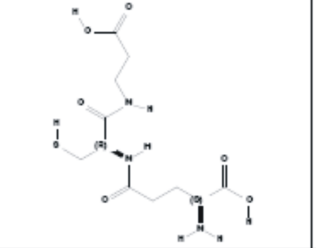
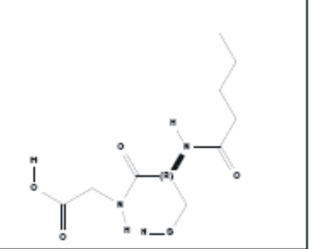
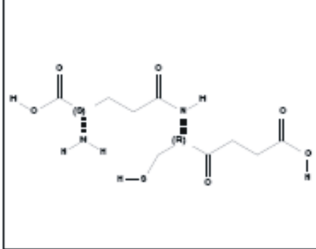
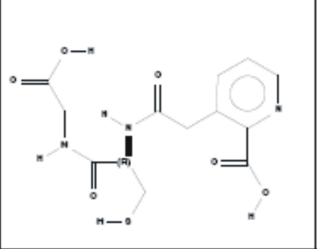
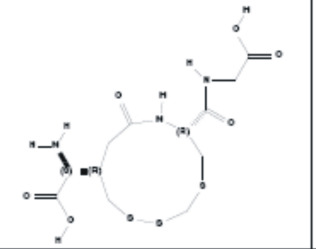
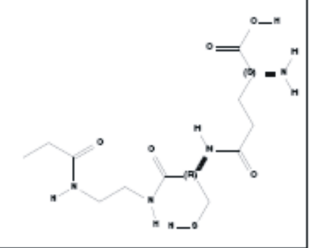
Based on **AutoDock**'s performance when comparing RMSD values, a good correlation of the energy values with the ranking of the ligands would play a crucial part in the next part of the study where different ligands would be ranked according to their affinity towards PfgST. Since libraries in the region of hundreds of compounds needed to be docked and scored, a docking suite was required that needed little or no user input and was computationally inexpensive. **AutoDock** was the only docking suite that complied with these prerequisites. **DOCK** needed too much user input for active site definition and grid calculations, therefore large scale compound screening would have been very time consuming. The definition of fragment placement increased the computational cost of **FlexX** and hence **DOCK** and **FlexX** did not meet these requirements.

Table 2.1: Comparison of **AutoDock**, **FlexX** and **DOCK** with regards to the energy values and RMSD values of the top 5 ligands. The RMSD values were determined between the atomic coordinate set of the docked ligands and the ligand present in the crystal structure.

	AutoDock		FlexX		DOCK	
Rank of Ligand	Energy value	RMSD	Energy value	RMSD	Energy value	RMSD
1	-14.96	2.54	-28.62	2.7	-48.48	14.36
2	-14.68	2.52	-28.38	2.91	-47.15	7.16
3	-14.22	1.88	-28.23	2.78	-46.06	14.64
4	-13.18	2.19	-27.24	2.8	-45.21	8.17
5	-11.58	1.54	-28.24	2.75	-39.67	24.28

2.4.2. Ligand design

			
LIGC1101	A-	GTX11	A-
			
LIGC3001	A-	LIGC66	A-
			
LIGMK6	A-	LIGS3401	A-
			
LIGS8101	A-	LIGSH901	A-
			
LIGU20011	A-	LIGUA601	A-
			
LIGSX0011	A-	LIGU20 2011	A-

			
LLIGSF301	A-	LIGUA6011	A-
			
LIGSO1011	A-	LIGSH9011	A-
			
LIGU1901	A-	LIGSR901	A-
			
LIGS6201	A-	LIGS3801	A-
			
LIGME501	A-	LIGCA9	A-
			
LIGC2401	A-	LIGC30 101	A-
		LIGC3301	A-
		LIGM1301	A-

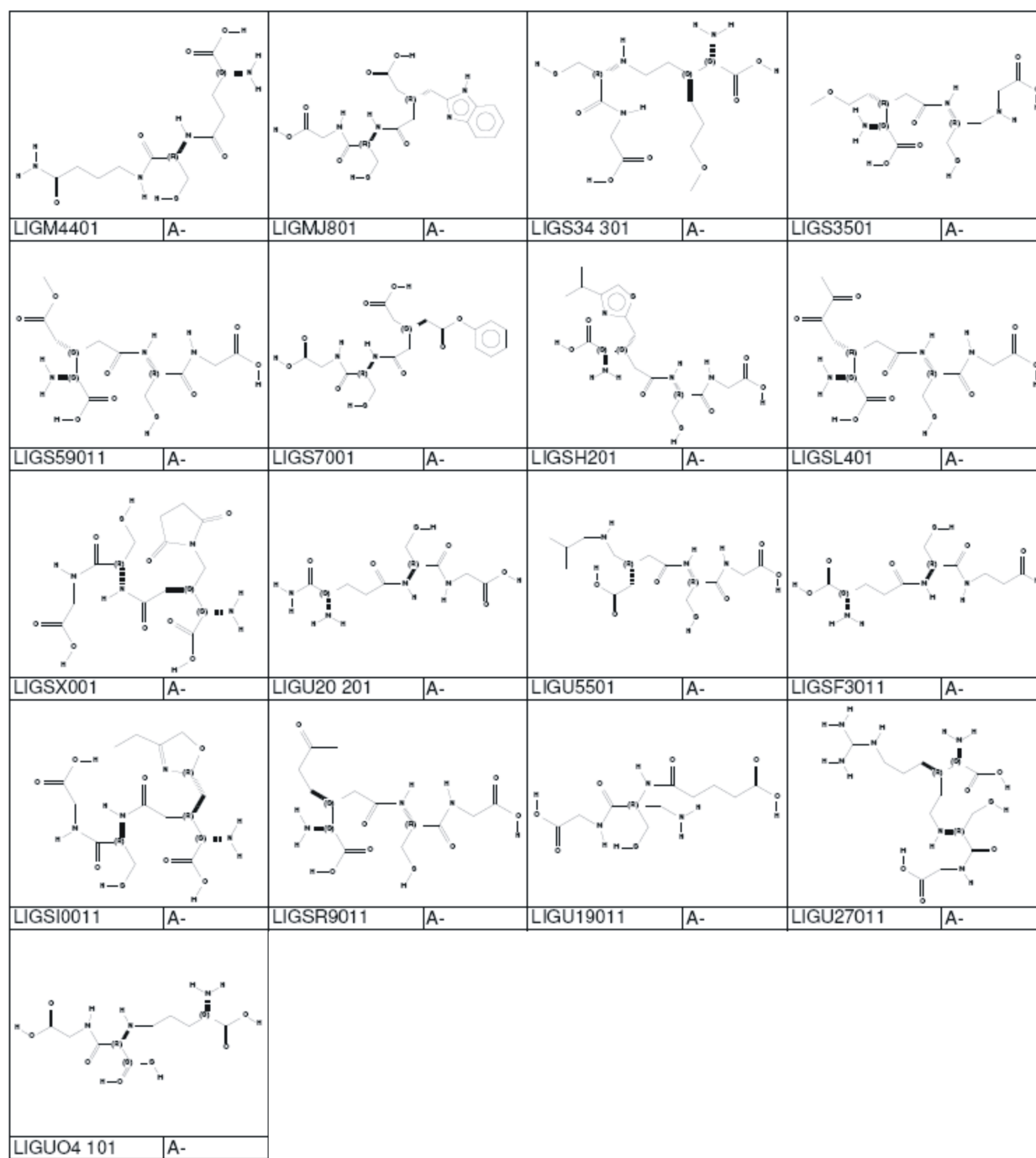
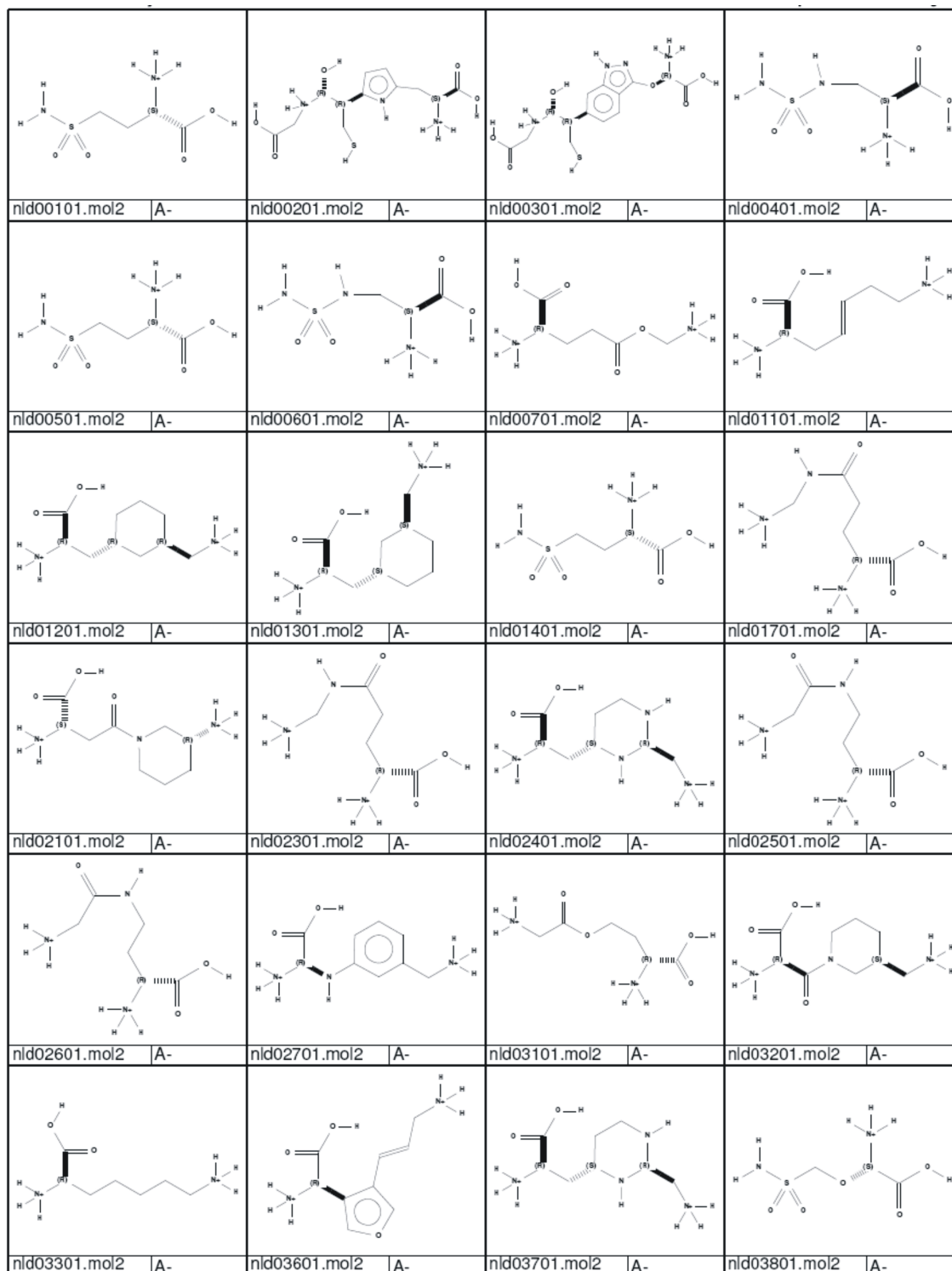


Figure 2.5: Ligands designed by LUDI, before the duplicates were removed.

From the initial LUDI runs, 65 possible compounds were identified. The library is shown in Figure 2.5. Of the 65 compounds 14 were removed due to duplication, consequently 51 compounds were generated by using modifications of GTX. These 51 compounds were the

first addition to the focused library of PfGST inhibitors. NEWLEAD provided 46 putative inhibitory ligands to PfGST, there were no duplicates and all these compounds were added to the library of putative inhibitors. This assembly shown in Figure 2.6.



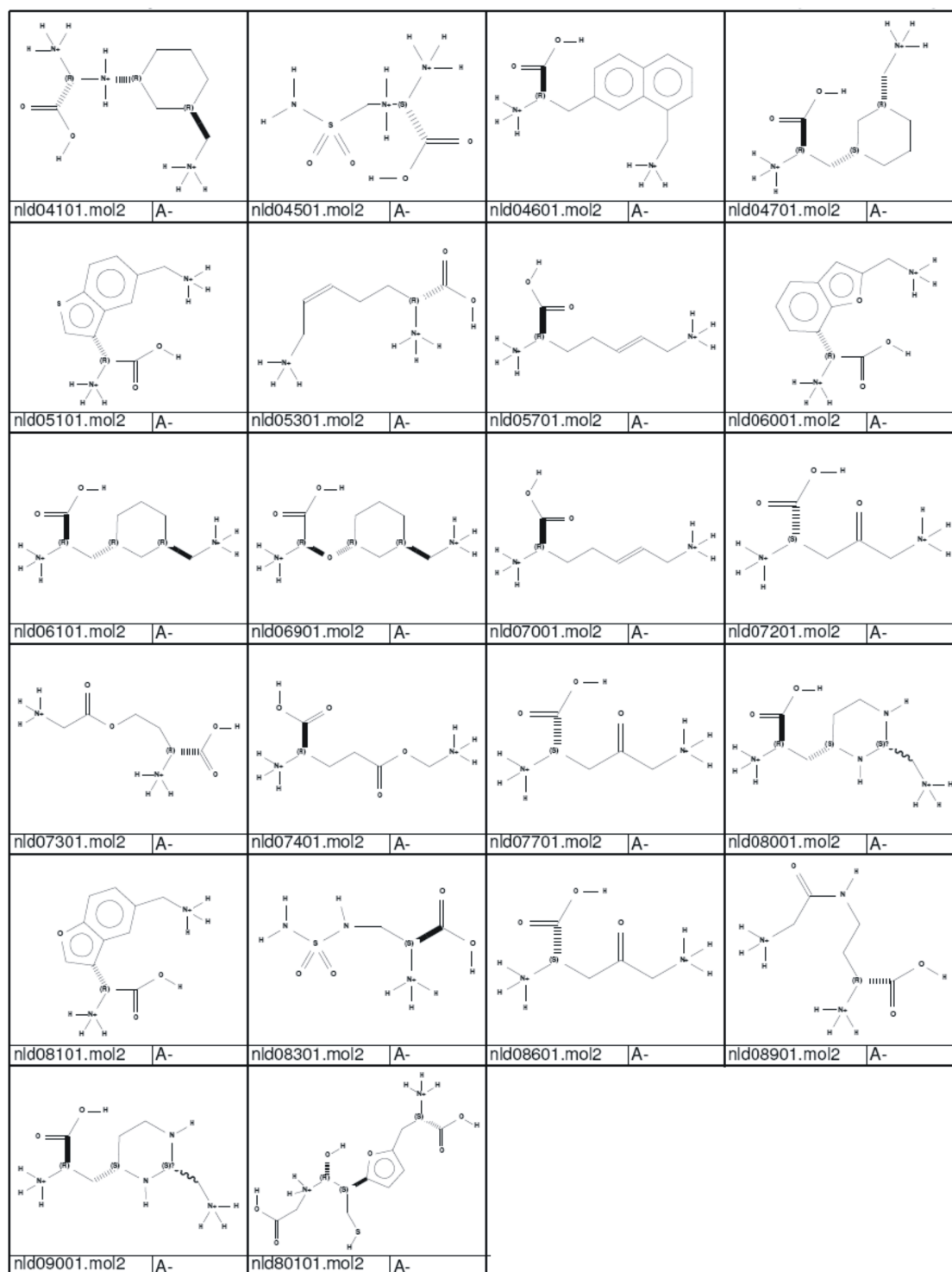


Figure 2.6: The 46 ligands designed by NEWLEAD.

Ligand-based library design can be subdivided into approaches that are based on similarity and those based on diversity. The library designed by LUDI and NEWLEAD were similarity

based. A diversity approach would be near impossible because the binding mode of only one ligand was known, which necessitates the exploration of a vast chemical space. By using similarity as a foundation the properties important for ligand activity were captured.

2.4.3. *In silico* screening: Estimation of protein binding affinities using consensus scoring based on results from AutoDock, LUDI and XScore

The aim of *in silico* screening was to use computational tools to predict the ligand's activity that can be related to inhibitory capacity thus ranking ligands to be carried forward in a discovery pipeline. Due to the size of the library that needed to be screened the high throughput version of AutoDock was used (kindly provided by the co-author of AutoDock Ruth Huey, The Scripps Research Institute, La Jolla, CA, USA). The output of this was an estimation of free energy of binding as well as an estimate of the docked energy (both of these were given in kcal/mol). In a wet laboratory experiment it is standard practice to include a positive control to be able to benchmark the work. In this study GTX was used as a benchmark and positive control to validate the inhibitors *in silico* as well as in the wet lab. AutoDock provided the following for all the ligands: binding energies that ranged between -3.11 kcal/mol and -9.67 kcal/mol. The docked energy that ranged between -3.85 kcal/mol and -15.06 kcal/mol; where GTX scored -9.67 kcal/mol and -15.01 kcal/mol, respectively.

The AutoDock scores for the ligands designed by LUDI were on average lower than the ligands designed by NEWLEAD. The reason for this may be size since the LUDI inhibitors were on average larger than those designed by NEWLEAD. The increase in size may cause the ligands not to fit either into the G site or the H site and will therefore be placed in a space that did not form part of the catalytic site, leading to very low docking energies. The LUDI scores of this library ranged from 22 kcal/mol to 807 kcal/mol with GTX scoring 587 kcal/mol ($K_i \sim 1\mu\text{M}$). The LUDI score can be interpreted as $100\log K_i$, where $\log K_i$ is the inhibition constant. The magnitude of the LUDI score is highly dependent on the number of hydrogen bonds that would form, appropriately the ligands with lower numbers of hydrogen bonding sites scored worse than those ligands with more hydrogen bond sites. The binding affinity predicted by XScore for GTX had a mean of 6.03 where the range of the other compounds was between 3.68 and 6.04 pK_d units.

Using average values of the estimates from different scoring functions should be a better predictor than a single scoring function. Unfortunately, the various scores from the docking and scoring functions have different scales, hence no average could be determined. Conse-

quently, it was decided to create a consensus score by the summation of all the scores. The energy value given by **AutoDock** was made a positive in order for it to contribute to the score. **LUDI** scores and **XScores** were added since they were positive already. The **LUDI** score was divided by 100 to equalize the weight it carried with respect to the other scoring functions. The **LogP** values were also incorporated. As a result, the consensus score for **GTX** was 27.98 and the range of the consensus score was 5.6 to 79.06 No unit was given to the consensus score since absolute values were used.

The ligand library was ranked according to consensus score values, the higher the consensus score the better the ligand would bind to the enzyme. These ligands were prioritized for purchase or synthesis and biological testing. Often filters like hydrogen bonds, electrostatic interactions, stability of the configuration, lipophilicity and **ADMET** predictions are used to choose the starting scaffold.

2.4.4. Hierarchical clustering of database molecules based on molecular fingerprints

Ward's clustering produced 8 clusters. The maximum common substructure (scaffold) from each cluster can be seen in Figure 2.7. Ward's clustering also produced singletons (clusters with only one member), 10 of these ligands were manually curated and added to cluster 8. Cluster 1 was the smallest cluster with only 3 members, cluster 4 was the biggest cluster with 23 members (including **GTX**) and cluster 8 had 16 members excluding all the singletons. Cluster 8 had 26 members after the singletons were added. The cluster distribution can be seen in Table 2.2. Of these ligands, twelve were commercially available as represented in Figure 2.8. The distribution of 11 of these inhibitors is shown in Table 2.2 but **N-acetyl-L-methionyl,methyl ester-glycine** was a singleton. Due to financial constraints only four could be obtained, **L-3-aminomethylphenylalanine (LAP)**, **ethyl 4-amino-2- [(4-ethoxy-2, 4-dioxobutyl) thio]-5-pyrimidine carboxylate (EDP)** and **3-(2-Naphthyl)-D alanine (NDA)**. **S-hexyl glutathione (GTX)** was purchased as well to be the positive control. **GTX** belongs to cluster 4 that represent the group of glutathione analogs, like homoglutathione. Cluster 1 was the smallest cluster with a pyrimidine-like scaffold. **EDP** belonged to this cluster and was selected as a representative for the biological assay. A benzene-like scaffold was observed for cluster 5. **LAP** belonged to this cluster and it was the only commercially available compound belonging to this cluster. **NDA** belonged to cluster 7 with a naphthalene-based

scaffold. Praziquantel (*a Schistosoma japonica* (SjGST) glutathione S-transferase inhibitor) also shared molecular fingerprints with cluster 7.

Table 2.2: The cluster distribution of the ligands in the library as well as the commercially available compounds.

Cluster number	Members	Commercially available compounds	Biologically tested compounds
1	3	2	EDP
2	13	None	
3	7	2	
4	23	2	GTX
5	10	1	LAP
6	6	None	
7	7	1	NDA
8	26	3	

2.4.5. *In silico*-determined properties of the four commercially available compounds

Before the assays commenced a more detailed analysis of the computationally-determined properties of the four commercially available compounds was done. Table 2.3 and Table 2.4 were created based on computationally-derived ligand binding data. An in depth analysis of the intermolecular interactions that formed between the ligands and PfGST was done.

Comparison of the commercially available inhibitors revealed that NDA had the highest docked energy of -9.58 kcal/mol followed by LAP (-9.16kcal/mol) and EDP (-8.54 kcal/mol). The binding affinity score given by LUDI is proportional to an inhibition constant value, a measurement of activity. The LUDI score predicted NDA to have the higher binding affinity followed by LAP and EDP. LogP was used to estimate the water solubility of the three inhibitors. LAP was estimated to be water soluble but solubility problems were predicted for EDP and NDA. XScore produced binding affinity predictions for the three inhibitors. NDA was predicted to have the highest affinity for PfGST (5.93) followed by LAP (5.41) and EDP (4.83).

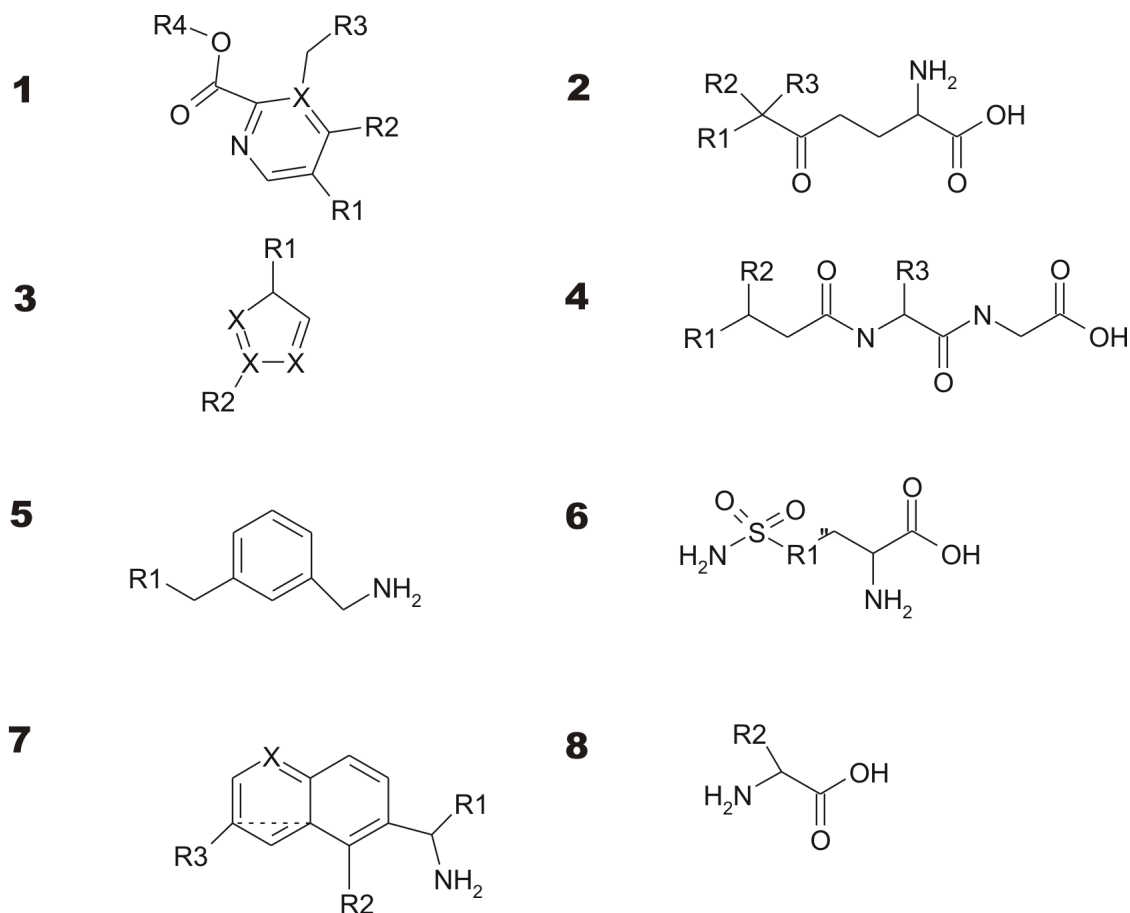


Figure 2.7: **The maximum common substructure (scaffold) from each cluster produced by Ward's hierarchical clustering.** Cluster 1 was the smallest cluster with only 3 members, cluster 4 was the biggest cluster with 23 members (including GTX) but after the singletons were added, cluster 8 had 26 members .

Table 2.3: The *in silico* derived screening data of the four commercially available compounds.

Compounds	Docked energy	LUDI score	XScore	LogP
GTX	-15.01 kcal/mol	5.87	6.03	-1.07
LAP	-9.16 kcal/mol	5.26	5.41	-0.09
EDP	-8.54 kcal/mol	5.22	4.83	1.17
NDA	-9.58 kcal/mol	5.35	5.93	2.05

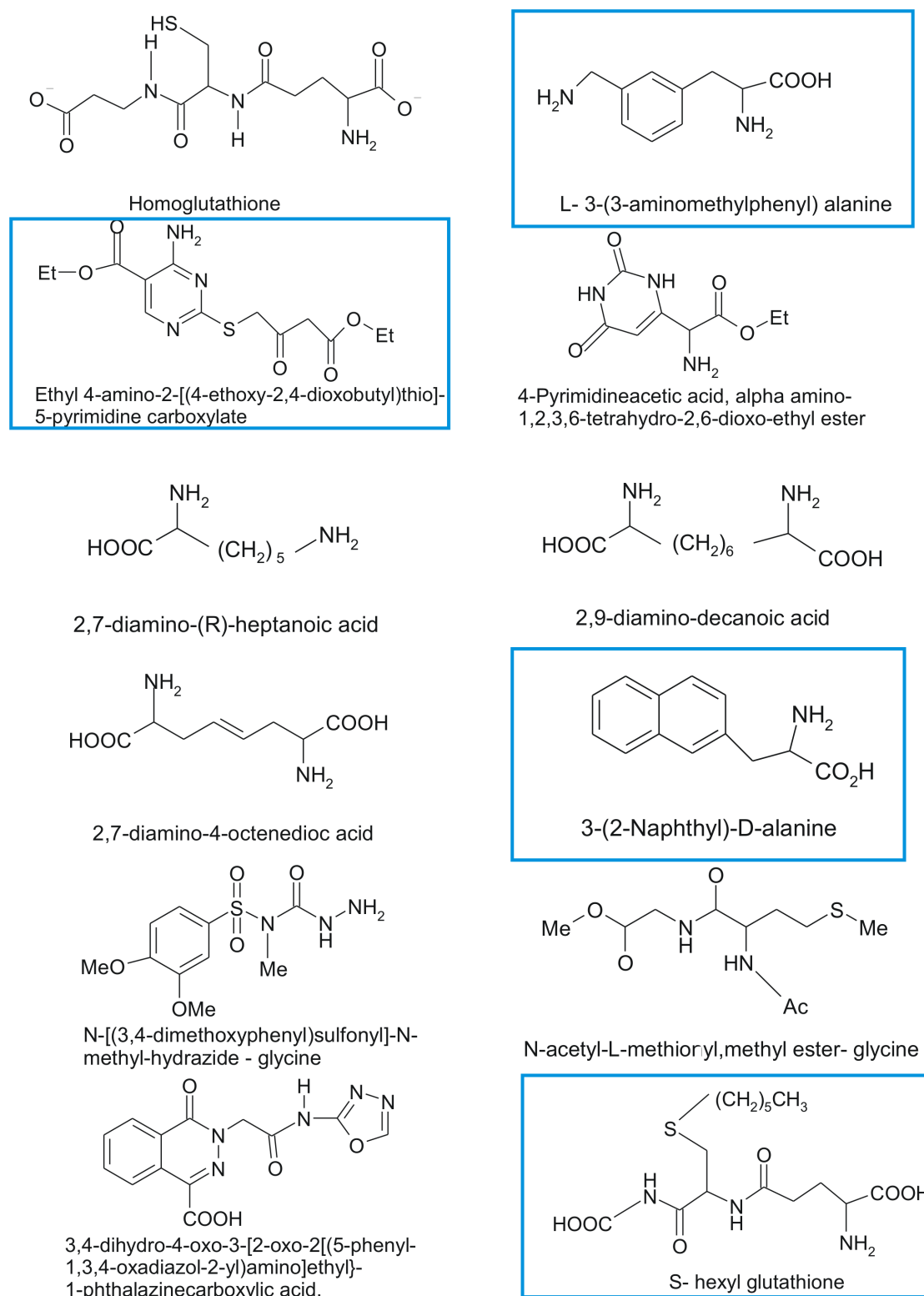


Figure 2.8: **The twelve ligands from the design process that could be obtained commercially.** The four ligands from the design process that were purchased are framed in blue.

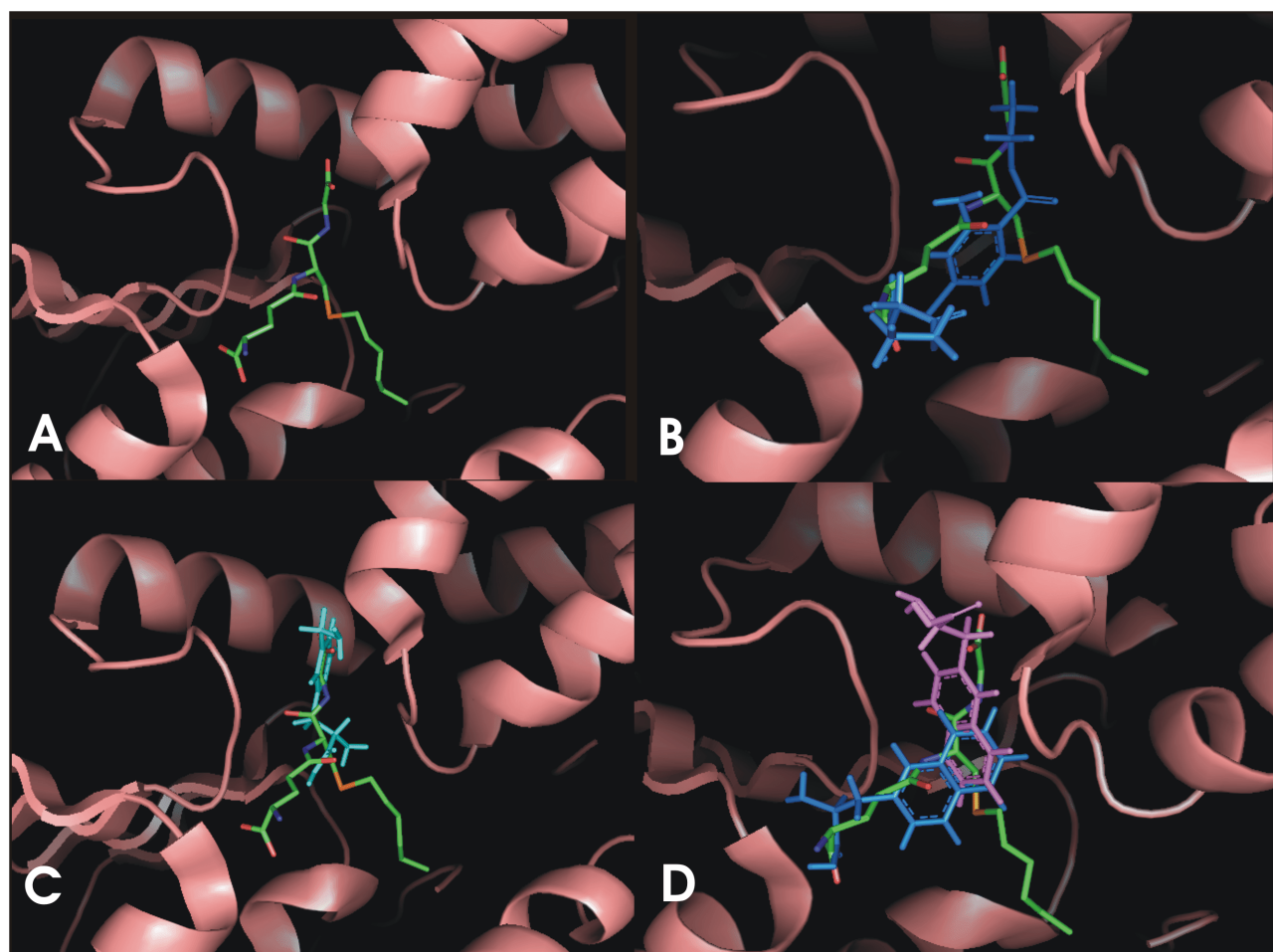


Figure 2.9: **AutoDock docking of GTX, EDP, LAP and NDA.** A Pymol presentation of the orientation in which GTX binds to PfGST as seen in the crystal structure (A). Proposed orientations in which EDP (B), LAP (C) and NDA (D) would bind to PfGST as predicted by AutoDock.

Hence, there was a positive correlation between the predictions of LUDI, AutoDock and XScore. It was expected that the biological assay would show NDA to be the best inhibitor followed by LAP and EDP. The visual inspection of how AutoDock proposed the four commercial ligands would bind to PfGST is shown in Figure 2.9.

For EDP, LAP and NDA the binding orientations are shown with regard to GTX in its crystallized orientation. During the visual inspection of the binding mode for NDA (Figure 2.9 D) it was evident that there were two conformations that had equally good scores in AutoDock and seemed similarly likely to bind to PfGST. Therefore both the conformations

were included in Figure 2.9. Based on the overall consensus score and **AutoDock** ranking the blue configuration was chosen as the best one and was used in the rest of the *in silico* screens. A comparison of binding interactions between the inhibitors (GTX, LAP, EDP and NDA) and PfgGST was visualized using **LIGPLOT** and can be seen in Table 2.4.

LIGPLOT was used to gain some information on the amino acids that these inhibitors would use for molecular interactions. A complete map of these interactions was deduced by using different conformations of the inhibitors, at various stages of minimization, one of these **LIGPLOT** maps is shown in Figure 2.10. Since the active site amino acids are highly conserved for the binding of GSH, it was interesting to monitor which of these conserved interactions were maintained by the newly-designed inhibitors. **LIGPLOT** uses a program **HBPLUS** to define hydrogen bonds and therefore the focus was on the comparison between the conserved hydrogen bonds. For NDA it was found that the ligand maintained at least five of the hydrogen bonds, being the highest. This binding mode can be positively correlated with the consensus score. LAP maintained four hydrogen bonds and EDP only three. This is again in good agreement with the affinity scores predicted by **AutoDock**, **Xscore** and **LUDI**. The hydrophobic interactions were more difficult to quantify.

In literature there are two major concerns regarding the use of scoring functions: firstly, it is known that scoring functions often have difficulty in determining relative binding affinity values, which can be directly correlated to an inhibition constant or experimentally determined binding free energy. Secondly, scoring functions seldom have the ability to distinguish between structurally-similar ligands. These two concerns were addressed in this study as well. A consensus score (summation of three different scores) was used to prioritize the ligands for future studies. Supported by the range of scores produced by **AutoDock**, **FlexX** and **XScore** it seems that these three scorings functions were able to distinguish between members of a highly similar library.

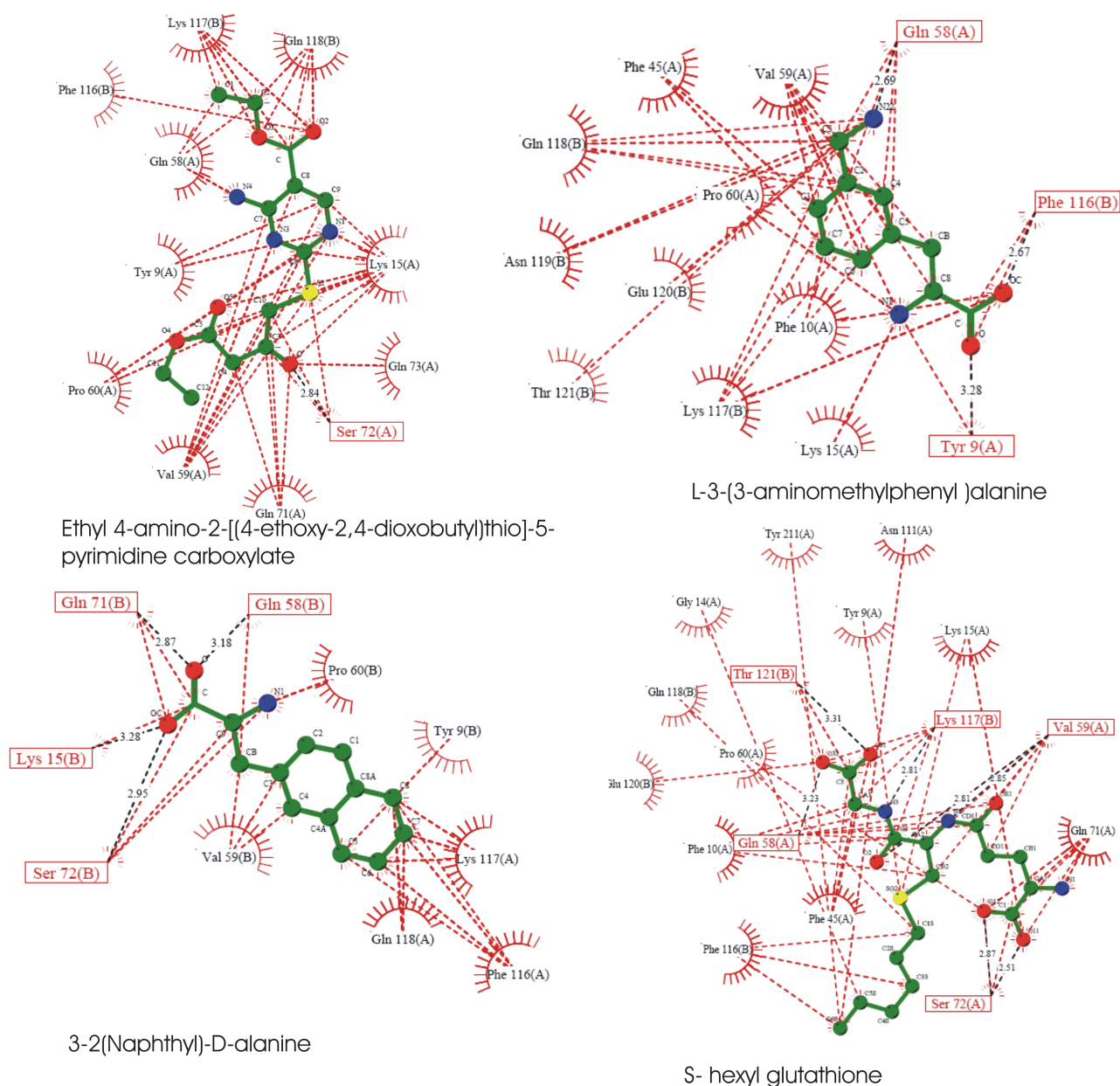


Figure 2.10: **Representation of the bonds that the inhibitors GTX, EDP, LAP and NDA form with PfgST.** Hydrogen bonds are indicated in black and the hydrophobic interactions in red. Carbon atoms were coloured green, oxygen red, nitrogen blue and the sulphur atom of GTX was coloured yellow. A and B denote the A and B chain of the heterodimer, A and A' denote the A chains of the homodimer.

Table 2.4: Interactions that formed between PfGST and the commercially available inhibitors (GTX, LAP, EDP and NDA). Interactions were derived from the **AutoDock** docking data as well as **LIGPLOT** visualization. Hydrophobic interactions are represented by the plain text and hydrogen bonds are represented by bold text. A and B denote the A and B chain of the heterodimer, A and A' denote the A chains of the homodimer.

Residues	GTX	LAP	EDP	NDA
Tyr 9 (A)	Tyr 9 (A)	Tyr 9 (A)		
Phe 10 (A)	Phe 10 (A)		Phe 10 (A)	Phe 10 (A)
Gly 14 (A)	Gly 14 (A)			
Lys 15 (A)	Lys 15 (A)	Lys 15 (A)	Lys 15 (A)	Lys 15 (A)
Phe 45 (A)	Phe 45 (A)			
Gln 58 (A)	Gln 58 (A)	Gln 58 (A)	Gln 58 (A)	Gln 58 (A)
Val 59 (A)	Val 59 (A)	Val 59 (A)	Val 59 (A)	Val 59 (A)
Gln 71 (A)	Gln 71 (A)	Gln 71 (A)	Gln 71 (A)	Gln 71 (A)
Ser 72 (A)	Ser 72 (A)	Ser 72 (A)	Ser 72 (A)	Ser 72 (A)
Asn 111 (A)	Asn 111 (A)			
Gly 120 (A)	Gly 120 (A)			
Asp 105 (A')	Asp 105 (A')			
Pro 60 (B)	Pro 60 (B)	Pro 60 (B)	Pro 60 (B)	Pro 60 (B)
Phe 116 (B)	Phe 116 (B)	Phe 116 (B)	Phe 116 (B)	Phe 116 (B)
Lys 117 (B)	Lys 117 (B)	Lys 117 (B)	Lys 117 (B)	Lys 117 (B)
Gln 118 (B)	Gln 118 (B)	Gln 118 (B)	Gln 118 (B)	Gln 118 (B)
Thr 121 (B)	Thr 121 (B)			

2.5. Discussion

2.5.1. Program validation

Docking accuracy is very dependent on the type of receptor used (Wang *et al.*, 2003). The ultimate goal of this study was to validate and select the best design, docking and scoring functions for the design of pfGST inhibitors. The first step in this study was therefore to assess the docking accuracy of docking programs available and optimize the parameters specifically for PfGST. The docking programs **AutoDock**, **FlexX** and **DOCK** were evaluated. Ideally the docking program should accurately locate the catalytic site of the receptor and use that as the sample space; **DOCK** failed to do so and was eliminated as an option for docking. **DOCK** did not manage to place the GTX molecule in an extended conformation as seen in the crystal structure. Although the active site was defined prior to docking, **DOCK** placed GTX in the cleft between the monomers and not in the G or H site. Since libraries of more than a hundred compounds were docked a high throughput system was required. Although **FlexX** and **AutoDock** both produced reasonably good results, **AutoDock**

was preferred to **FlexX** because it was less computationally expensive and no user input was required once the method was optimized. The accuracy of **AutoDock** was validated by its ability to correctly reposition GTX in the active site of PfGST. The method of binding free energy predictions proved to be satisfactory. When the docked poses of GTX produced by **AutoDock** were visualized in **InsightII**, the computer-predicted complexes were in good agreement with the crystallographic structures. These results suggest that **AutoDock** as docking method could be exploited to predict the binding orientation of ligands to the G and H sites of PfGST. Two major limitations of structure-based docking experienced in this study were the inability of **AutoDock** to account for induced-fit interactions, another was the low sensitivity towards specific water interactions (Brooijmans and Kuntz, 2003). In **AutoDock**, ligand flexibility was addressed by allowing the ligand to be flexible in a rigid binding site. Hopefully, faster computers and more efficient algorithms will enable the future docking of flexible ligands into proteins with flexible side chains, to reduce the number of ligands that were overlooked due to rigid docking. These limitations were accepted to be incorporated into the false-positives and false-negatives produced. Scoring functions can not be compared in this way, as scoring functions are highly dependent on the accuracy of the docking programs used to position the ligand.

2.5.2. Ligand design

Binding hot spots were incorporated in the design process. A hot spot can be defined as a small subset of amino acids that contribute to the affinity of the enzyme for the substrate (Gohlke *et al.*, 2000b). These hot spots were translated into geometrical features known as pharmacophore fragments. This was the approach followed by **NEWLEAD**. NDA and LAP were examples of inhibitors designed by **NEWLEAD**. Both these inhibitors scored very well, NDA was predicted to be the better one of the three inhibitors with LAP, second. Depending on the biological assay, **NEWLEAD** might prove to be one of the best ways to design future inhibitors. To increase the parasite specificity the initial pharmacophore fragments from different inhibitors could be used.

2.5.3. EDP, LAP and NDA as inhibitors of PfGST

To date most GST inhibitors can be divided into five major classes (i) analogues of the electrophilic substrates (binding to the H site) (ii) Glutathione conjugates (double headed compounds) that will bind competitively to the G site but also protrude into the H site

(iii) Ligands that bind non-competitively to non-substrate sites of GST (iv) GSH peptide analogues, and (v) Bivalent inhibitors. GSH analogues have been widely tried as inhibitors (Mahajan and Atkins, 2005). One of the major pitfalls of this was the hydrolysis of the unusual peptide linkages by γ -glutamyl transpeptidase, meaning that the ligand will be destroyed before reaching the enzyme (Mahajan and Atkins, 2005). To overcome this limitation, the whole tripeptide was used for modification by LUDI and no substitutions of whole amino acids took place.

LUDI added or replaced functional groups on GSH and GTX to produce larger structures that would saturate other interaction sites present in the G site as well as the H site. Based on the visualization of the compounds produced by LUDI it was clear that only four of these designed inhibitors explored the H site, by maintaining the hexyl chain from GTX. None of the changes made by LUDI were done in the H site. Consequently, this design was not successful in exploring the H site of the enzyme. This may lead to a decrease in binding affinity due to the loss of additional hydrogen bonds that could form to amino acids in the H site. The alterations were mainly in the G site, meaning that the specificity towards PfGST might not have increased very much since the G site is highly conserved and the deviation from the other GST enzymes lies within the more solvent accessible H site. These newly-designed compounds were scored for their affinity towards PfGST by an empirical scoring function included in the LUDI package. The scoring function of LUDI was very fast, even on large ligand-protein complexes. Unfortunately due to the geometric configuration of this fitness function it tends to miss interactions due to delocalized electrostatic and van der Waals interactions (Joseph-McCarthy, 1999). EDP was the only example of a LUDI-designed inhibitor that was biologically tested.

2.5.4. Clustering and biological testing criteria

An *in silico* design strategy was used to minimize the cost, consequently not all 96 designed compounds could be tested. Therefore the library were clustered to obtain a group of scaffold structures that would represent the complete library.

Ward's hierarchical clustering produced 8 clusters with a average of 10 members per cluster. Half of the clusters had aromatic scaffolds and the other half had substituted hydrocarbon chains as basic scaffold. The 4 compounds selected were chosen to represent 4 different clusters. Bemis and Murcko, (1996) analyzed the most common features found in

drug molecules, and concluded that acyclic molecules account for only 6% of total drug molecules. The study proved that 8.5% of the drug molecules studied had benzene as molecular framework.

The 10 best scoring molecules belonged only to 4 clusters. Three of these molecules belonged to cluster 1, like EDP. Four of these molecules belonged to cluster 4, like GTX. The other 3 belonged to cluster 4 (like LAP), cluster 6 and cluster 7 (like NDA). Commercially available molecules were chosen to represent the clusters with the top scoring molecules. Cluster 6 were not represented because there were no commercially available molecules belonging to that class. Based on the study by Bemis and Murcko, (1996), preference was given to cyclic molecules.

2.5.5. *In silico* binding affinity determination of EDP, LAP and NDA

Key amino acids in the G site of PfGST have been identified by observing the binding mode of glutathione and GTX. The γ -glutamyl moiety of GSH has a highly polar structure therefore it is most critical for binding. The interaction that this moiety forms with PfGST will therefore be important to the binding affinity (Burg and Mulder, 2002). The γ -glutamyl part of GSH forms hydrogen bonds with Lys 15, Val 59, Gln 71 and Ser 72 (Burg and Mulder, 2002). Perbandt *et al.* (2004) showed other important interactions formed with Tyr 9 and Gln 58. Hiller *et al.* (2006) showed that Tyr 9 was responsible for the deprotonation of GSH together with Lys 15 and Gln 71 being important for GSH binding. Subsequently, the hypothesis was that those inhibitors that exploited these bonds will have a higher affinity for PfGST.

XScore produced binding affinity predictions for the three inhibitors, NDA was predicted to have the highest affinity for PfGST followed by LAP and EDP. Hence, there is a correlation between the predictions of AutoDock and XScore. The XScore values, as well as LUDI scores, are highly dependent on the hydrogen bonds that form between the ligands and the protein. A comparison of the intermolecular interactions were done. The interactions formed by the inhibitors and PfGST can be seen in Table 2.4. These predicted interactions support the scores given by LUDI and XScore. For example, NDA had the highest consensus score and this was supported by the 5 hydrogen bonds to key amino acids.

LIGPLOT was used to visualize the interaction between GTX and PfGST. The LIGPLOT

predicted interactions between GTX and PfGST were also highly correlated with the crystallographical interactions, namely the nine molecular interactions that were observed and depicted in Figure 1.8 (Perbandt *et al.*, 2003; Fritz-Wolf *et al.*, 2003). In this study two additional hydrogen bonds were predicted: the carboxyl group of the glycine moiety interacts with Thr 121(B) and the α -amino group of the glycine moiety forms a hydrogen bond with Lys 117(B). The γ -glutamyl part of GTX formed hydrogen bonds with Gln 58, Gln 71 and Val 59 and there was a close hydrophobic interaction with Lys 15 and Gln 71, as seen in (Figure 2.10).

EDP formed hydrogen bonds with Phe 10, Gln 58 and Ser 72 but hydrophobic interactions with Lys 15, Val 59 and Gln 71 were predicted as well. EDP was the biggest of the three inhibitor molecules but formed the least amount of interactions with the protein receptor. EDP formed only three hydrogen bonds with the protein. Structurally there was the occurrence of a six carbon ring in the form of a pyrimidine. The amino group on the pyrimidine formed a hydrogen bond with Gln 58. Ser 72 and Phe 10 formed bonds with the two oxygen groups of the 2,4-dioxobutyl moiety, respectively. Except for GTX, EDP was the only molecule with a sulphur atom. The sulphur atom on GSH plays an important role in the conjugation of GSH to electrophilic compounds, this reaction is catalysed partially by the Tyr 9 amino acid (Armstrong, 1997). The sulphur atom on EDP was predicted to form an interaction with Lys 15, as is seen in Figure 2.10.

When comparing the orientation in which LAP was docked with the orientation of GSH, LAP was positioned in such a way that it lies in the G site of PfGST. Consequently the interactions involved are very similar. The carboxyl group of the γ -glutamyl moiety of GTX is aligned and formed two hydrogen bonds with the main chain as well as the side chain of Ser 72. The carboxyl group of LAP also formed a hydrogen bond with Ser 72 as well as Tyr 9 and Phe 119 (from the B subunit). The α -amino group of the phenylalanine residue (LAP) formed a hydrogen bond with Lys 15. This is an important bond since Lys 15 is not conserved in the human GST enzymes. The aminomethyl group formed a hydrogen bond with Gln 58, which was known to bind to the glycine moiety of GTX. LAP formed two important hydrogen bonds with Gln 58 and with Tyr 9. LAP was the only ligand that formed a hydrogen bond with Tyr 9. Tyr 9 plays a huge role in the catalytic mechanism of GST enzymes (Armstrong, 1997).

Hydrogen bonds with the γ -glutamyl moiety of GTX were formed with Val 59, Gln 58,

Gln 71 and the Lys 15, similar bonds also forms with the alanine moiety of NDA. NDA formed hydrogen bonds with Lys 15, Gln 58, Gln 71 and Ser 72 as well as close hydrophobic interactions with Val 59 and Pro 60. The hydrophobic interactions with Tyr 9 should also be noted since Tyr 9 plays an important catalytic role.

Based on this visualization in Figure 2.10 the GTX molecule should have a higher affinity for PfGST than NDA, LAP and EDP. As GTX has the most predicted hydrogen bonds, followed by NDA, LAP and EDP. This trend was represented by the empirical scoring functions (Table 2.3), since LUDI and XScore uses hydrogen bond data to calculate binding affinity (Bohm, 1992; Wang *et al.*, 1998).

Table 2.5: Comparison of GTX binding between PfGST and the human μ -class glutathione S-transferase (HmGST) enzyme. Hydrophobic interactions are represented by the plain text and hydrogen bonds are represented by bold text.

PfGST	PfGST and HmGST	HmGST
		Trp 7(A)
	Tyr 9 (A)	
Phe 10 (A)		
Gly 14 (A)		
Lys 15 (A)		Arg 15 (A)
Phe 45 (A)		Trp 45 (A)
		Lys 49 (A)
	Gln 58 (A)	
	Val 59 (A)	
		Thr 68 (A)
Gln 71 (A)	Gln 71 (A)	Gln 71 (A)
	Ser 72 (A)	
Asn 111 (A)		
Gly 120 (A)		
	Asp 105 (B)	
	Pro 60 (B)	
Phe 116 (B)		
Lys 117 (B)		
Gln 118 (B)		
Thr 121 (B)		
		Arg 131 (B)

In the Swiss-Prot protein (Watanabe and Harayama, 2001) knowledgebase there are 19 different human GST isozymes belonging to 8 different classes. In the μ -class alone there is

5 different isozymes. From the crystal structures available (PDB codes: 6GSS, 9GSS, 1YJ6 and 1XW6) the key amino acids that human μ -class GSTs use to bind GTX were identified, (Table 2.5). A comparison between the interactions of GTX with PfGST and HmGST is shown in Table 2.5. There are a few noticeable differences, the Lys 15 in PfGST was replaced by a Arg 15 in HmGST, also a positive amino acid. Conserved hydrogen bonds were identified with Tyr 9, Gln 58, Val 59, Gln 71, Ser 72 and Asp 105 (B), and hydrophobic interactions to Pro 60. In Table 2.4 PfGST formed a set of conserved hydrophobic interactions with Phe 10, Phe 116, Lys 117, Gln 118, Thr 121 and the inhibitors. None of these interactions were present in HmGST binding, Table 2.5. Hypothetically, these molecular interactions can be exploited to make parasite specific inhibitors. All the crystal structures of all the human GST class enzymes were not available, a more comprehensive study would have to be launched to determine more parasite-specific interactions.

LogP was used to estimate the water solubility of the three inhibitors. The LogP value of LAP (-0.09) suggested LAP to be soluble but solubility problems were predicted for EDP (1.17) and NDA (2.05). This suggests that LAP should be water soluble but also infers that organic solvents like DMSO or dimethylformamide (DMFO) should be used to dissolve EDP and NDA. GTX has a LogP value of -0.08 and is known to be soluble in water at low concentrations but in DMSO in high concentration. NDA is structurally similar to LAP with the presence of a ring system combined with the alanine moiety. The benzene ring of LAP was replaced with a naphthalene ring on NDA. It was known that benzene derivatives tend to be insoluble in water, somewhat soluble in ethanol, soluble in benzene, and very soluble in ether, chloroform, or carbon disulfide. Consequently, it was suspected that LAP would be more soluble than NDA, as suggested by the LogP values (Table 2.3).

Chapter 3 describes the *in vitro* binding studies of the four commercially available compounds: L-3-aminomethylphenylalanine, ethyl 4-amino-2- [(4-ethoxy-2, 4-dioxobutyl) thio]--5-pyrimidine carboxylate, 3-(2-Naphthyl)-D alanine and S-hexyl glutathione. During these biochemical assays the set of representative molecules were validated for their ability to bind and inhibit PfGST. Leadlike clusters were identified and are to be pursued for future drug development.

Chapter 3

In vitro evaluation of inhibitors

3.1. Introduction

In order to gain trust in computational inhibitor design as well as the method of affinity prediction, it is necessary to test some of the designed inhibitors experimentally. Based on these experimental results, one can demonstrate the method's ability to predict the affinity between an inhibitor and the protein *in silico*. Furthermore, these results can be used as guidelines to further optimize and modify current inhibitors. If this strategy of *in silico* ligand design and affinity predictions prove to be successful, this strategy will reduce the cost and time of future drug development. *Chapter 3* will therefore address the testing of the four commercially available ligands (GTX, LAP, EDP and NDA) to add confidence to the *in silico* methodology of designing ligands, determination of binding mode by docking and scoring the binding affinity as applied to PfGST.

3.1.1. Correlation between *in silico* and biological data

In the literature there are various examples of authors attempting to correlate *in silico* derived molecular docking and affinity scoring data with biologically obtained results, and some of these methods and results will be briefly described as examples in the following section.

In a study conducted by Gradler *et al.* (2001), the X-ray structure of *Zymomonas mobilis* tRNA gaunine glucosylase was used with LUDI to search for new putative inhibitors by screening the Available Chemical Directory (ACD). The hits identified by LUDI were further screened according to affinity computed by Bohm's scoring function as incorporated into the LUDI score. One of the compounds (4-aminophthalhydrazide) had a LUDI score of 542, hence a predicted inhibition constant of 8.3 μM . This compound could not be synthesized,

hence a derivative, 3,5-diaminophthalhydrazide, was synthesized and tested in a biological assay. 3,5-diaminophthalhydrazide and 4-aminophthalhydrazide had the same molecular fingerprints. The inhibition constant of this compound was found to be $0.2\mu M$. There was thus a positive correlation to the ability of LUDI to find a compound that could bind to the enzyme but the LUDI score was also significant enough to rank the derivatives according to putative binding affinity. LUDI was validated by Gradler *et al.* (2001) to be able to predict binding affinity for compounds when applied to tRNA gaunine glucosylase. This study will determine if LUDI can be used to accurately predict the binding affinity of inhibitors as applied to PFGST.

In contrast, Enyedy *et al.* (2001) screened the NCI 3D database to find inhibitors for the B-cell CLL/lymphoma 2 (Bcl-2) protein that plays an important role in programmed cell death in humans. DOCK screened 206,875 compounds. From the 80 candidate compounds identified by DOCK only 35 were available for testing. During biological assays only 7 of these molecules were active as inhibitors. The authors concluded that the compound rankings performed by DOCK were not satisfactory and different means of ranking should be used in future studies. These results confirm the findings that DOCK cannot be used to screen ligands for affinity to certain enzymes like PFGST and Bcl-2.

Li *et al.* (2004) conducted a similar study to Enyedy *et al.* (2001) by screening the NCI 3D database to find inhibitors to human 5-aminoimidazole-4-carboxamide ribonucleotide transformylase. AutoDock was used for the docking simulations. AutoDock identified 16 inhibitors that were predicted to bind to the enzyme with binding free energy within the range of -10 kcal/mol to -14.7 kcal/mol. Of these 16 inhibitors, 8 were active inhibitors that inhibited in the micromolar range. One of the inhibitors that was tested in a biological assay had an inhibition constant of 154 nM and an IC_{50} value of 600 nM. The authors concluded that AutoDock was able to correctly rank the inhibitors with a standard deviation in binding free energy of 2.1 kcal/mol, except when the molecules were structurally very similar.

The English saying that “a chain is only as strong as its weakest link” can also be applied to the docking and scoring of ligands. The scoring of the ligands by a state-of-the-art scoring function is only as powerful as the docking function used to position the ligand in the active site of the enzyme. Hence, the scoring function’s ability to discriminate between a good and bad inhibitor is highly dependent on the quality of docking that was performed on the

inhibitor prior to scoring.

Zang *et al.* (2004) compared the ability of various scoring functions to predict binding affinity between ligands and their respective protein receptors in comparison to experimentally--determined binding affinities. The Spearman correlation coefficient was used for convenient comparison between the different scoring functions. XScore was ranked second to the DFIRE scoring function with correlation scores of 0.64 and 0.66, respectively. Unfortunately, DFIRE was not freely available therefore XScore was used.

3.1.2. Previous cloning and recombinant expression of PfGST

The recombinant PfGST was successfully cloned and expressed at high levels by the groups of Perbandt *et al.* (2002) and Harwaldt *et al.* (2002). The cloning vectors and method of purification used by the two groups were different and are compared in Table 3.1. The method that was used by Prof. Liebau (Liebau *et al.* 2002; Perbandt *et al.* 2002) is described in more detail since the PfGST-pJC20 plasmid used in this study was provided and prepared accordingly, by Prof. Liebau.

Table 3.1: Comparison of cloning and expression between the methods used by Perbandt *et al.* (2002) and Harwaldt *et al.* (2002).

Factor	Perbandt <i>et al.</i> (2002)	Harwaldt <i>et al.</i> (2002)
Expression vector	pJC20	pQE20
Expression host	<i>E. coli</i> BL21(DE3)	<i>E. coli</i> M15
Growth Media	Terrific Broth	LB media
Purification Column	GSTrap - Glutathione Sepharose	Ni-NTA column (His-Tag protein)

The PfGST gene occurs on Chromosome 14 of the *P. falciparum* genome. It contains two exons and consists of 636 bp including the start and stop codons with a GenBank accession number of AY014840. Liebau *et al.* (2002) determined that the cDNA of PfGST encodes a functional protein. The cDNA of the whole coding region was amplified using the enhanced Avian RT-PCR Kit (Sigma) (Liebau *et al.*, 2002). The PCR product was digested with HindIII and BamHI, and was cloned into the high expression vector pJC20. The pJC20-PfGST plasmid was transformed into *E. coli* BL21(DE3) cells. Expression was carried out in a 2L high-density fermentor with Terrific Broth containing 50 μ g/ml ampicillin (Liebau *et al.*, 2002). After affinity purification with glutathione-Sepharose the identity of

the protein was checked by SDS-PAGE as well as immunoblotting (Perbandt *et al.*, 2002; Harwaldt *et al.*, 2002). Burmeister *et al.* (2003) obtained protein expression of 7 mg/ml for crystallization purposes.

3.1.3. Testing the activity of PfGST using a colourimetric assay

The PfGST enzyme was found to catalyze the conjugation of glutathione (GSH) to 1-chloro-2,4-dinitrobenzene (CDNB). The method for detecting enzyme activity of PfGST is a colourimetric assay, based on the catalyzed formation of CDNB-GSH, shown in Figure 3.1. The dinitrophenyl thioester that is formed can be detected with a spectrophotometer at 340nm. One unit of GST activity is defined as the amount of enzyme producing 1 μ M of GSH-CDNB conjugate per minute at 25°C and pH 6.5. The GSH-CDNB conjugation can be used to test any GST activity because the reaction is independent of the isozyme type present. Under these laboratory conditions the spontaneous conjugation of GSH and CDNB cannot take place.

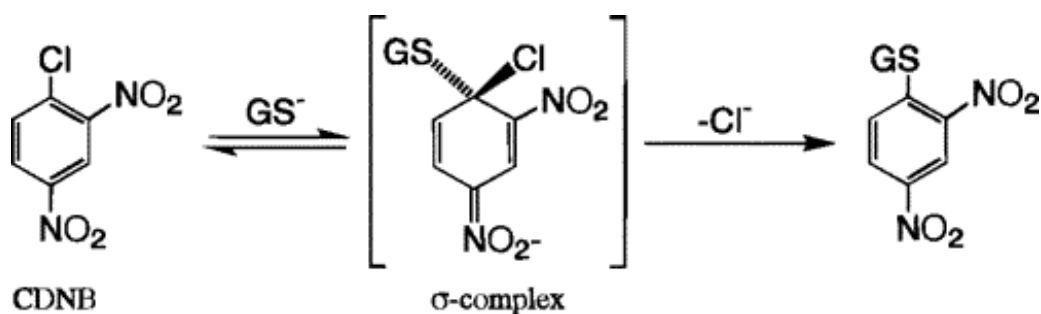


Figure 3.1: **Addition of glutathione (GS⁻) to 1-chloro-2,4-dinitrobenzene, catalyzed by glutathione S-transferases (Armstrong, 1997).** The addition of glutathione to 1-chloro-2,4-dinitrobenzene results in the formation of dinitrophenyl thioester that can be detected with a spectrophotometer at 340nm, to follow GST activity.

To determine the change in absorbance, the slope of the graph of absorbance versus time was determined. The rate of the reaction ($\Delta A_{340}/\text{min}$) was substituted into the Beer-Lamberts Law (Lambert 1760; Beer, 1854) to determine specific PfGST activity.

$$A = \varepsilon CD$$

A was the rate of the reaction ($\Delta A_{340}/\text{min}$), ε the molar extinction coefficient of PfGST ($\varepsilon_{340} = 9600 \text{ M}^{-1} \cdot \text{cm}^{-1}$), C was the specific activity of PfGST in $\mu\text{mol}/\text{min}/\text{mg}$ enzyme and D was the path length determined by the cuvette that was used. Harwaldt *et al.* (2002)

determined specific activity to be 0.2 U/mg at pH 6.5. The pH profile indicated that at pH 8.1 the optimum activity was reached (0.64 U/mg). However, at this pH a spontaneous reaction between GSH and CDNB occurred (Harwaldt *et al.*, 2002) and the inhibition assays were done at pH 6.5.

3.1.4. Additional kinetic data available for PfGST

Harwaldt *et al.* (2002) tested the following inhibitors: Cibacron Blue, ethacrynic acid, hemin, GTX and protoporphyrin IX and these results are shown in Table 3.2.

Table 3.2: Inhibition of PfGST by selected compounds (Harwaldt *et al.*, 2002).

Inhibitors	IC ₅₀	Type of inhibition	K _i
Cibacron Blue 3GA	4 μM	Competitive	0.5 μM
Ethacrynic Acid	30 μM	Not determined	Not determined
Hemin	2.5 μM	Uncompetitive	6.5 μM
GTX	20 μM	Competitive	35 μM
Protoporphyrin IX	>40 μM	Mixed type	10 μM

The following secondary substrates for PfGST were identified: CDNB, ethacrynic acid and o-nitrophenol in GSH-dependent reactions. The K_m value for CDNB could not be determined due to the strong absorption of CDNB at 340nm (Harwaldt *et al.*, 2002). The K_m for GSH was determined by Harwaldt *et al.* (2000) to be $164 \pm 20 \mu M$ and by Liebau *et al.* (2002) to be $156 \pm 13 \mu M$.

3.1.5. Criteria for validation of methodology

The criteria used in this study to classify a ligand as a putative inhibitor were: firstly, the inhibitor must have an acceptable docking score, which means that the inhibitor will fit into the active site; the cut-off value for this was based on energy values (given by AutoDock in Table 2.1) as well as visual inspection of the binding site (LIGPLOT in Figure 2.10). Secondly, an energy-based affinity score from LUDI and XScore was taken into account. By combining the scores from LUDI, AutoDock and XScore a consensus score of the binding affinity was established. Since most of these scoring functions work on an arbitrary energy value the actual values were not explicitly used. The consensus score was used to rank these inhibitors according to their ability to bind PfGST. Lastly, the availability and cost implications were taken into consideration as well, since none of these inhibitors form part of a mass production

scheme but were synthesized on request in milligram quantities.

The LUDI scores for all four inhibitors were above 500. Based on the above mentioned literature, Gradler *et al.* (2001), the K_i values for these inhibitors should be below 500 μM . If the AutoDock scores for the compounds were between -10 kcal/mol and -14.9 kcal/mol, the molecules could inhibit in the nanomolar range (Li *et al.*, 2004). GTX was the only inhibitor that scored better than -14.9 kcal/mol and it is known that GTX acts as a competitive inhibitor to PfGST with a K_i value of 35 μM (Harwaldt *et al.*, 2002). Since EDP, LAP and NDA had scores of between -8.54 kcal/mol and -9.58 kcal/mol, these inhibitors were expected to inhibit in the micromolar ranges. From the study by Zang *et al.* (2004), XScore was expected to rank the molecules similarly to the biological data, according to their affinity for PfGST. If the biological data support these *in silico* results it would confirm that the consensus score can be used to accurately predict the binding affinity ligands have for PfGST. Overall enzyme inhibition will prove that LUDI and NEWLEAD can be used in conjunction to produce ligands that can act as PfGST inhibitors. GTX was used as a benchmark in this study because it is a known inhibitor of glutathione S-transferase and was docked and scored together with all the other ligands.

3.2. Goals

Chapter 2 focused on the design of ligands to fit into the active site of PfGST with a high enough binding affinity for the receptor, enabling the inhibitor to compete with the substrate in the biological system.

Based on the *in silico* analysis, S-hexyl glutathione (GTX) was predicted to offer the best inhibition followed by 3-(2-Naphthyl)- D alanine, L-3-aminomethylphenylalanine and ethyl 4-amino-2- [(4-ethoxy-2, 4-dioxobutyl) thio]-5-pyrimidine carboxylate (Figure 3.2).

The main aim of *Chapter 3* was to validate the strategy of ligand design and affinity determination methods used, by testing commercially available compounds. The following objectives were set:

- Transformation of the plasmid pJC20-PfGST into BL21(DE3) *E. coli* cells followed by recombinant protein expression.

- Affinity purification of PfGST and use in a standard assay with GSH as primary substrate and CDNB as secondary substrate.
- Testing of the inhibitors (GTX, EDP, LAP and NDA) in the standard CDNB assay for activity against PfGST.
- The determination of enzyme kinetics such as K_m and K_i using the Lineweaver-Burk double reciprocal plots.

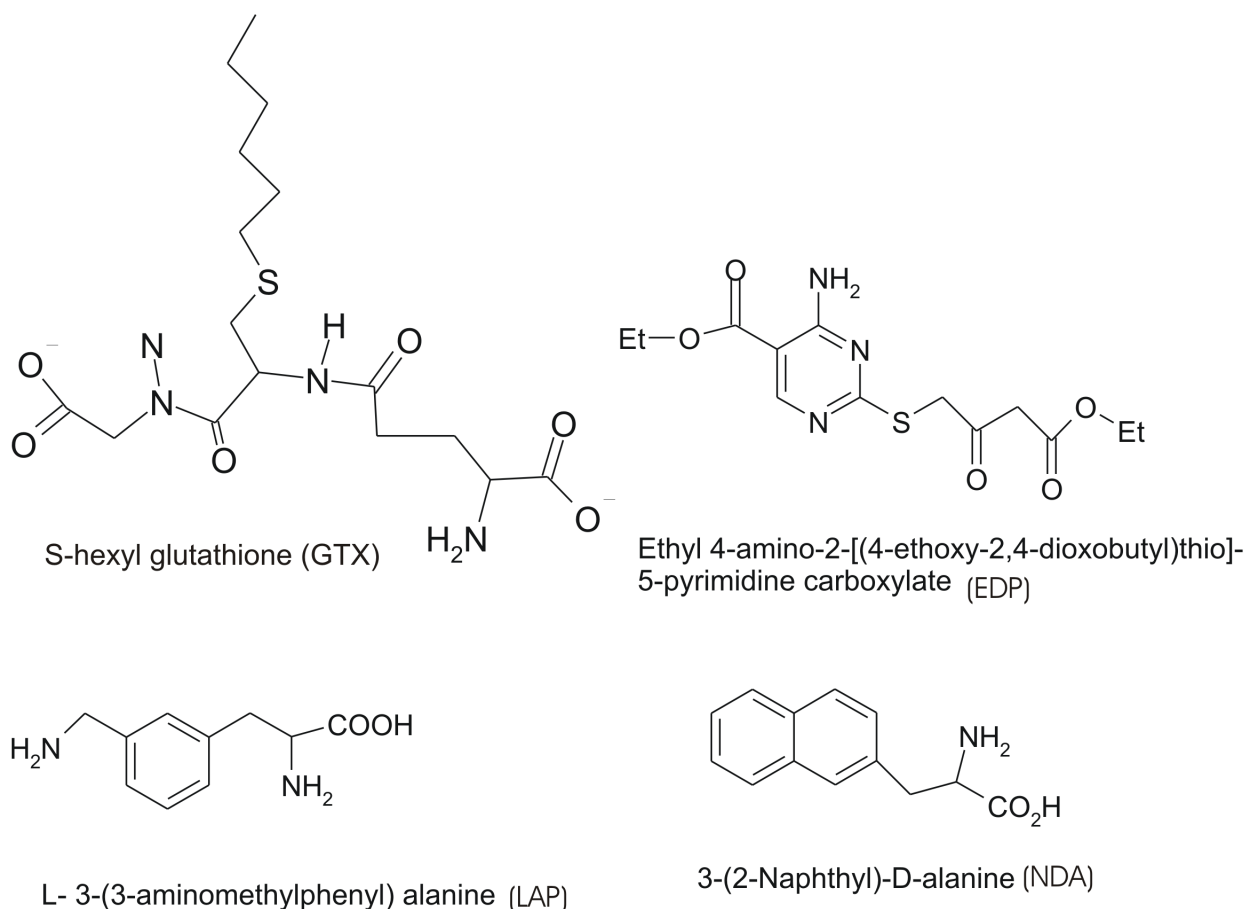


Figure 3.2: **The four commercially available inhibitors GTX, EDP, LAP and NDA.** GTX was used as a positive control. NDA and LAP were examples of inhibitors designed by NEWLEAD and EDP was designed by LUDI.

3.3. Materials and Methods

3.3.1. Restriction enzyme digestion

The PfgST gene cloned into pJC20 (Clos and Brandau, 1994), a high expression vector, was kindly provided by Prof. E. Liebau (Liebau *et al.*, 2002; Perbandt *et al.*, 2004). In order to confirm the integrity of the plasmid, the following restriction enzyme digestion was set up: 5 μ l of the plasmid DNA (397 ng/ml) was digested with 1 μ l BamHI (10U, Fermentas) and 1 μ l HindIII (10U, Fermentas), in the appropriate 1 \times buffer at 37°C for 2 hours.

3.3.2. Agarose gel electrophoresis

The digested products were separated by electrophoresis on a 0.8% agarose (Promega, Wisconsin, USA) gel. TAE buffer (0.04M Tris-Acetate, 1mM, pH 8.0) was used for electrophoresis. Ethidium bromide, a DNA base intercalator (1.25 ng/ml) was included in the gel solution for DNA visualization. Samples were electrophoresed at 8 V.cm⁻¹ and the bands were visualized at 312 nm, on a Bio-Rad Chemidoc Gel Documentation System (Bio-Rad laboratories, USA), using the QuantityOne software.

3.3.3. Preparation of electrocompetent cells

Electrocompetent cells were prepared by the following method (Sambrook *et al.*, 1989): A single colony of BL21(DE3) from *E. coli* (Novagen, Madison, Wisconsin, USA) was inoculated into 15 ml Luria-Bertani (LB) liquid media (1% Tryptone, 0.5% yeast extract, 1% NaCl at pH 7) and grown overnight at 37°C with shaking (300 rpm). Fresh LB-Broth (250 ml) was inoculated with 1 ml of the overnight culture and grown at 37°C with shaking (220 rpm) until the cells reached early to mid-log phase (OD at 600nm of 0.3-0.6). To harvest, the cells were transferred to two cold centrifuge tubes and pelleted at 5,000 rpm in a Sorvall RC-5C Plus (Sorvall, UK) for 10 min at 4 °C. All the subsequent steps were done at 4°C. After the supernatant was discarded, the cells were washed with 250 ml ice-cold water. The suspension was centrifuged at 5,000 rpm in a Sorvall RC-5C Plus (Sorvall, UK) for 10 min at 4 °C. This washing step was repeated twice. After the final centrifugation step the supernatant was immediately removed from the loose pellets. The pellets were resuspended into 10 ml of ice-cold 10% glycerol and incubated on ice for 30 minutes. Cells were subsequently pelleted (5,000 rpm for 10 min at 4°C), the supernatant removed by vacuum suction and the pellets were resuspended in 800 μ l of 10% ice-cold glycerol. This was divided into 90 μ l aliquots and frozen at -70°C.

3.3.4. Transformation of BL21(DE3) using electrocompetent cells

The electrocompetent BL21(DE3) cells were thawed on ice. To $2\mu\text{l}$ of plasmid DNA (pJC20-PfGST) $100\mu\text{l}$ of electrocompetent cells were added. This was then transferred to a pre-chilled electroporation cuvette and a pulse of 2500 V applied for 5ms in a Bio-Rad MicroPulser electroporator (Bio-Rad Laboratories, Hercules, California) (Dower *et al.*, 1988). LB liquid media (1 ml) was added directly after electroporation and the cells were incubated for 1 hour at 37°C with shaking (180 rpm), and plated on LB solid media (1% noble agar, 0.5% yeast extract, 1% NaCl, pH 7.0) supplemented with $100\mu\text{g}/\text{ml}$ ampicillin.

3.3.5. Conventional miniprep plasmid isolation

The protocol from Sambrook *et al.* (1989) was used when plasmid DNA was isolated from a 2 ml overnight culture (37°C) of the transformed cells in LB-Broth with the appropriate antibiotic ($100\mu\text{g}/\text{ml}$ ampicillin for pJC20 expression vectors). The cells were harvested by centrifugation at 13,000 rpm for 2 min (4°C) and the pellet was resuspended in $200\mu\text{l}$ of Solution I (50mM glucose, 10mM EDTA and 25mM Tris-HCl, pH 8.0). $300\mu\text{l}$ of ice cold Solution II (0.2N NaOH and 1% SDS) was added followed by incubation on ice for 5 min. $300\mu\text{l}$ of ice cold Solution III (60% 5M potassium acetate, 11.5% glacial acetic acid) was added to the reaction and incubated on ice for 15 min. The tubes were centrifuged at 13,000 rpm for 30 minutes at 4°C and the supernatant added to $900\mu\text{l}$ of phenol:chloroform:isoamyl alcohol (25:24:1). The phases were separated by centrifugation for 2 minutes at 13,000 rpm. The upper aqueous phase was transferred to a clean microcentrifuge tube and $900\mu\text{l}$ of chloroform:isoamyl alcohol (24:1) was added to remove traces of phenol. The upper DNA-containing aqueous phase was precipitated with 1 ml cold absolute ethanol for 30 min at -70°C . The plasmid DNA was sedimented by centrifugation at 13,000 rpm for 15 min (4°C), and the pellet washed with 70% ethanol. The pellet was resuspended in $50\mu\text{l}$ TE buffer (10 mM Tris-HCl, 1mM EDTA, pH 7.5) with $0.5\text{mg}/\text{ml}$ RNase A (QIAGEN, Chatsworth, California) and subjected to restriction enzyme digestion and agarose gel electrophoresis as described in Sections 3.3.1 and 3.3.2.

3.3.6. Protein expression

The system used for protein expression was the pJC20 vector for the expression of recombinant genes in *Escherichia coli* under the control of T7 RNA polymerase (Clos and Brandau, 1994). Electrocompetent expression hosts (Section 3.3.3) were freshly transformed with the pJC20 plasmid containing the PfGST gene construct and plated onto LB plates

supplemented with 100 μ g/ml ampicillin to select for the cells containing the gene.

One positive colony was picked and grown for 16 hours in 50 ml LB liquid media supplemented with 100 μ g/ml ampicillin and grown overnight with shaking (220 rpm) for population expression. This population expression took place at 25°C, 30°C and 37°C. Protein expression at 37°C was found to be the best, hence protein expression was continued at this temperature. The overnight cultures were diluted by the addition of 250 ml LB liquid medium (supplemented with 100 μ g/ml ampicillin) to 2.5 ml of the overnight culture. This was followed by growing at 37°C with shaking until an optical density of approximately 0.5 (logarithmic growth phase) at 600 nm was reached. Consequently, 0.238mg/ml isopropyl- β -D-thiogalactopyranoside (IPTG) (pEQ Lab Biotechnology, GmbH) was added to induce protein expression. After the addition of rifampicin (0.1 g/l) to the 250 ml cultures (after they were grown at 37°C for 1 hour), the cultures were left to grow for 16 hours, with shaking (220 rpm) at 37°C. Rifampicin is an antibiotic known to inhibit bacterial RNA polymerases, but not the T7 RNA polymerase. Therefore by adding rifampicin to the cultures the contaminating host cell protein concentration would be decreased (Kuderova *et al.*, 1999). This would optimize the protein expression of the glutathione S-transferase protein because it is under the control of the T7 promoter.

3.3.7. Protein extraction

Protein extraction and purification was done utilizing the methodology described by Liebau *et al.* (2002) as well as personal communications. The cells were harvested by centrifugation at 10,000 rpm for 10 minutes (Sorvall Plus RC-5C, GSA rotor, Sorvall, UK) and the pellet frozen for overnight storage (-20°C). The cells were allowed to thaw on ice and resuspended in 10 ml HEPES/EDTA (0.1M HEPES, 1mM EDTA, pH 6.5) buffer. Phenylmethylsulphonyl fluoride (PMSF), a protease inhibitor was added to a 1mM final concentration during sonication. The cells were sonicated for 10 cycles of 30 seconds pulsed sonication followed by 30 seconds incubation on ice water (Vibracell sonicator, output control 5, and duty cycle 50). The cell debris was removed by centrifugation at 10,000 rpm for 10 minutes (Sorvall Plus RC-5C, rotor SS34, Sorvall, UK). The soluble protein-containing supernatant was transferred to a pre-chilled clean centrifuge tube. A 1% streptomycin sulphate solution was added drop wise to the supernatant while slowly stirring at 4°C (Amyes and Smith, 1976). The nucleic acid debris was precipitated by centrifugation at 10,000 rpm for 10 minutes at 4°C. The protein containing supernatant was transferred to a clean tube for

overnight storage at 4°C before purification commenced.

3.3.8. Protein purification

3.3.8.1. GSTrap column specifications

The glutathione ligand was coupled via a 10-carbon linker to highly cross-linked 6% agarose. The column dimensions were 0.7cm x 2.5cm. The ligand concentration was 1.5-3.5 mg glutathione/ml (based on Glycine analysis) allowing a total binding capacity to be around 10 mg pure GST protein/ml (GST Gene Fusion System Handbook, GE Healthcare, 18-1157-58, www.amershambiosciences.com).

3.3.8.2. Affinity purification methodology

The affinity purification of the recombinant protein using the 1 ml pre-packed GSTrap column (Glutathione Sepharose, Amersham) (Liebau *et al.*, 2002) entailed the following steps. The 1 ml GSTrap (Glutathione Sepharose, Amersham) was connected to the liquid chromatography system AKTAexplorer (Pharmacia Biotech, Buckinghamshire, England) with a Frac900 fraction collector. All the subsequent steps were done at 4°C. A constant flow rate of 1 ml/min was maintained throughout the purification. The column was equilibrated with 5 column volumes of PBS binding buffer (140mM NaCl, 2.7mM KCl, 10mM Na₂HPO₄, 1.8mM KH₂PO₄ at pH 7.3). After 2 ml of the crude protein extract was filtered through a 0.45µm filter (Sartorius, Goettingen, Germany), the sample was applied to the column, at a constant flow rate of 1 ml/min. Unbound proteins were washed off with 5 column volumes of binding buffer. The enzyme was eluted by using 5 column volumes of elution buffer (50mM TRIS-HCl, 10mM reduced glutathione, pH 8.0). Fractions of 0.5 ml were collected. The column was washed with 5 column volumes of binding buffer to remove the residual elution buffer and to restore the column (Liebau *et al.*, 2002; GST Gene Fusion System Handbook, GE Healthcare, 18-1157-58).

3.3.8.3. SDS-PAGE analysis

The Laemmli method for SDS (Sodium dodecyl-sulphate) poly-acrylamide gel electrophoresis (PAGE) was used to analyse the protein separation based on molecular mass (Laemmli, 1970). Gels were prepared as follows: 4% stacking gel (4% Bio-Rad Acrylamide-Bisacrylamide Mix, 0.1% SDS, 0.05% ammonium persulphate, 0.1% TEMED, 0.05M Tris-HCl, pH 6.8), and a 12% running gel (12% Bio-Rad Acrylamide-Bisacrylamide Mix, 0.1% SDS 0.05% ammo-

nium persulphate, 0.1% TEMED, 0.375M Tris-HCl, pH 8.8) with 10 wells. A 15 μ l purified protein sample from each of the collected fractions was transferred to a clean microcentrifuge tube. Subsequently, an equal volume of denaturing buffer (1.2% SDS, 30% glycerol, 15% β -mercaptoethanol, 0.18 mg/ml bromophenol blue, 0.15M Tris, pH 6.8) was added. The protein samples were denatured at 90°C for 5 minutes. 10 μ l of each sample was loaded onto the gel. Electrophoresis was performed in a 0.025M Tris-0.2 M Glycine buffer (pH 8.3) and separated at 200V in a Bio-Rad Mini Protean 3 Electrophoresis system. Protein bands were visualized with Coomassie Blue G250 staining solution (0.1 g Coomassie Blue G250 in 40% methanol, 10% acetic acid), and destaining solution (40% methanol, 10% acetic acid).

3.3.8.4. Sample preparation for enzyme assay

All the subsequent steps were done at 4°C. Depending on the concentration of the fractions, samples with the highest protein concentration were pooled and a final protein concentration of 0.675 mg/ml was obtained. The pooled samples were transferred into SnakeSkin Dialysis Tubing (Pierce Biotechnology, Rockford, Illinois, USA) membrane. The dialysis tube was emerged in 1:100 volumes PBS buffer and slowly stirred overnight at 4°C, to remove the glutathione present in the elution buffer in order to reduce interference in the protein concentration determination as well as the enzyme assay.

3.3.8.5. Protein concentration determination

Protein concentrations were determined according to the Folin-Lowry method (Lowry *et al.*, 1951), using calibration curves constructed with bovine serum albumin (BSA). A stock solution of 0.3 mg/ml was diluted to obtain a standard concentration range of 300, 240, 210, 180, 150, 120, 90, 60 and 30 μ g/ml BSA protein. To each BSA standard reaction and protein sample, 300 μ l of Solution ABC (20:1:1) was added (Solution A: 2% Na₂CO₃ in 1M NaOH; Solution B: 1% CuSO₄ · H₂O; Solution C: 2% Potassium tartrate). After a 15 minute incubation period, 900 μ l of 10% 2N Folin-Ciocalteu (mixture of phosphotungstic acid and phosphomolybdic acid in phenol) reagent was added. The Copper(II) ion in alkaline solution B reacts with the protein to form complexes with functional groups of tyrosine, tryptophan and cysteine. These complexes react with the Folin-Ciocalteu reagent. The product becomes reduced to molybdenum/tungsten blue and can be detected colorimetrically by absorbance at 660 nm (Lowry *et al.*, 1951). Presence of strong acids or ammonium sulfate can interfere with the assay. After a 45 minute incubation period in the dark, the absorbance at 660 nm was read. After the standard curve for BSA was obtained, the slope and the y-intercept of the BSA standard curve were used to extrapolate the concentrations

for the PfGST protein samples.

3.3.9. Enzyme assays

The enzyme assays were performed using the methods described by Liebau *et al.* (2002) and Harwaldt *et al.* (2002). PfGST activity was tested at 25°C with 1mM of reduced glutathione substrate and 0.05M CDNB (1-chloro-2, 4-dinitrobenzene) as a second substrate in standard HEPES/EDTA (0.1M HEPES, 1mM EDTA, pH 6.5) buffer. To start the reaction, glutathione S-transferase enzyme was added (0.0675 mg PfGST/ml reaction). The formation of the chromogenic compound S-2,4-dinitrophenyl glutathione was followed spectrophotometrically at 340nm ($\epsilon_{340} = 9.6 \text{ mM}^{-1} \cdot \text{cm}^{-1}$). The control for the reaction was 1mM of reduced glutathione and 0.05M CDNB as a second substrate in a standard HEPES/EDTA buffer. The spectrophotometer used was a Perkin Elmer, Lambda 35 UV/Vis Spectrophotometer (Perkin Elmer, Wellesley, USA).

3.3.9.1. DMSO solvent controls

Due to the solubility problems, GTX and the other inhibitors were dissolved in absolute DMSO (dimethyl sulphoxide). The solvent controls were designed to have similar concentrations in the control and experimental samples. The inhibitor stock solutions were made up to 5mM. Hence when a 1mM inhibitor concentration was used 200 μl of DMSO was present in 1 ml sample. Table 3.3, shows the volumes of DMSO that were added to the reaction corresponding to each inhibitor concentration.

Table 3.3: The volumes of DMSO that were added to the reaction corresponding to each inhibitor concentration.

Inhibitor concentration	1mM	500 μM	300 μM	200 μM	100 μM
Volume of DMSO added	200 μl	100 μl	60 μl	40 μl	20 μl
Percentage of total reaction volume	20%	10%	6%	4%	2%

3.3.9.2. Standard GTX inhibition assay

5mM GTX was dissolved in absolute DMSO. GTX was added in five different concentrations (1mM, 500 μM , 300 μM , 200 μM and 100 μM) to the standard running assay at 25°C. The glutathione stock solution of 5mM was diluted with water to concentrations of 1mM, 500 μM , 300 μM , 200 μM , 100 μM and 75 μM . The reaction was started by adding 100 μl of 0.7

mg/ml PfGST. The formation of the chromogenic compound S-2,4-dinitrophenyl glutathione was followed spectrophotometrically at 340nm.

3.3.10. Inhibition assays

The inhibitory capacity of L-3-aminomethylphenylalanine (Peptech 57213-47-5), ethyl 4-amino-2-[(4-ethoxy-2, 4-dioxobutyl) thio]-5-pyrimidine carboxylate (Chemstep 54991) and 3-(2-Naphthyl)-D alanine (Sigma N5387) were tested. Each reaction of every assay was done in triplicate.

3.3.10.1. Inhibition of PfGST by L-3-aminomethylphenylalanine (LAP)

The LAP stock solution of 5mM was made up in absolute DMSO. LAP was added to the standard running assay in five different concentrations: 1mM, 500 μ M, 300 μ M, 200 μ M and 100 μ M. In order to be able to determine the necessary enzyme kinetics the first substrate GSH was also added in six different concentration: 1mM, 500 μ M, 300 μ M, 200 μ M, 100 μ M and 75 μ M. The reaction was started by adding PfGST and the formation of the chromogenic compound as described before over a period of 15 minutes, with reading intervals of 30 seconds.

3.3.10.2. Inhibition of PfGST by ethyl 4-amino-2-[(4-ethoxy-2, 4-dioxobutyl) thio]-5-pyrimidine carboxylate (EDP)

The EDP stock solution of 5mM was made up in absolute DMSO. LAP was added to the standard running assay, in five different concentrations: 1mM, 500 μ M, 300 μ M, 200 μ M and 100 μ M and GSH was added in the concentrations as before. The reaction was started by adding PfGST, the formation of a chromogenic was followed spectrophotometrically at 340nm as described before.

3.3.10.3. Inhibition of PfGST by 3-(2-Naphthyl)-D alanine (NDA)

NDA was dissolved in absolute DMSO and the solution was acidified with HCl to increase the solubility. The pH of the HEPES/EDTA buffer was accordingly increased with the addition of NaOH to keep the assay running pH at 6.5. The pH of the HEPES/EDTA buffer changed with each concentration of inhibitor that was used. NDA was added to the standard running assay in five different concentrations: 1mM, 500 μ M, 300 μ M, 200 μ M and 100 μ M., and the assays were conducted as described above.

3.3.11. Lineweaver-Burk double reciprocal plots

The method used for the construction of the Lineweaver-Burk double reciprocal plots was done as described by Lineweaver and Burk, (1934). Lineweaver-Burk double reciprocal plots lead to a straight line, which is preferable to the rectangular hyperbola of a Michaelis-Menten plot (Michaelis and Menten, 1913). The Lineweaver-Burk plots were constructed according to the following equation:

$$\frac{1}{V_o} = \frac{K_m}{V_{max}} \cdot \frac{1}{[S_o]} + \frac{1}{V_{max}}$$

The Lineweaver-Burk plots were drawn for each inhibitor separately. V_o was the initial velocity also known as specific activity that was determined by Beer-Lambert's law (Section 3.1.3). V_{max} was the maximum rate of the reaction, K_m was the Michaelis constant and S_o was the primary substrate concentration used. The slope of these graphs was relational to $\frac{K_m}{V_{max}}$. The x-axis intercept was used to determine the dissociation constant (K_m value) of the enzyme. The y-axis intercept was used to determine the maximum reaction rate of the enzyme (V_{max}).

For *competitive inhibition* the enzyme's affinity for the substrate changes, because a competitive inhibitor competes with the substrate for the same binding site on the enzyme (Palmer, 2001). Therefore the K_m (x-axis intercept) would change in the presence of a competitive inhibitor but the V_{max} (y-axis intercept) would stay constant. K_m would be altered so that $K'_m = K_m(1 + \frac{[I_o]}{K_i})$, where K'_m is the apparent K_m in the presence of the inhibitor at concentration I_o . By rearranging this equation the K_i value could be determined (Lineweaver and Burk, 1934).

A *non-competitive* inhibitor would inhibit the enzyme in the presence or the absence of the substrate, as the inhibitor binds at a site different from the substrate (Palmer, 2001). The inhibition constant or dissociation constant (x-axis intercept) would not be affected. The amount of active enzyme would be reduced, therefore decreasing the V_{max} (y-axis intercept) for each concentration but not altering the K_m value since neither inhibitor nor substrate hinder the binding of one another. V_{max} will be altered so that $\frac{1}{V'_{max}} = \frac{1}{V_{max}} (1 + \frac{[I_o]}{K_i})$, where V'_{max} is the apparent V_{max} in the presence of the inhibitor at concentration I_o . By rearranging this equation the K_i value could be determined (Lineweaver and Burk, 1934).

Mixed inhibition refers to a combination of two different types of reversible enzyme inhibition; competitive inhibition and uncompetitive inhibition. The term 'mixed' is used when

the inhibitor can bind to either the free enzyme or the enzyme-substrate complex. In mixed inhibition, the inhibitor binds to a site different from the active site where the substrate binds (Palmer, 2001).

3.3.12. *In silico* comparison between the binding of NDA and Praziquantel (PZQ) to PfGST

NDA was found to show mixed inhibition of PfGST, meaning that NDA did not bind to the PfGST catalytic site as predicted by AutoDock. In Section 2.4.4 the similarity in molecular fingerprints between praziquantel (PZQ) and NDA was noticed. PZQ is a known non-substrate inhibitor of *Schistosoma japonica* glutathione S-transferase (SjGST) (McTigue *et al.*, 1995). McTigue *et al.* (1995) crystallized PZQ binding to SjGST at a non-substrate site. Based on the similarity of NDA and PZQ, the binding site of NDA was further explored by superimposing the crystal structures and further energy minimizations.

The crystal structure of SjGST and its complex with PZQ (PDB ID code 1GTA) was superimposed on PfGST (PDB ID code 1Q4J) using InsightII with the HOMOLGY module. The ligand PZQ was unmerged from the SjGST structure and merged with the PfGST structure. The PZQ_PfGST structure was then subjected to 1000 steps of minimization using a Polak-Ribiere conjugate gradient algorithm as applied in the DISCOVERY3 module of InsightII. PZQ was replaced by NDA, after NDA was superimposed onto PZQ. The NDA_PfGST structure was then subjected to 1000 steps of minimization using a Polak-Ribiere conjugate gradient algorithm as applied in the DISCOVERY3 module of InsightII.

3.4. Results

3.4.1. Restriction enzyme digestion

Upon arrival the plasmid concentration was 397 ng/ μ l. The plasmid was successfully transformed into BL21(DE3) *E. coli* cells *via* electroporation. The presence of the plasmid in the cells was confirmed by antibiotic selection with 100 μ g/ml ampicillin as the pJC20 plasmid carries an ampicillin resistance marker gene. After alkaline lysis and plasmid isolation, the presence of the PfGST gene in the plasmid was confirmed with restriction enzyme digestion using BamHI and HindIII. The gel in Figure 3.3, shows the two bands. The bigger band is of

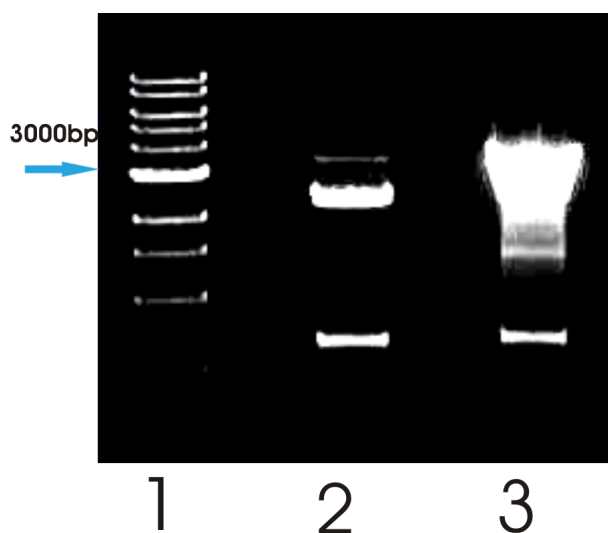


Figure 3.3: **BamHI and HindIII restriction analysis of the pJC20-PfGST plasmid.** Lane 1: 1kb DNA Ladder (New England Biolabs, Boston, USA). Lane 2: Restriction digestion of the plasmid received from Prof. Liebau. Lane 3: Restriction digestion of the plasmid propagated in the BL21(DE3) cells. The 3000bp band of the 1kb DNA Ladder is indicated.

size 2340 bp that represents the pJC20 plasmid back-bone and the smaller band represents the PfGST gene with a size of 636 bp. This was an indication that the plasmid had been transformed successfully into the BL21(DE3) cells.

3.4.2. Protein expression

Overnight protein expression took place at three different temperatures: 25°C, 30°C and 37°C. It was concluded that the protein expression was optimal at 37°C. A thick band at 24.8 kDa in Figure 3.4 shows the analysis, including the difference between the soluble and insoluble fractions at 25°C, 30°C and 37°C, on a 12% SDS-PAGE gel. This comparison in

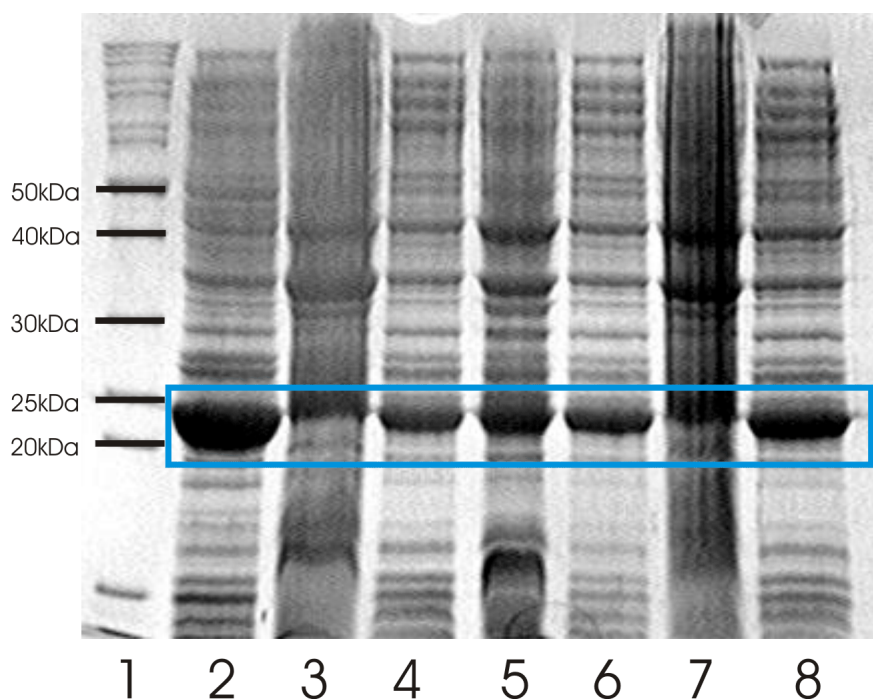


Figure 3.4: **SDS-PAGE analysis of protein expression at different temperatures.** Lane 1: PageRuler Protein Ladder (Fermentas). Lane 2: Total protein fraction at 37°C. Lane 3 (insoluble) and Lane 4 (soluble) at 25°C. Lane 5 (insoluble) and Lane 6 (soluble) at 30°C. Lane 7 (insoluble) and Lane 8 (soluble) at 37°C. PfGST indicated in blue as a band at 24.8 kDa. Lanes 3-8 were loaded quantitatively.

Figure 3.4 revealed that there was enzyme present in the insoluble fraction (Lane 3) but there was enough protein present in the soluble fraction (Lane 4) to continue with the experiment by using just the soluble fraction, at 24.8 kDa.

A comparison between 24.8 kDa bands of the soluble IPTG-induced fraction (Lane 2) and uninduced fractions (Lane 3) is shown in Figure 3.5, on a 12% SDS-PAGE gel.

3.4.3. Protein purification

After the presence of the enzyme had been confirmed on a 12% SDS-PAGE gel, the soluble fraction was used in the protein GSTrap purification. From the 12% SDS-PAGE gel in Figure 3.6 it was evident that the protein purification was successful, since there was a single band at 24.8 kDa. The fractions with the highest protein concentration (based on intensity of the SDS-PAGE band) were pooled and dialyzed overnight in PBS buffer before using it in the assay.

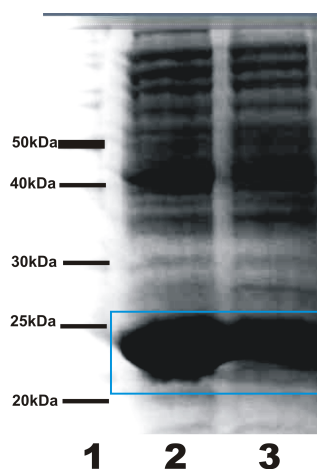


Figure 3.5: **SDS-PAGE analysis induced and uninduced protein expression.** Lane 1: PageRuler Protein Ladder (Fermentas). Lane 2: Induced by IPTG. Lane 3: Uninduced. The 50kDa band of the PageRuler Protein Ladder (Fermentas) is indicated. The expected size of PfGST was 24.8 kDa, as shown in blue.

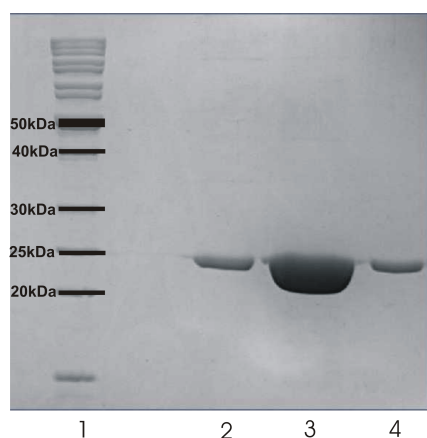


Figure 3.6: **SDS-PAGE analysis of affinity protein purification with a GSTrap column.** Lane 1: PageRuler Protein Ladder (Fermentas). Lane 2: Fraction 2. Lane 3: Fractions 3. Lane 4: Fractions 4. Fractions of 0.5 ml were collected by the liquid chromatography system AKTAexplorer (Pharmacia Biotech, Buckinghamshire, England) with a Frac900 fraction collector. The PfGST band can be seen at 24.8 kDa.

3.4.4. Protein concentration determination

A bovine serum albumin standard curve for protein concentration was produced and quantified by the Folin-Lowry method (Lowry *et al.*, 1951). The standard curve shown in Figure 3.7 has an R-square value of 0.998, this is an indicator of how well the linear regression model fits the data. By the method of extrapolation of the linear part of the standard curve

the protein concentration was determined to be 13.5 PfGST mg/L culture. PfGST was known to express at concentrations of 14mg PfGST enzyme/L cell culture (Harwaldt *et al.*, 2002).

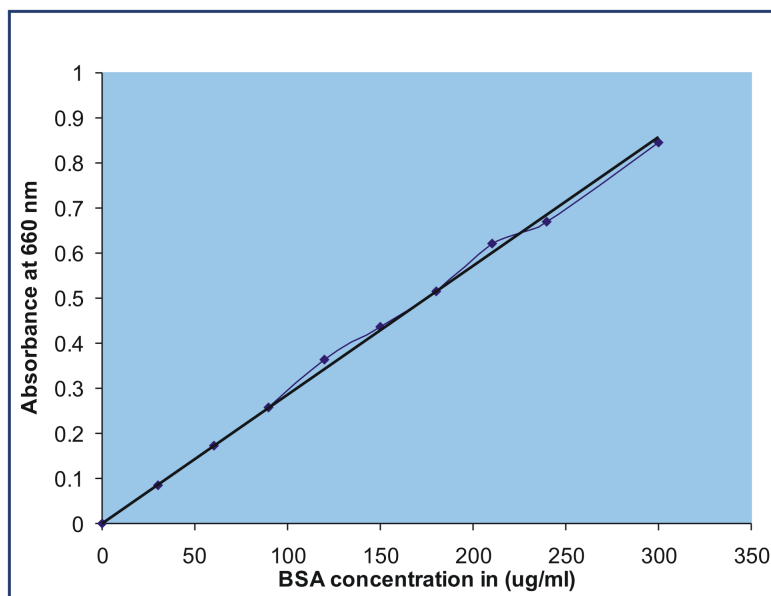


Figure 3.7: **Standard curve for protein concentration determination constructed by the Folin-Lowry assay.** Blue line: Bovine Serum Albumin (BSA) protein standards. Black line: Linear regression of the data provided by BSA protein standards.

3.4.5. Enzyme assay

A Perkin Elmer, Lambda 35 UV/VIS Spectrophotometer was used to measure the absorbance of S-2,4-dinitrophenyl glutathione at 340nm every 30 seconds for 15 minutes. By using the Beer-Lambert law this rate of the reaction (Δ Absorbance/time) was converted into specific activity of the enzyme during the linear reaction in $\mu\text{mol}/\text{min}/\text{mg}$ or U/mg. The average specific activity of PfGST towards GSH and CDNB in a standard running assay was determined to be 0.132 U/mg (data not shown). This was lower than the 0.2 U/mg that Harwaldt *et al.* (2002) obtained.

3.4.6. Inhibition assays

Since solubility problems arose, EDP and NDA had to be dissolved in DMSO. Consequently, the effect of DMSO on the enzyme had to be evaluated. The maximum volume of absolute DMSO added to the reaction would be $200\mu\text{l}$ when a 1mM final concentration of inhibitor was tested. When $200\mu\text{l}$ 100% DMSO was added, the specific activity

had a standard deviation of 0.132 ± 0.00055 U/mg between those samples with and without DMSO. Therefore, it was concluded that DMSO had no statistically significant influence on enzyme activity and all inhibitors were dissolved in DMSO. In Figure 3.8 PfGST activity was tested with 1mM of reduced glutathione substrate. Figure 3.8 shows the inhibitory effect that GTX, EDP, LAP and NDA had on PfGST at inhibitor concentrations of 1mM (blue), 500 μ M (purple) and 100 μ M (green). From this graph can be deduced that NDA showed the highest percentage inhibition of PfGST, followed by GTX, LAP and EDP.

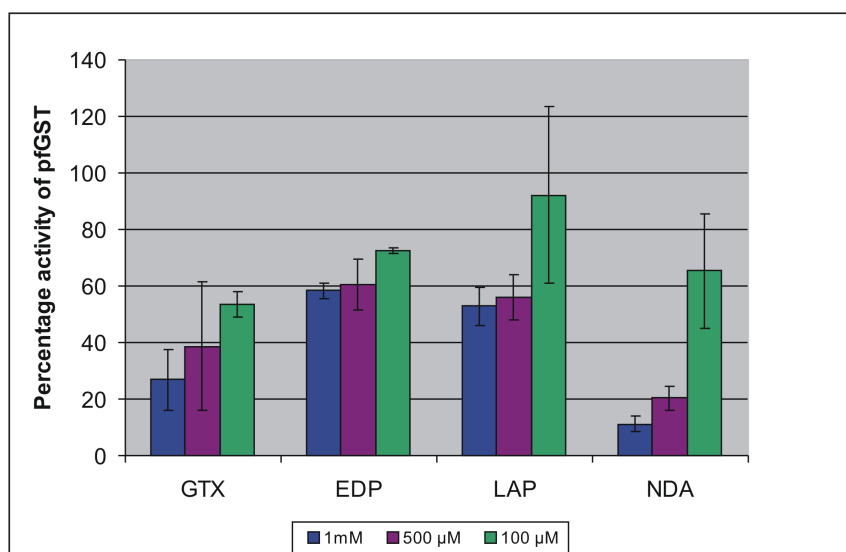


Figure 3.8: **The effect of the inhibitors GTX, EDP, LAP and NDA on PfGST activity.** Decrease in activity as a result of inhibitors, original PfGST activity (0.132 U/mg) was regarded as 100% activity and total inhibition as 0% activity. Concentrations of inhibitors: 1mM (blue), 500 μ M (purple) and 100 μ M (green).

3.4.7. Enzyme kinetics

Figure 3.8 showed inhibition of PfGST by GTX, EDP, LAP and NDA, but the type of inhibition was still undetermined. All the assays were repeated at different inhibitor concentration (1mM, 500 μ M, 300 μ M, 200 μ M and 100 μ M). The GSH concentrations were varied accordingly (1mM, 500 μ M, 300 μ M, 200 μ M, 100 μ M and 75 μ M). The Lineweaver-Burk plots in Figure 3.9 were constructed according to the method described by Lineweaver and Burk, (1934). NDA could be classified as a non-competitive inhibitor because the V_{max} changed but the K_m was not influenced by the addition of inhibitor.

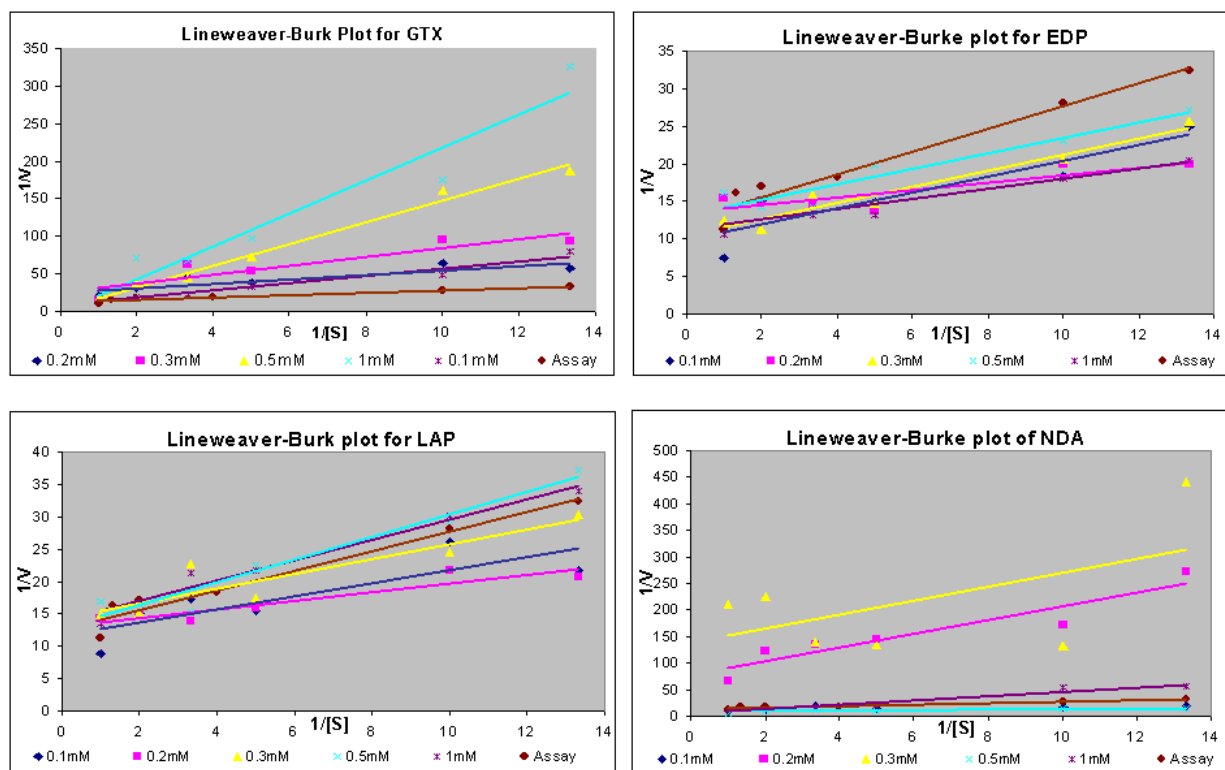


Figure 3.9: **Lineweaver-Burke plots for GTX, LAP, EDP and NDA.** The double reciprocal plots of Initial velocity (V_0) versus Substrate concentration (S_0) for inhibitors A) GTX, B) LAP, C) EDP and D) NDA.

Although Lineweaver-Burk is still one of the most widely used kinetics models it has been criticized on the following. Firstly, it has the tendency to compress the data points at the higher substrate concentrations, consequently shifting the weight towards the data points at lower substrate concentrations where the results are more likely to be erroneous. Secondly, the extrapolation across the $1/V_0$ axis can lead to the alteration of the axis. Lastly, the deviation from linearity is less obvious in Lineweaver-Burk than in Eadie-Hofstee plots (Eadie, 1942; Fersht, 1943; Hofstee, 1959). Hence, the graphical data representation of the inhibition kinetics was done according to Eadie-Hofstee plots as well in Figure 3.10.

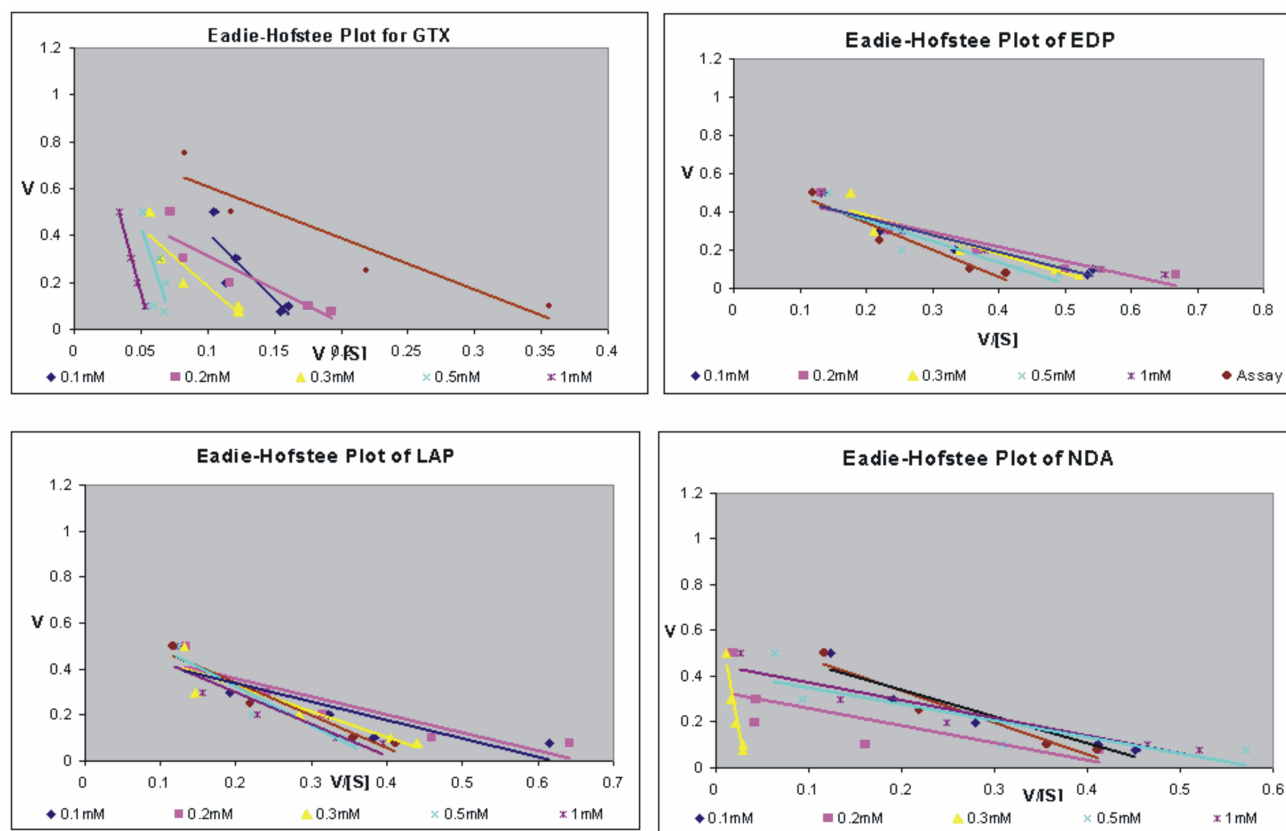


Figure 3.10: **Eadie-Hofstee plots for GTX, LAP, EDP and NDA.** The plots of Initial velocity (V) versus Initial velocity (V_0) divided by Substrate concentration (S) for inhibitors A) GTX, B) LAP, C) EDP and D) NDA.

When extracting kinetic parameters nonlinear regression analysis should be used on data to determine if the changes in V_{max} is significant. However, the graphical analysis of the data (Figure 3.9 and Figure 3.10) is adequate for a preliminary screen because the values span a large range so a trend could be picked up.

3.4.8. *In silico* comparison between the binding of NDA and Praziquantel (PZQ) to PfGST

The PZQ ligand binds to SjGST in the dimer interface groove adjoining the two catalytic sites (McTigue *et al.*, 1995). Figure 3.11 (A) represents a stereo view of the SjGST dimer with PZQ located at the non substrate binding site. The tyrosine residue that forms part of the substrate binding site is indicated in green. In Figure 3.11 (B), the PZQ ligand was merged into PfGST in a structurally similar orientation as in the SjGST. NDA ligand (shown in purple) superimposed on the PZQ ligand (shown in orange) in the PfGST dimer, can be

seen in Figure 3.11 (C). The hypothetical binding orientation of NDA to PfGST can be seen in Figure 3.11 (D). The hypothetical binding site of NDA to PfGST as presented in Figure 3.11, can be validated only by future site directed mutation or crystalization studies.

3.5. Discussion

3.5.1. Overexpression and purification of recombinant PfGST

PfGST was known to express at concentrations of 14 mg PfGST enzyme/L cell culture (Harwaldt *et al.*, 2002), initially protein concentrations of 6.7 mg/L cell culture were observed. Two changes were made to the original method used by Harwaldt *et al.* (2002), in order to optimize protein expression and purification. The first was the addition of 0.1 g/L rifampicin to inhibit bacterial RNA polymerases and decrease contaminating host cell proteins (Kuderova *et al.*, 1999). Secondly, a streptomycin sulphate precipitation step was added to precipitate some of the contaminating nucleic acids (Ameyes and Smith, 1976). Protein concentrations of 13.5 mg PfGST enzyme/L culture were observed.

3.5.2. Enzyme assays

GTX as an inhibitor of PfGST (Harwaldt *et al.*, 2002) inhibits PfGST competitively at a K_i value of 0.035 mM. In this study GTX served as a positive control for the inhibition assays. GTX was also found in this study to competitively inhibit PfGST at a K_i value of 0.01909 ± 0.0132 mM.

Kinetic studies were done to characterize the interactions between the inhibitors and PfGST in more detail. EDP and LAP showed competitive inhibition but the K_i values were significantly higher than that of GTX. Consequently, none of these inhibitors showed the potential to be explored as lead anti-malarials. The possibility exists to couple these G site inhibitors to H site inhibitors, forming double-headed inhibitors.

With respect to GSH and CDNB, NDA was found to be a non-competitive inhibitor. Historically, non-competitive GST inhibitors have been named ligandin inhibitors (Lyon *et al.*, 2003), which are hydrophobic planar compounds with anionic functional groups or steroids (Mahajan and Atkins, 2005). Therefore, it was suggested that NDA binds to a non-substrate binding site that may lead to conformational changes of the enzyme, hence leading to a loss in enzyme activity (Palmer, 2001). It is widely speculated that these ligandin type inhibitors bind the H site partially but also span the cleft between the two

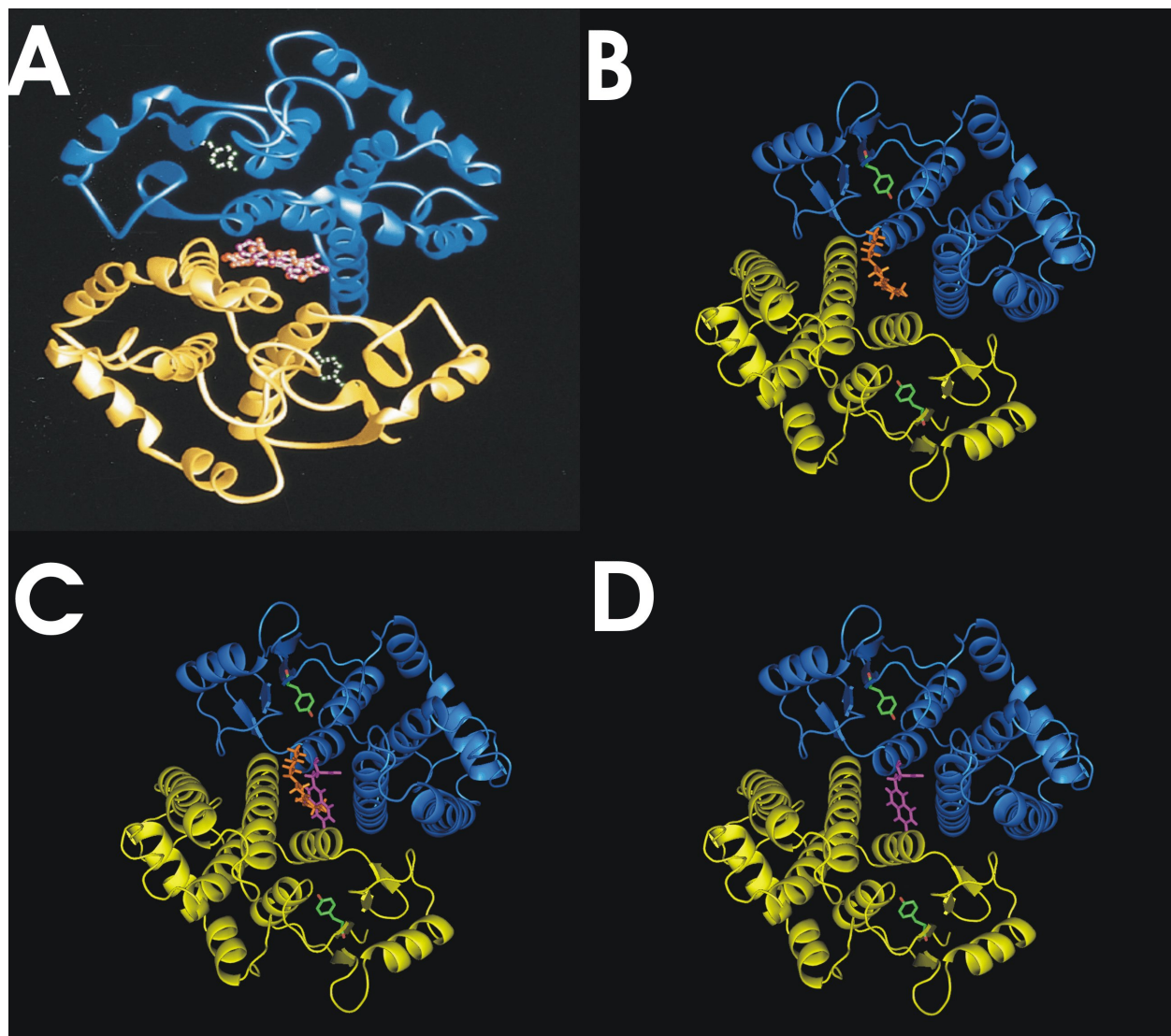


Figure 3.11: **PfGST and SjGST dimer with PZQ and NDA located at the non-substrate binding site.** **A:** The two SjGST monomers are shown in blue and yellow, with the tyrosine residue in green and the PZQ ligand shown in orange in a ball and stick representation. **B:** The two PfGST monomers are shown in blue and yellow, with the tyrosine residues in green and the PZQ ligand shown in orange in stick representation. **C:** The NDA ligand (shown in purple) superimposed on the PZQ ligand (shown in orange) in the PfGST dimer. **D:** The two PfGST monomers are shown in blue and yellow, with the tyrosine residues in green and the NDA ligand shown in purple stick representation (PYMOL).

subunits (Lyon *et al.*, 2003; McTigue *et al.*, 1995). This was not noticed when the binding mode was predicted by AutoDock, the reason for this may have been because there were reasonable interactions that formed between NDA and PfGST in the G site and therefore the program did not explore the cleft between the two monomers as a putative binding site. It might be possible that NDA has the same pattern of recognition as praziquantel, a ligandin inhibitor. Praziquantel is a known glutathione S-transferase inhibitor of *Schistosoma japonica*. The PZQ and NDA structures are depicted in Figure 3.12, and the maximum

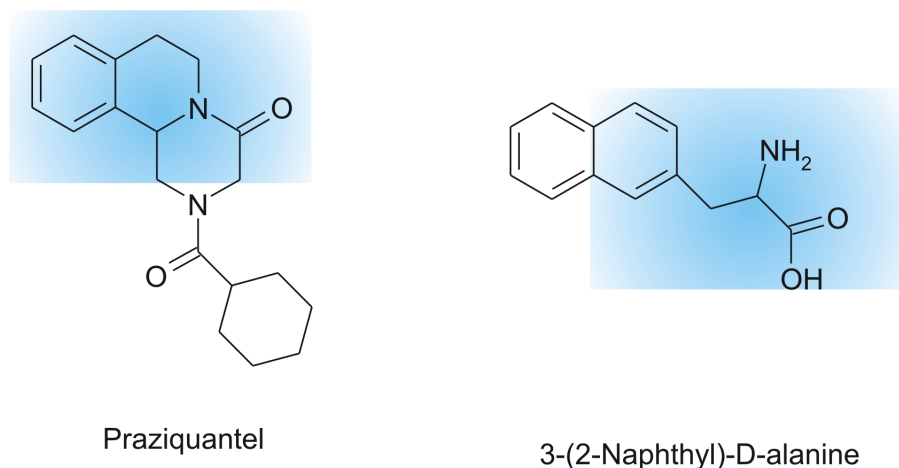


Figure 3.12: **Chemical structures of Praziquantel and NDA.** The structural comparison between Praziquantel and NDA with the maximum common substructure highlighted by the blue boxes.

common substructure in the blue boxes. A maximum common substructure is a scaffold of molecular fingerprints that two molecules share (Stahl and Mauser, 2005). Based on the similarity between PZQ and NDA the following hypothetical binding site was elucidated and can be seen in Figure 3.11. The elucidation of the mechanism of inhibition as well as the site where NDA binds will be further investigated but did not form part of this study.

3.5.3. Comparison between *in silico* predictions and biological results

Can this study validate the structure-based inhibitor design methodology used in application to *Plasmodium falciparum* glutathione S-transferase? The main aim of this chapter was therefore, to validate the strategy of ligand design that was used as well as the consensus scoring affinity prediction methods. Validation took place by means of testing the designed compounds and relating the *in silico* predictions with the inhibition data.

Table 3.4: Comparison between *in silico* and biological assay data of GTX, NDA, EDP and LAP. The PfgGST activity was expressed as percentages in the presence of 1mM and 100 μ M of the inhibitor.

Inhibitors	AutoDock scores	LUDI scores	xScore	LogP	Consensus scores	Ki values
GTX	-15.01 kcal/mol	587	6.03	-1.07	27.98	19.1 μ M
LAP	-9.16 kcal/mol	526	5.41	-0.09	19.83	974 μ M
EDP	-8.54 kcal/mol	522	4.83	1.17	18.59	712 μ M
NDA	-9.58 kcal/mol	535	5.93	2.05	20.86	52.1 μ M

When relating the values obtained from the *in silico* ligand screening in Table 3.4 with the experimentally determined inhibition constants there is a positive correlation between the *in silico* scores and the data obtained from biochemical assays. The consensus scores were calculated as described in Section 2.4.3. GTX was expected to have the highest affinity for PfgGST with a score of 27.98, then NDA with 20.86, LAP with 19.83 and lastly EDP with 18.59.

When comparing these compounds *in silico* the consensus score was used as a guide to ranking of the compounds according to binding affinity. Similarly, the inhibition constants (K_i) were used to rank the inhibitory action of the compounds. *In silico* NDA was ranked second to GTX. This is a perfect correlation to the *in silico* data. *In silico* LAP was predicted to be the better inhibitor but the biological data contradict the *in silico* data because EDP has a lower inhibition constant than LAP. This implies that the docking and scoring functions need to be optimized to rank the ligands more accurately.

The LogP method of determination was found to be accurate since LAP (LogP -0.09) could be dissolved in water but EDP (LogP 1.17) could only be dissolved in DMSO. NDA (LogP -2.05) was insoluble in the organic solvents such as DMSO and as a result the DMSO solution had to be acidified (pH 4) until the compound dissolved. Consequently, the running buffer had to be adjusted in order to keep the running pH the same but at the same time keeping NDA in solution.

The biological data derived from these inhibition studies had two major implications. Firstly, EDP, LAP and NDE were novel ligands and were predicted to have complementary interactions with PfgGST, but were also shown to have relative low binding constants in the biological assays. Hence it can be concluded that these experimental results add confidence to the discriminative power of the structure-based ligand design strategy, as applied to PfgGST.

Since the designed ligands that were tested did inhibit PfGST, it proved that **NEWLEAD** and **LUDI** can be used to design inhibitors to bind to PfGST. The K_i values of all four inhibitors were in the micromolar range, resultingly **AutoDock**, **LUDI** and **XScore** can be used to do affinity predictions for ligands binding to PfGST. Secondly, **NDA** was identified as the first novel inhibitor to PfGST that explores a non-substrate binding site. If this inhibitor is parasite specific, **NDA** might prove to be a lead compound for the design of future antimalarial drugs.

It can therefore be hypothesized that by using the already existing data on ligand design, ligand protein docking and binding energy scoring with PfGST as model protein even better inhibitors could be synthesized.

Chapter 4

Concluding Discussion

The increase in resistance of malaria parasites to available drugs lead to the consideration of new chemotherapeutic targets. It is known that the antioxidant defense system is very important to the parasite, although the interactions are not completely understood yet (Yeh and Altman, 2006). The *P. falciparum* glutathione S-transferase enzyme belongs to a super family of multi-functional, dimeric, phase II detoxification enzymes that can bind various xenobiotic electrophilic substrates. GST activity plays an important role in cancer resistance, cell proliferation, oxidative stress as well as parasitic diseases like malaria and schistosomiasis (Armstrong, 1997). Most organisms possess more than one GST isozyme, the *P. falciparum* parasite however possesses only one isozyme. Srivastava *et al.* (1999) proved that the malarial parasites died if PfGST was inhibited because it leads to the creation of a hazardous milieu inside the parasitic cells due to the accumulation of antimalarial and other toxic metabolites. Upon classifying PfGST into a structural class it was seen that the μ -loop is significantly shorter than most other μ -class GSTs. This discrepancy leaves room for exploitation of the more solvent accessible active site of PfGST as well as enabling the design of PfGST specific inhibitors (Harwaldt *et al.*, 2002; Liebau *et al.*, 2002; Fritz-Wolf *et al.*, 2003).

There are various factors that may influence the success of designing an inhibitor *in silico*. The three major contributors to the failure of drug design efforts are the lack of information, infrastructure (hardware, software, trained scientists, etc.) and cost (Congrewe *et al.*, 2005). In this study these three problems were encountered as well, and partial solutions were provided as will be discussed below.

One of the primary information requirements of a structure-based ligand design strategy is a 3D structure of the target protein. Most *Plasmodium* proteins are very difficult to express in high enough quantities for crystallization purpose, therefore most studies have to

suffice with homology models of the protein. Fortunately, the crystal structure of PfGST was available in January 2004, when this masters study commenced (PDB ID code 1Q4J). A structure-based ligand design strategy seemed the best option to utilize in finding inhibitors to bind to PfGST. In order to design novel drugs, specific enzyme information regarding the active site and binding mode was used. In most cases this information can be obtained from inhibitors or known substrates of the protein. Very little was known about the substrates (Harwaldt *et al.*, 2002) as well as inhibitors of the PfGST enzyme (Harwaldt *et al.*, 2002; Fritz-Wolf *et al.*, 2003). Ligand design was therefore based on the structure of GTX that was crystallized within the PfGST active site. The binding mode and kinetic data for GTX were available (Harwaldt *et al.*, 2002; Liebau *et al.*, 2002). Therefore, GTX was not only used as the fundamental substrate for modification, but also used as a positive control in the *in silico* and biochemical screening of the ligands.

Infrastructure and economic considerations can be seen as the two sides of a coin where research is concerned. The most critical consideration in antimalarial drug development is the resources and financial constraints. Since malaria is mostly a disease that targets developing countries, antimalarial drugs need to be very inexpensive. Additionally, investments in antimalarial drug discovery and development have been small (Rosenthal, 2003). Hence, this process is reliant on methods that may prevent unnecessary costs. Current efforts to find cost effective antimalarials include the optimization of available therapies, usage of combination therapies, developing analogs of current drugs, natural product development, agents that were developed against other diseases, resistance reversers and exploitation of novel targets (Rosenthal, 2003). Amongst all these efforts, structure-based drug design is one of the fastest growing research fields, the reason being that most of the work can be done *in silico* to eliminate the compounds that do not show the potential to become antimalarial drugs, leads to a cost effective conquest. Regrettably, there is no standard methodology that can be followed that will guarantee a successful drug. Apart from Accelrys, all the programs employed were freely available. Since commercialization has been a huge driving force in the field of drug design, most of the highly developed and very functional programs were only commercially available. Programs like NEWLEAD, AutoDock, LIGPLOT and XScore were freely available. Some of the biggest pitfalls of open source software are the lack of documentation, user manuals, clear examples as well as poorly written graphical user interfaces (Geldenhuys *et al.*, 2006). It is very time consuming to install these programs and perform parameter optimization before utilizing the program.

Consensus scoring was used to prioritize the ligands for synthesis or purchase. These filters were also necessary to reduce the costs by eliminating inhibitors that would not bind to the enzyme. When comparing the relationship between the LUDI scores and the K_i values from literature to this study, the LUDI algorithm was programmed to be: $\text{LUDI score} = 100 \log K_i$ (LUDI User Manual, March 2000, MSI, San Diego). Gradler *et al.* (2001), found experimentally that an inhibitor with a score of 541 had an inhibition constant value of 8.3 μM . GTX was proven to have a K_i of 35 μM (Harwaldt *et al.*, 2002). In this study GTX had a LUDI score of 587 and a K_i of 19.1 μM , this value is higher than the expected K_i based on the LUDI score but lower than the previously proven experimental value. In literature no direct correlation between XScore, AutoDock and K_i values were found. XScore and AutoDock were proven in various studies (Wang *et al.*, 2003; Zang *et al.*, 2005) to be able to rank ligands according to their affinity for a protein. In this study both XScore and AutoDock ranked the inhibitors according to their biologically determined K_i value.

Interesting findings that were discovered during this study were that two of the ligands that were tested, EDP and LAP both inhibited PfGST competitively and based on the docking studies done with AutoDock, it is hypothesized that EDP and LAP compete with GSH to bind to the G site. Inhibitors that compete with GSH will influence the GSH levels in the cells as well as disrupt enzyme action. This will lead to an enhanced effect of interference in the parasite metabolism.

There are other inhibitors known to bind to GST enzymes at non-substrate sites and it is hypothesized that NDA may also be binding at a site other than the G and H site. It might be possible that NDA has the same pattern of recognition than praziquantel, a ligandin inhibitor. Praziquantel was a known glutathione S-transferase inhibitor from *Schistosoma japonica* and its complex leads to anti-schistosomal activity (McTigue *et al.*, 1995). By exploring the binding mode of this inhibitor, it might be possible to determine a non-substrate binding site. Future studies would be to crystallize the NDA and PfGST complex or do mutation studies to determine the key amino acids to which NDA binds. The advantage of having a ligandin inhibitor is that it may slow down the development of resistance against the drug, as well as that specificity towards the parasitic PfGST could be increased.

The activity of EDP and LAP could be enhanced by making use of either a bivalent inhibitor design strategy or a conjugation strategy. The principle of polyvalency is to use a single molecule to inhibit multiple binding domains or active sites of multi-subunit proteins

(Lyon *et al.*, 2003). One design would imply identical molecules and the other will use different molecules. The first application of bivalency to PfGST would be to use EDP and LAP to inhibit the G site of the enzyme, then use linker fragments to bind the two identical molecules together. This will imply that the G site on both monomers in the dimer will be inhibited by a symmetrical molecule. This will be similar to a study conducted by Meado *et al.* (2006). Ethacrynic acid was known to have a IC_{50} value of $13\mu M$ when binding to SjGST. When two ethacrynic molecules were linked to form a bivalent inhibitor the IC_{50} value decreased to 13.7 nM. An increase in selectivity towards the SjGST α -class of 75 fold was noted (Meado *et al.*, 2006). The careful design of linker fragments for LAP and EDP could increase the inhibitory capacity and parasite specificity.

Another bivalent strategy would be to couple these G site inhibitors to inhibitors of the H site, consequently both active sites would be inhibited in a competitive manner. This strategy would depend on the design of competitive inhibitors to the H site of the PfGST enzyme, this might be accomplished by modifications of other known secondary substrates like CDNB. Considering this results it suggests that the H site should be better exploited in order to find more potent inhibitors (Lyon *et al.*, 2003).

Joa *et al.* (2006) used glutathione analogs (lowest IC_{50} $147\mu M$) that bind the G site of the SjGST and coupled them to steroids (lowest IC_{50} $128\mu M$) that bind the intermonomer cleft (like PZQ). The resultant compounds had a significant increase in inhibitory capacity (lowest C_{50} $7\mu M$). By coupling GSH analogs, LAP or EDP to PfGST specific steroids (or NDA after the binding site was validated) could increase the inhibitory capacity since there is a greater variation in the amino acids present in the cleft, this might help making the inhibitors parasite-specific.

The successful design and ranking of inhibitors that inhibit PfGST at K_i values below 1mM, has some major implications. Firstly, it provided an *in silico* methodology than could be developed further to form a ligand design pipeline that would provide a scientist with ligands that could be used as scaffolds for future drug design. This methodology was very cost effective since most of the programs used were freely available. Minimal information that needed to be available was a 3D structure of the receptor protein and a single ligand or inhibitors. Availability of more inhibitors would increase the diversity of the ligands designed by this methodology. Molecular docking and interaction scoring functions could also be employed in the form of AutoDock, LUDI score and XScore to rank the designed

ligands for biological assays. Secondly, it provided inhibitors to PfGST that could be used as chemical scaffold for future antimalarial drug development.

In conclusion, the future prospect does exist that the better scoring inhibitors can be synthesized and tested. The hypothesis would then be that those designed inhibitors that scored marginally better than GTX *in silico* during this study, would have a lower inhibition constants when tested in biological assays as well, keeping the prospect of finding an antimalarial lead compound alive. This study was used to provide a proof-of-principle that this approach can be used to design ligands that bind PfGST and predict a consensus score that can be used to rank these ligands according to their affinity for PfGST.

Summary

The primary aim of this study was to use a computational structure-based ligand design strategy in finding novel ligands that could act as inhibitors of PfGST as basis for future antimalarial drug development. Since there is only one PfGST isoenzyme present in the parasite and the architecture of the binding site differs significantly from its human counter part, PfGST is considered a highly attractive drug target. Inhibition of PfGST is expected to interfere at more than one metabolic site in synergy: it is likely to disrupt the glutathione-dependent detoxification process, which will lead to an increase in the cytotoxic peroxide concentration and most likely lead to an increase in the levels of ferriprotoporphyrin IX and hemin as well. S-hexyl glutathione was co-crystallized with PfGST (Harwaldt *et al.*, 2004), consequently it was seen as one of the most important lead compounds in the development of PfGST inhibitors.

The first step in the rational drug design strategy was to modify GTX, concentrating on its ability to bind competitively to the G site and the hydrocarbon chain protrudes into the H site as well. Considering the 3D structure of the enzyme, modifications to GTX were made by LUDI and NEWLEAD, resulting in a library of active site binding ligands ranked by AutoDock according to their ability to optimally bind to PfGST. Additionally, the ligands were ranked according to their affinity for binding to PfGST produced by AutoDock, LUDI and XScore.

Once all the compounds were ranked by these *in silico* methods they were screened for acquisition or synthetic accessibility and those available were experimentally screened for activity against recombinantly expressed PfGST. Based on *in silico* predictions NDA was the best inhibitor followed by LAP and EDP. From the biological assay and Lineweaver-Burk analysis the order of inhibition was NDA as the best inhibitor tested, followed by LAP and EDP. EDP and LAP showed competitive inhibition but the inhibition constant values were significantly lower than GTX. With respect to GSH and CDNB, NDA was found to be a non-competitive inhibitor. It was suggested therefore that NDA binds to a non-substrate

binding site that may lead to conformational change of the enzyme and hence lead to a loss in enzyme activity. This data leads to the conclusion that the H site should be better exploited in order to find more potent inhibitors or non-substrate binding sites.

It was concluded that the experimental results add confidence to the discriminative power of the structure-based ligand design strategy and that these inhibitors could form scaffolds for future antimalarial drug development.

References

- Abagyan, R. and Totrov, M. (2001). High throughput docking for lead generation. *Current Opinion in Chemical Biology* **5**, 375-382.
- Ajay, A. and Murcko, M.A. (1995). Computational methods to predict binding free energy in ligand receptor complexes. *Journal of Medicinal Chemistry* **38**, 4953-4967.
- Amyes, S.G.B. and Smith, J.T. (1976). The purification and properties of the trimethoprim-resistant dihydrofolate reductase mediated by the R-factor, R 388. *European Journal of Biochemistry* **61** (2), 597-603.
- Anderson, A.C. and Wright, D.L. (2005). The design and docking of virtual compound libraries to structures of drug targets. *Current Computer Aided Drug Design* **1**, 103-127.
- Arav-Boger, R. and Shapiro, T.A. (2005). Molecular mechanisms of resistance in antimalarial chemotherapy, the unmet challenge. *Annual Reviews in Pharmacology and Toxicology* **45**, 565-585.
- Armstrong, R.N. (1997). Structure, catalytic mechanism and evolution of the glutathione transferases. *Chemical Research in Toxicology* **10**, 2-18.
- Biagini, G.A., O'Neill, P.M., Nzila, A., Ward, S.A. and Bray P.G. (2003). Antimalarial chemotherapy, young guns or back to the future? *Trends in Parasitology* **11b** 479-448.
- Becker, K., Rahlfs, S., Nickel, C. and Schirmer, R.H. (2003). Glutathione, functions and metabolism in the malarial parasite *Plasmodium falciparum*. *Biological Chemistry* **384** (4), 551-566.
- Becker, K. and Kirk, K. (2004). Of malaria, metabolism and membrane transport. *Trends in Parasitology* **20** (12), 590 -596.
- Becker, K., Tilley, L., Vennerstrom, J.L. and Roberts, D. (2004). Oxidative stress in malaria parasite-infected erythrocytes, host-parasite interactions. *International Journal of Parasitology* **34** (2), 163-189.
- Beer, A. (1854). Braunischweig. *Grundriss des Photometrischen Calculs*.

- Bemis, G.W. and Murcko, M.A. (1996). The properties of known drugs. Molecular frameworks. *Journal of Medicinal Chemistry* **39**, 2887-2893.
- Bohm, H.J. (1992). LUDI, rule-based automatic design of new substituents for enzyme inhibitor leads. *Journal of Computer-Aided Molecular Design*. **6** (6), 593-606.
- Bozdech, Z. and Ginsburg, H. (2004). Antioxidant defense in *Plasmodium falciparum*, data mining of the transcriptome. *Malaria Journal* **3**, 23-27.
- Brooijmans, N. and Kuntz, I.D. (2003). Molecular recognition and docking algorithms. *Annual Review of Biophysics and Biomolecular Structure* **32**, 335-373.
- Brown, M.W. and Van der Jagt, H. (2004). Creating artificial binding pocket boundaries to improve the efficiency of flexible ligand docking. *Journal of Chemical Information and Computer Sciences* **44**, 1412-1422.
- Burg, D. and Mulder, G.J. (2002). Glutathione conjugates and their synthetic derivatives as inhibitors of glutathione dependent enzymes involved in cancer and drug resistance. *Drug Metabolism Reviews* **34** (4), 821-863.
- Burmeister, C., Perbandt, M., Betzel, C.H., Walter, R.D. and Liebau, E. (2003). Crystallization and preliminary X-ray diffraction studies of the glutathione S-transferase from *Plasmodium falciparum*. *Acta Crystallographica Biological Crystallography* **59** (8), 1469-1471.
- Caffisch, A., Walchli, R. and Ehrhardt, C. (1998). Computer-Aided Design of Thrombin Inhibitors. *News in Physiological Sciences* **13**(4), 182-189.
- Cardoso, R.F.M., Daniels, D.S., Bruns, C.M. and Tainer, J.A. (2003). Characterizations of the electrophile binding site and substrate binding mode of 26kDa glutathione S-transferase from *Schistosoma japonicum*. *Proteins* **5**, 137-146.
- Clark, R.D., Strizhev, A., Leonard, J.M., Blake, J.F. and Matthew, J.B. (2002). Consensus scoring for ligand/protein interactions. *Journal of Molecular Graphics and Modelling* **20**, 271-295.
- Clos, J. and Brandau, S. (1994). pJC20 and pJC40—two high-copy-number vectors for T7 RNA polymerase -dependent expression of recombinant genes in *Escherichia coli* *Protein Expression and Purification* **5** (2), 133-137.
- Congrewe, M., Murry, C.W., and Blundell, T.L. (2005). Structural biology and drug discovery. *Drug Discovery Today* **10** (13), 895-907.
- Delfino, R.T., Santos-Filho, O.A. and Figueroa-Villar, J.D. (2002). Molecular modeling of wild-type and antifolate resistant mutant *Plasmodium falciparum* DHFR. *Biophysical*

Chemistry **98**, 287-300.

De Witte, R.S. and Shakhnovich, E.L. (1996). SMOG, *de novo* design method based on simple, fast and accurate free energy estimates. *Journal of the American Chemistry Society* **118**, 11733-11744.

Dirr, H.W. (2001). Folding and assembly of glutathione transferases. *Chemico-Biological Interactions*. **1333**, 19-23.

Dixon, P.D., Laphorn, A. and Edwards, R. (2002). Plant glutathione transferases. *Genome Biology* **3**, 1-10.

Dower, W.J., Miller, J.F. and Ragsdale, C.W. (1988). High efficiency transformation of *E.coli* by high voltage electroporation. *Nucleic Acids Research* **16** (13), 6127-6145.

Dubois, V.L., Platel, D.F., Pauly, G. and Tribouley-Duret, J. (1995). *Plasmodium berghei*, implication of intracellular glutathione and its related enzyme in chloroquine resistance *in vivo*. *Experimental Parasitology* **81** (1), 117-124.

Dym, O., Xenariosi, I., Ke, H. and Colicelli, J. (2002). Molecular docking of competitive phosphodiesterase inhibitors. *Molecular Pharmacology* **61**, 20-25.

Eadie, G.S. (1942). The inhibition of cholinesterase by physostigmine and prostigmine. *Journal of Biological Chemistry* **146**, 85-93.

Enyedy, I.J., Ling, Y., Nacro, K., Tomita, Y., Wu, X., Cao, Y., Cuo, R., Li, B., Zhu, X., Huang, Y., Roller, P.P., Yang, D. and Wang, S. (2001). Discovery of small molecule inhibitors of Bcl-2 through structure-based computer screening. *Journal Medicinal Chemistry* **44**, 4313-4324.

Erickson J.A., Jalaie, M., Robertson, D.H., Lewis, R.A. and Vieth, M. (2003). Lessons in molecular recognition, the effects of ligand and protein flexibility on molecular docking accuracy. *Journal Medicinal Chemistry* **47** (1), 45 -55.

Ferrara, P., Gohlke, H., Price, D.J., Klebe, G. and BrooksIII, C.L. (2004). Assessing scoring functions for protein-ligand interactions. *Journal Medicinal Chemistry* **47**, 3032-3047.

Fersht, A. (1993). Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. H.W. Freeman and Company, New York 103-131.

Fritz-Wolf, K., Becker, A., Rahlfs, S., Harwalt, P., Schirmer, R.H., Kansch, W. and Becker, K. (2003). X-ray structure of glutathione S-transferase from the malarial parasite *Plasmodium falciparum*. *PNAS* **24**, 13821-13826.

Gardner, M.J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R.W., Carlton, J.M.,

Pain, A., Nelson, K.E., Bowman, S., Paulsen, I.T., James, K., Eisen, J.A., Rutherford, K., Salzberg, S.L., Craig, A., Kyes, S., Chan, M.S., Nene, V., Shallom, S.J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M.W., Vaidya, A.B., Martin, D.M., Fairlamb, A.H., Fraunholz, M.J., Roos, D.S., Ralph, S.A., McFadden, G.I., Cummings, L.M., Subramanian, G.M., Mungall, C., Venter, J.C., Carucci, D.J., Hoffman, S.L., Newbold, C., Davis, R.W., Fraser, C.M., and Barrell, B. (2002). Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 531-534.

Ginsburg, H. and Kruglaik, M. (1992). Chloroquine, some open questions on its antimalarial mode of action and resistance. *Drug Resistance Updates* **2**, 180-187.

Gohlke, H., Hendlich, M. and Klebe, G. (2000a). Knowledge based scoring function to predict protein-ligand interactions. *Journal of Molecular Biology* **295**, 337-356.

Gohlke, H., Hendlich, M. and Klebe, G. (2000b). Predicting binding modes, binding affinities and "hot spots" for protein-ligand complexes using a knowledge-based scoring function. *Perspectives in Drug Discovery and Design* **20**, 115-144.

Gradler, U., Gerber, H.D., Goodenough-Lashau, D.M., Carcia, G.A., Ficner, R. and Reuter, K. (2001). A new target for shigellosis, rational design and crystallographic studies of inhibitors of tRNA-gaunine tranlycosylase. *Journal of Molecular Biology* **306**, 455-467.

Geldenhuis, W.J., Gaasch, K.E., Watson, M., Allen, D.D. and Van der Schyf, C.J. (2006). Optimizing the use of open-source software applications in drug discovery. *Drug Discovery Today* **11** (3), 127-132.

Giordanetto, F., Cotesta, S., Catana, C., Trosset, J.Y., Vulpetti, A., Stouten, P.W.F. and Kroemer, T.R. (2004). Novel scoring functions comprising QXP, SASA and protein side chain entropy terms. *Journal of Chemical Information and Computer Sciences* **44**, 882-893.

Harwaldt, P., Rahlfs, S. and Becker, K. (2002). Glutathione S-transferase of the malarial parasite *Plasmodium falciparum*, characterization of a potential drug target. *Biological Chemistry* **383** (5), 821-830.

Hay, S.I., Guerra, C.A., Tatem, A.J., Noor, A.M. and Snow, R.W. (2004). The global distribution and population at risk of malaria, past, present and future. *The Lancet Infectious Diseases* **4** (6), 327-336.

Hiller, N., Fritz-Wolf, K., Deponte, M., Wende, W., Zimmermann, H., and Becker, K. (2006). *Plasmodium falciparum* - Structural and mechanistic studies on ligand binding and enzyme inhibition. *Protein Science* **15**, 281-289.

Hillisch, A., Pineda, L.F. and Hilgenfeld, R. (2004). Utility of homology models in the drug discovery process. *Drug Discovery Today* **9** (15), 659-669.

Hofstee, B.H.J. (1959). Non-inverted versus inverted plots in enzyme kinetics. *Nature* **184**, 1296-1298.

Honma, T. (2003) Recent advances in *de novo* design strategy for practical lead identification. *Medicinal Research Reviews*. **23** (5), 606-632.

Hu, X., Balaz, S. and Shelver, W.H. (2004). A practical approach to docking of Zinc metalloproteinase inhibitors. *Journal of Molecular Graphics and Modeling* **22**, 293-307.

Jones, G., Willett, P., Glen, R.C., Leach, A.R. and Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *Journal of Molecular Biology* **267**, 727-748.

Joseph-McCarthy, D. (1999). Computational approaches to structure based ligand design. *Pharmacology and Therapeutics* **84**, 179-191.

Kelly, L.A., Gardner, S.P. and Sutcliffe, M.J. (1996). An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Engineering* **9** (11), 1063-1065.

Kirk, K. (2004). Channels and transporters as drug targets in the *Plasmodium*-infected erythrocyte. *Acta Tropica* **89** (3), 285-298.

Koehler, R.T., Villar, H.O., Bauer K.E. and Higgins D.L. (1997). Ligand based protein alignment and isozyme specificity of glutathione S-transferase inhibitors. *Proteins* **28**, 202-216.

Kramer, B., Rarey, M. and Lengauer, T. (1999) Evaluation of the FlexX incremental construction algorithm for protein-ligand docking. *Proteins* **37** (2), 228-241

Kremsner, P.G. and Krishna, S. (2004). Antimalarial combinations. *Lancet* **364** (9447), 1754-1755.

Krovat, E.M., Steindl, T. and Langer, T. (2005) Recent advances in docking and scoring. *Current Computer Aided Drug Design* **1**, 93-102.

Krumrine, J., Raubacher, F. and Brooijmans, N. (2003). Principles and methods of docking and ligand design. *Methods Biochemical Analysis* **44**, 443-476.

Kuderova, A., Nanak, E., Truksa, M. and Brzobohaty, B. (1999). Use of rifampicin in T7 RNA polymerase -driven expression of a plant enzyme, rifampicin improves yield and assembly. *Protein Expression and Purification* **16** (3), 405-409.

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

Lambert, J.H. (1760) *Photometria sive de mensura de gradibus luminis, colorum umbrae. Eberhard Klett.*

Lew, A. and Chamberlin, R.A. (1999). Blockers of human T cell Kv1.3 potassium channels using de novo ligand design and solid-phase parallel combinatorial chemistry. *Bioorganic and Medicinal Chemistry Letters* **9** (23), 3267-3272.

Li, C., Xu, L., Wolan, D.W., Wilson, I.A. and Olson, A.J. (2004). Virtual screening of human 5-Aminoimidazole - 4-carboxamide ribonucleotide transformylase against the NCI diversity set by use of AutoDock to identify novel nonfolate inhibitors. *Journal of Medicinal Chemistry* **47**, 6681-6690.

Liebau, E., Bergmann, B., Campbell, A.M., Teesdale-Spittle, P., Brophy, P.M., Luersen, K. and Walter, R.D. (2002). The glutathione S-transferase from *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* **124**, 85-90.

Lineweaver, H. and Burk, D. (1934). The determination of enzyme dissociation constants. *Journal of the American Chemical Society* **56**, 658-666.

Lowry, O.H., Rosebrough, N.J., Farr, A.L. and Randall, R.J. (1951). Protein measurement with Folin phenol reagent. *Journal of Biological Chemistry* **193** (1), 265-275.

Lyon, R.P., Hill, J.J. and Atkins, W.M. (2003). Novel class of bivalent glutathione S-transferase inhibitors. *Biochemistry* **42**, 10418-10428.

Macreadie, I., Ginsburg, H., Sirawaraporn, W. and Tilley, L. (2000). Antimalarial drug development and new targets. *Parasitology Today* **16** (10), 438-444.

Maeda, D.Y., Mahajan, S.S., Atkins, W.M. and Zebala, J.A. (2006). Bivalent inhibitors of glutathione S-transferase: The effect of spacer length on isozyme selectivity. *Bioorganic and Medicinal Chemistry Letters* **16** (14), 3780-3783.

Mahajan, S. and Atkins, W.M. (2005). The chemistry and biology of inhibitors and prodrugs targeted to glutathione S-transferases. *Cellular and Molecular Life Sciences* **62**, 1221-1233.

Majeux, N., Scarsi, M., Apostilakis, J., Ehrhardt, C. and Caffisch, A. (1999). Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins* **37**, 88-105.

Marsden, P.M., Puvamendrapillai, D., Mitchell, J.B.O. and Glen, R.C. (2004). Predicting protein-ligand affinities, a low scoring game? *Organic and Biomolecular Chemistry* **2**, 3267-3273.

May, J. and Meyer, C.G. (2003). Chemoresistance in *falciparum* malaria. *Trends in Parasitology* **19** (10), 432-435.

- McTigue, M.A., Williams, D.R. and Tainer, J.A. (1995). Crystal structures of a schistosomal drug and vaccine target, glutathione S-transferase from *Schistosoma japonica* and its complex with the leading antischistosomal drug praziquantel. *Journal of Molecular Biology* **246** (1), 21-27.
- Meshnick, S.R., (2002). Artemisinin, mechanisms of action, resistance and toxicity. *International Journal for Parasitology* **32** (13), 1655-1660.
- Michaelis, L. and Menten, M.L. (1913). Die kinetik der invertinwirkung. *Biochemistry* **49**, 333-369.
- Miller, L.H., Baruch, L.D., Marsh, K., Doumbo, O.K. (2002). The pathogenic basis of malaria. *Nature* **415**, 673-679.
- Mitchell, J.C., Kerr, R. and Ten Eyck, L.F. (2001). Rapid atomic density methods for molecular shape characterization. *Journal of Molecular Graphics and Modeling* **19**, 325-330.
- Mitchell, J.C., Shahbaz, S. and Ten Eyck, L.F. (2004). Interfaces in molecular docking. *Molecular Simulation* **30**, 97-106.
- Moore, S.A., Surgey, E.G. and Cadwgan, A.M. (2002). Malaria vaccines, where are we and where are we going? *Lancet Infectious Diseases* **2** (12), 737-743.
- Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry* **14**, 1639-1662.
- Muegge, I. and Martin, Y.C. (1999). A general and fast scoring function for protein-ligand interactions, a simplified potential approach. *Journal of Medicinal Chemistry* **42**, 791-804.
- Muller, S. (2004). Redox and antioxidant systems of the malaria parasite *Plasmodium falciparum*. *Molecular Microbiology*. **53** (5), 1291-305.
- Olliaro, P. (2001). Mode of action and mechanism of resistance for antimalarial drugs. *Pharmacology and Therapeutics* **89**, 207-219.
- Olsen, L., Jost, S., Adolph, H., Pettersson, I., Hemmingsen, L. and Jorgensen, F.S. (2006). New leads of metallo- β -lactamase inhibitors from structure-based pharmacophore design. *Bioorganic and Medicinal Chemistry* **14**, 2627-2635.
- Palmer, T. (2002). Enzymes. Biochemistry, biotechnology, clinical chemistry. Horwood Publishing Limited.
- Pearlman, R.S. and Smith, K.M. (1998). Novel software tools for chemical diversity. *Perspectives in Drug Design* **9**(11), 339-353.

Perbandt, M., Burmeister, C., Walter, R.D., Betzel, C. and Liebau, E. (2003). Native and inhibited structure of a Mu class related glutathione S- transferase from *Plasmodium falciparum*. *The Journal of Biological Chemistry* **2**, 1336-1342.

Perola, E., Walters, W.P. and Charifson, P.S. (2004). A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* **56**, 235-249.

Platel, D.F.N., Mangou, F. and Tribouley-Duret, J. (1999). Role of glutathione in the detoxification of ferriprotoporphyrin IX in chloroquine resistant *Plasmodium berghei*. *Molecular and Biochemical Parasitology* **98**, 215-223.

Rahlfs, S., Schirmer, R.H. and Becker, K. (2002). The thioredoxin system of *Plasmodium falciparum* and other parasites. *Cellular Molecular Life Science* **59** (6), 1024-1041.

Ramage, P., Hemming, R., Mathis, B., Cowan-Jacobs, S.W., Rondeau, J.M., Kallen, J., Blommers, M.J.J., Zurini, M. and Rudisser, S. (2002). Snags with tags, Some observations made with (His)₆-tagged proteins. *Life Science News (Amersham Biosciences)* **11**, 1-5.

Rarey, M., Kramer, B., Lengauer, T. and Klebe, G. (1996a). A fast flexible docking method using an incremental construction algorithm. *Journal of Molecular Biology* **261**, 470-489.

Rarey, M., Wefing, S., Lengauer, T. (1996b). Placement of medium-sized molecular fragments into active sites of proteins. *Journal of Computer Aided Molecular Design* **10**, 41-54.

Rarey, M., Kramer, B., Lengauer, T. (1999). Docking of hydrophobic ligands with interaction based matching algorithms. *Bioinformatics* **15** (3), 243-250.

Riganese, G.M., Cardoso, R.F.M., Daniels, D.S., Bruns, C.M. and Tainer, J.A. (2003). Characterizations of the electrophile binding site and substrate binding mode of 26kDa glutathione S-transferase from *Schistosoma japonicum*. *Proteins* **5**, 137-146.

Rishton, G.M. (2003) Nonleadlikeness and leadlikeness in biochemical screening. *Drug Discovery Today* **8**(2), 86-96.

Rosenthal, P.J. (2003) Antimalarial drug discovery, old and new approaches. *The Journal of Experimental Biology* **206**, 3735-3744.

Schneider, G. and Bohm, H.J. (2002). Virtual screening and fast automated docking methods. *Drug Development Today* **7**, 64-70.

Schneider, G. and Fechner, U. (2005). Computer based *de novo* design of drug like molecules. *Nature Reviews. Drug Discovery* **4**, 649-663.

Schneidman-Duhovny, D., Nussinov, R. and Wolfson, H.J. (2004). Predicting molecular

interactions *in silico*, protein-protein and protein-drug docking. *Current Medicinal Chemistry* **11**, 97-110.

Sheenan, D., Meade, G., Foley, V.M. and Dowd, C.A. (2001). Structure function and evolution of glutathione transferases, implication for classification of non-mamalian members of an ancient super family. *Biochemistry Journal* **360**, 1-16.

Srivastva, P., Puri, S.K., Kamboj, K.K. and Pandey, V.C. (1999). Glutathione S-transferase activity in malaria parasites. *Tropical Medicine and International Health* **4**, 251-254.

Staines, H.M., Ellory, J.C. and Chibale, K. (2005). The new permeability pathways, targets and selective routes for the development of new antimalarials. *Combinatorial Chemistry and High Throughput Screening* **8**, 81-88.

Stahl, M. and Mauser, H. (2005). Database clustering with a combination of fingerprints and maximum common substructure methods. *Journal of Chemical Information and Modeling* **45**, 542-548

Stoichet, B.K., Leach, A.R. and Kuntz, I.D. (1999). Ligand solvation in molecular docking. *Proteins* **34**, 4-16.

Stoichet, B.K., Stroud, R.M., Santi, D.V., Kuntz, I.D. and Perry, K.M. (1993). Structure-based discovery of inhibitors of thymidylate synthase. *Science* **259** (5100), 1445-1450.

Strange, R.C., Jones, P.W. and Fryer, A.A. (2000). Glutathione S-transferase, genetics and role in toxicology. *Toxicology Letters* **15**, 357-363.

Tamta, H. and Mukhopadhyay A.K. (2003). Biochemical targets for malaria chemotherapy. *CRIPS* **4** (1), 6-9.

Terstappen, G.C. and Reggiani, A. (2001). In silico research in drug discovery. *Trends in Pharmacological Science* **4**(1), 23-26

Tschinke, V. and Cohen, N.C. (1993). The NEWLEAD program, a new method for the design of candidate structures from pharmacophoric hypotheses. *Journal of Medicinal Chemistry* **36** (24), 3863-3870.

Wang, R., Lui, L., Lai, L. and Tang, Y. (1998). SCORE, A new empirical method for estimating the binding affinity of a protein ligand complex. *Journal Molecular Modeling* **4**, 379-394.

Wang, R., Lai, L. and Wang, S. (2002). Further development and validation of empirical scoring functions for structure-based binding affinity predictions. *Journal of Computer Aided Molecular Design* **16**, 11-26.

Wang, R., Lu, Y. and Wang, S. (2003). Comparative evaluation of 11 scoring functions for molecular docking. *Journal Medicinal Chemistry* **46**, 2287-2303.

Ward, J.H. (1963). Hierarchical grouping to optimize an objective function. *Journal of American Statistical Association* **58**, 236-244.

Watanabe K., and Harayama S. (2001). SWISS-PROT: The curated protein sequence database on Internet. *Protein, Nucleic Acid and Enzyme* **46**, 80-86.

White, N.J. (1997). Assessment of the pharmacodynamic properties of antimalarial drugs in vivo. *Antimicrobial Agents Chemotherapy* **41** (7), 1413-1422.

White, N.J. (2004). Antimalarial drug resistance. *The Journal of Clinical Investigation* **133** (8), 1084-1092.

Winstanley, P.A. (2000). Chemotherapy for *falciparum* malaria , the armory, the problems and the prospects. *Parasitology Today* **16**, 146-153.

Wolf, K. and Dormeyer, M. (2003). Information-based methods in the development of antiparasitic drugs. *Parasitology Research* **90**, 91-96.

Wongsrichanalai, C., Pickard, A.L., Wernsdorfer, W.H. and Meshnick, S.R. (2002). Epidemiology of drug-resistant malaria. *Lancet Infectious Diseases* **2** (4), 209-218.

Yew, I. and Altman, R.B. (2006). Drug targets for *Plasmodium falciparum*, a post-genomic review/survey. *Mini-Reviews in Medicinal Chemistry* **6**, 177-202.

Zanders, E.D., Bailey, D.S. and Dean, P.M. (2002). Probes for chemical genomics by design. *Drug Discovery Today* **7**, 711-717.

Zang, C., Liu, S., Zhu, Q. and Zhou, Y. (2005). A knowledge-based energy function for protein-ligand, protein-protein and protein-DNA complexes. *Journal of Medicinal Chemistry* **48**, 2325-2335.