

# **Structural modelling of therapeutic targets and inhibitors of the malaria parasite**

by

Fourie Joubert

Submitted in partial fulfilment of the degree

Ph.D.

in the Department of Biochemistry

University of Pretoria

July 2000

"Molecular docking is a serious challenge for bio-chemists. There are many ways to fit molecules together but only a few juxtapositions that bring them close enough to bond. On a molecular level success may mean discovering what synthetic structure, what chemical, will form a union with, say, the protein shape on a tumor cell. If you make this high-risk jigsaw work you may have found a cure for carcinoma. But molecules and the human beings they are part of exist in a universe of possibility. We touch one another, bond and break, drift away on force-fields we don't understand."

A quotation from the novel *'Written on the Body'* by the British writer Jeanette Winterson (1992).

# Acknowledgements

I would like to express my sincere gratitude to:

- Prof. Braam Louw, my supervisor, who always supported and motivated me, even when things were looking hopeless. Without his ideas and inputs, this project could not have succeeded. He has truly become part of my family.
- Prof. Albert Neitz, co-supervisor and Head of the Department of Biochemistry. He strongly supported all my new ventures into Bioinformatics and Structural modelling with his usual good humour and patience, and has led the whole Department to new heights.
- Prof. Robin Crewe, who was the first to give us financial and moral support to venture into new research directions.
- All my friends and colleagues in the Biochemistry, Genetics and Microbiology Departments for their support, assistance and good humour.
- Ben, Deborah, Karen, Katherine and Lyn-Marie for support, input and enthusiasm through the years.
- My brother Louis, his wife Anita, brother-in-law Michael, sister-in-law Clarel and parents-in-law Michael and Susan for patience, love, support and understanding.
- My wife Annie, who has loved and carried me through the toughest of times. Without her, I would probably not be alive today.
- My Father in Heaven, for giving me the will, strength and good health to complete this thesis.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Malaria	1
1.2	Metabolic pathways	5
1.2.1	Glycolysis	5
1.2.2	Purine salvage pathway	7
1.2.3	Pyrimidine biosynthesis	8
1.3	The target enzymes	9
1.3.1	Dihydrofolate reductase	9
1.3.2	Triosephosphate isomerase	21
1.4	The approach	29
<b>2</b>	<b>Expression of putative malaria drug target proteins</b>	<b>30</b>
2.1	Introduction	30
2.2	Materials and Methods	34
2.2.1	Expression of dihydrofolate reductase	34
2.2.1.1	The pET17 system	34
2.2.1.2	The pTrxFus system	35
2.2.1.3	The pET32 system	37
2.2.2	Expression of triosephosphate isomerase	38
2.2.2.1	The pET15b system	38
2.2.2.2	Analysis of recombinant TIM	39
2.3	Results	41

2.3.1	Expression of dihydrofolate reductase . . . . .	41
2.3.2	Expression of triosephosphate isomerase . . . . .	43
2.4	Discussion . . . . .	50
<b>3</b>	<b>Homology modelling of putative malaria drug target proteins</b>	<b>53</b>
3.1	Introduction . . . . .	53
3.2	Methods . . . . .	57
3.2.1	Homology modelling of dihydrofolate reductase and triosephosphate isomerase . . . . .	57
3.3	Results . . . . .	58
3.3.1	Homology modelling of dihydrofolate reductase . . .	58
3.3.2	Homology modelling of triosephosphate isomerase .	67
3.4	Discussion . . . . .	74
<b>4</b>	<b>Ligand discovery for malaria triosephosphate isomerase</b>	<b>77</b>
4.1	Introduction . . . . .	77
4.2	Methods . . . . .	81
4.3	Results . . . . .	82
4.4	Discussion . . . . .	92
<b>5</b>	<b>Concluding Discussion</b>	<b>98</b>

# List of Figures

1.1 Summarised geographical map of recent surveys in malaria drug resistance (Centers for Disease Control and Prevention, Atlanta, 1999). . . . .	4
1.2 The glycolytic pathway . . . . .	6
1.3 The postulated purine salvage pathway in <i>P. falciparum</i> . Red indicates the red blood cell pathway and brown that of the parasite. . . . .	8
1.4 <i>De novo</i> pyrimidine biosynthesis in <i>P. falciparum</i> . . . . .	9
1.5 Structure of folic acid moieties. . . . .	10
1.6 Reduction of folic acid. . . . .	10
1.7 Human DHFR (PDB ID: 1DHF) bound to folate. <i>Trp24</i> and the conserved loop is indicated in red, <i>Phe31</i> in yellow, <i>Phe34</i> in green and <i>Arg70</i> in cyan. Folate is shown in CPK atom colors. . . . .	12
1.8 <i>E. coli</i> DHFR bound to folate (PDB ID: 1DYI). <i>Asp27</i> is indicated in red, <i>Thr46</i> in green, <i>Ser49</i> in purple, <i>Val88</i> in cyan and the loop formed by residues 16-20 in blue. . . . .	14
1.9 Avian DHFR bound to biopterin (PDB ID: 1DR1). <i>Glu30</i> is shown in red, <i>Leu22</i> in cyan, <i>Ile7</i> in green, <i>Val8</i> in magenta, <i>Tyr31</i> in yellow and <i>Val115</i> in blue. . . . .	17
1.10 Dimeric human TIM with the active sites of both subunits indicated in green (PDB ID: 1HTI). PGA is shown in red in the active site of one subunit. . . . .	22

1.11 An overview of the human TIM active site region (PDB ID: 1HTI). PGA is shown in CPK colours. <i>Glu165</i> is shown in red, <i>His95</i> in green, <i>Lys12</i> in blue, <i>Asn10</i> in purple and <i>Glu97</i> in cyan. <i>Thr75</i> and its chain from the second subunit are shown in yellow. . . . .	23
1.12 The active site region of yeast TIM (PDB ID: 2YPI). <i>His95</i> is indicated in red, <i>Glu165</i> in green, <i>Lys12</i> in cyan, <i>Glu97</i> in blue and residues 71-77 in magenta. PGA is indicated in CPK colours. . . . .	24
1.13 The active site of avian TIM (PDB ID: 1TPH). <i>His95</i> is indicated in red, <i>Glu165</i> in green and residues 166-176 in magenta. PGH is shown in CPK colours. . . . .	26
1.14 Structure of <i>Trypanosoma</i> TIM (PDB ID: 4TIM). <i>Lys13</i> is indicated in cyan, and <i>His95</i> in yellow. Loop 1 is colored blue, loop 2 green, loop 3 red and loop 4 magenta. . . . .	27
2.1 SDS-PAGE analysis of pET17-DHFR expression. Lane 1 contained molecular mass markers, lane 2 was a negative control BL21(DE3) sample and lane 3 was an extract of BL21(DE3) expressing malaria DHFR. . . . .	42
2.2 Activity assay of crude malaria DHFR. The negative control (water) is indicated by a dotted line, the positive control (bovine DHFR) by a dashed line and malaria DHFR by a solid line. . . . .	43
2.3 Expression of the recombinant malaria DHFR fusion protein with pET32 in BL21(DE3). Lane 1 contained molecular mass markers, lanes 2&3 both contained control soluble phases, lanes 4&5 contained soluble phases from pET32-DHFR, lanes 6&7 contained control insoluble phases and lanes 8&9 contained insoluble phases from pET32-DHFR. . . . .	44
2.4 Electrophoresis of PCR products for malaria TIM. Lane 1 contained molecular mass markers, lanes 2&4 contained 1.5 $\mu$ l of cDNA template and lanes 3&5 contained 0.5 $\mu$ l of cDNA template. The PCR of samples in lanes 4&5 was performed for 35 cycles and lanes 2&3 for 25 cycles. . . . .	44
2.5 A part of the electrophoretogram for the TIM sequence. . . . .	45

2.6	Expression of recombinant TIM at 30°C. Lane 1 contained molecular mass markers, lanes 2-7 contained soluble fractions and lanes 8-13 contained insoluble fractions. Lanes 3, 5, 7, 9, 11 and 13 contain IPTG-induced samples, and lanes 2, 4, 6, 8, 10 and 12 were uninduced. Lanes 2, 3, 8 and 9 were at 0 hours post-induction, lanes 4, 5, 10 and 11 were at 8 hours and lanes 6, 7, 12 and 13 were at 16 hours. Lane 14 contained IMAC-purified recombinant malaria TIM. . . . .	45
2.7	His-tag purification of recombinant TIM. The eluted recombinant protein is indicated by black shading. . . . .	46
2.8	A model of malaria TIM with the subunits indicated in red and blue, the active sites in CPK-coloured spheres and the oligo-histidine tag in green. . . . .	47
2.9	MALDI analysis of purified recombinant TIM. The major peak of 30,258Da corresponds to the sum of the m/z monomer peak plus the 1/2m/z peak of the remaining dimer. Some dimer is still visible at m/z=60.606Da. . . . .	48
2.10	Inverse reciprocal plot for recombinant malaria TIM. $K_m$ was determined as 0.586mM and $V_{max}$ as 0.027 $\mu$ mole/min from assays performed in triplicate. . . . .	48
2.11	A pH optimum plot for recombinant TIM. The optimum was determined as approximately 8.5. . . . .	49
2.12	Temperature stability plot for recombinant TIM. The enzyme activity was stable to a temperature of 55°C, after which a sudden decrease in activity occurred. . . . .	49
3.1	Amino acid sequence alignment of malaria DHFR with other species DHFRs from the Brookhaven Protein Data Bank. Homology is indicated in blue and identity in red. Conserved active site residues are boxed. . . . .	58
3.2	Sequence alignment of modified malaria DHFR with <i>E. coli</i> and <i>L. casei</i> DHFR. Homology is indicated in blue and identity in red. Important catalytic residues are indicated by boxes. . . . .	59

3.3	Modified sequence alignment of truncated malaria DHFR with <i>E. coli</i> and <i>L. casei</i> DHFR after removal of large insertions from the malaria sequence. Homology is indicated in blue and identity in red. . . . .	60
3.4	Ramachandran plot for the homology model of malaria DHFR. . . . .	61
3.5	Main chain parameter analysis for homology modelled malaria DHFR. . . . .	62
3.6	Side-chain parameter analysis for homology modelled malaria DHFR. . . . .	63
3.7	Quality score indications of homology-modelled DHFR. Red regions indicate lower quality, green intermediate and blue regions indicate higher quality. . . . .	64
3.8	Fitted structures of modelled malaria DHFR (red) and human DHFR (green). . . . .	65
3.9	Superimposed active site residues of malaria DHFR (red) and human DHFR (green). . . . .	66
3.10	Sequence alignments of all TIM structures available in PDB. Homology is indicated in blue and identity in red. . .	67
3.11	Ramachandran plot for the homology model of malaria TIM. . . . .	68
3.12	Main chain parameter analysis for homology modelled malaria TIM. . . . .	69
3.13	Side-chain parameter analysis for homology modelled malaria TIM. . . . .	70
3.14	Quality score indications of homology-modelled TIM. Red regions indicate lower quality, green intermediate and blue regions indicate higher quality. . . . .	71
3.15	Fitted structures of modelled malaria TIM (red) and human TIM (green). . . . .	72
3.16	Superimposed active sites of malaria TIM (red) and human TIM (green). . . . .	73
3.17	Superimposed C $\alpha$ backbones of homology modelled TIM (red) and the X-ray structure of TIM (green). A R.M.S. deviation for carbon- $\alpha$ of 1.5Å was found. . . . .	75

3.18	Superimposed active site residues of homology modelled TIM (red) and the X-ray structure of TIM (green). The only major difference was the rotation angle of Phe96. . . . .	76
4.1	Region of malaria TIM surrounding the active site used for ligand docking. The active site was determined by the binding position of the inhibitor PGA. Yellow spheres indicate the active site surface, and the region of interest is bound by a red box. Active site residue <i>Lys12</i> is indicated in green, <i>His96</i> in magenta and <i>Glu165</i> in blue. . . . .	82
4.2	Detailed view of spheres defining the inverse of the active site cavity. . . . .	83
4.3	Inhibitor kinetic plot for Direct Red and malaria TIM. Squares indicate a concentration of 60 $\mu$ M, triangles 50 $\mu$ M and circles 40 $\mu$ M. . . . .	86
4.4	Parasite growth in the presence of Direct Red and Direct Violet. Red bars indicate red blood cell growth and cyan bars indicate parasite growth. . . . .	86
4.5	Direct Red and Violet bound in the active site of <i>P. falciparum</i> TIM. An accessibility surface is shown as mesh (top) or as solid (bottom). . . . .	88
4.6	Detailed contact map of the malaria TIM-Direct Red complex. . . . .	89
4.7	Detailed contact map of the malaria TIM-Direct Violet complex. . . . .	90
4.8	Comparison of the spatial positions of bound Direct Red and Violet. The common structural area is indicated by a white box. <i>Phe96</i> , <i>Glu165</i> and <i>Val212</i> show common contact points on both molecules. The top figure shows the exact orientation, while the bottom figure has been shifted slightly to distinguish the two molecules. . . . .	91
4.9	Structures of Suramin (top), and Direct Red (bottom). . . . .	93
4.10	A comparison of the three-dimensional structures of Suramin (yellow), Direct Red (red) and Direct Violet (magenta). . . . .	94
4.11	Detailed contact map for the TIM-Suramin complex. . . . .	95

4.12 Active site comparison for malaria TIM (red), <i>Trypanosoma</i> TIM (cyan) and human TIM (white). Direct Red as docked into the malaria TIM active site is shown in yellow. . . . .	96
5.1 Structures of Suramin and related compounds. Percentage inhibition of <i>Trypanosoma</i> TIM reactivation at 10 $\mu$ M is indicated. Similar structural motifs are shaded in grey (Gao <i>et al.</i> , 1998). . . . .	102

# List of Tables

2.1	Some malaria enzymes which have been successfully expressed in recombinant systems. . . . .	32
2.2	Properties of TIM from various species. . . . .	48
4.1	Commercially available compounds tested for TIM inhibition. . . . .	84
4.2	Synthesized compounds tested for TIM inhibition. . . . .	85

# List of Abbreviations

3D	Three dimensional
Å	Angstrom
A	Adenine
ADP	Adenosine diphosphate
Ala	Alanine
ALD	Fructose 1,6-bisphosphate aldolase
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
ATP	Adenosine triphosphate
bp	Base pairs
C	Cytidine
cDNA	Complementary DNA
CPK	Corey-Pauling-Koltun
CTP	Cytosine triphosphate
Cys	Cysteine
D-GAP	D-glyceraldehyde phosphate
DHAP	Dihydroxyactone phosphate
DHF	Dihydrofolate
DHFR	Dihydrofolate reductase

DHFR-TS	Dihydrofolate reductase-thymidylate synthase
DHODH	Dihydroorotase dehydrogenase
DNA	deoxy-Ribonucleic acid
dTMP	deoxy-Thymidine monophosphate
dUDP	deoxy-Uridine diphosphate
dUMP	deoxy-Uridine monophosphate
dUTP	deoxy-Uridine triphosphate
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylenediaminetetra-acetate
FDA	Food and Drug Administration
G	Guanine
GAP	Glyceraldehyde phosphate
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GDH	Glucose 3-phosphate dehydrogenase
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
GMP	Guanosine monophosphate
GNDA	Gossyclic nitrile diacetate
GPI	Phosphoglucoisomerase
His	Histidine
HGPRT	Hypoxanthine-guanine phosphoribosyltransferase
Ile	Isoleucine
IMAC	Immobilised metal affinity chromatography
IMP	Inosine monophosphate
IPTG	Isopropyl- $\beta$ -d-thiogalactopyranoside

$k_{cat}$	Catalytic constant
$K_I$	Inhibition constant
$K_M$	Michaelis-Menten constant
LB-broth	Luria-Bertani broth
<i>L. casei</i>	<i>Lactobacillus casei</i>
LDH	Lactate dehydrogenase
Leu	Leucine
Lys	Lysine
MALDI	Matrix assisted laser desorption ionisation
mdr	Multi-drug resistance
MePhSO <sub>2</sub> -Ph	Phenyl methanethiosulfonate
Met	Methionine
mTIM	Monomeric triosephosphate isomerase
NAD	Nicotinamide adeninedinucleotide
NADH	Nicotinamide adeninedinucleotide (reduced)
NCI	National Cancer Institute
NMR	Nuclear magnetic resonance
OD	Optical density
PABA	Para-amino butyric acid
PBS	Phosphate-buffered saline
<i>P. carinii</i>	<i>Pneumocystis carinii</i>
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PGA	Phosphoglycolic acid
PGH	Phosphoglycolohydroxamate
<i>P. falciparum</i>	<i>Plasmodium falciparum</i>

PGK	Phosphoglycerate kinase
Phe	Phenylalanine
PMSF	Phenylmethylsulphonylfluoride
Pro	Proline
<i>P. vivax</i>	<i>Plasmodium vivax</i>
RFLP	Restriction fragment length polymorphism
R.M.S.	Root mean square
RNA	Ribonucleic acid
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SDS	Sodiumdodecyl sulphate
SDS-PAGE	Sodiumdodecyl sulphate polyacrylamide gel electrophoresis
Ser	Serine
<i>S. pombe</i>	<i>Saccharomyces pombe</i>
S-Tag	Ribonuclease-S binding peptide tag
T	Thymine
<i>T. brucei</i>	<i>Trypanosoma brucei</i>
TIM	Triosephosphate isomerase
TIM-PGH	Triosephosphate isomerase-phosphoglycolohydroxamate complex
tRNA	Transfer RNA
Trp	Tryptophan
TS	Thymidylate synthase
Tyr	Tyrosine
UDP	Uridine diphosphate
UTP	Uridine triphosphate
Val	Valine
$V_{max}$	Maximum velocity
XMP	Xanthine monophosphate

## Summary

Anti-malarial drug resistance is increasing on a global level. With the progresses made in computerised drug design, developments in the field of therapeutic agents against malaria may be greatly accelerated in the near future. In this study, two malaria metabolic enzymes were identified as potential targets for computer-aided drug design. The enzymes dihydrofolate reductase (DHFR) and triosephosphate isomerase (TIM) were cloned, and expressed in bacterial expression systems. Large amounts of DHFR could be expressed, but not in soluble form. It was shown however, that native malaria genes could be expressed in *E. coli* under certain promoters. TIM was expressed in an active form and purified to homogeneity by means of immobilized metal affinity chromatography (IMAC).

The structures of both DHFR and TIM were modelled by homology based methods and analysed in terms of quality. DHFR was extremely complicated to model due to the presence of several large insertions, and did not yield a high quality model. The model of TIM was of adequate quality to be used for ligand studies.

When the X-ray structure for malaria TIM became available recently, it was compared to the homology model and was found to be virtually the same. The only important difference was the rotation of the *Phe96* side-chain at the active site. The X-ray structure was subsequently used for ligand screening against the NCI-3D small molecule database. Two of the 7 *in vitro* tested compounds yielded inhibition in the  $<100\mu\text{M}$  range. These two molecules Direct Red and Direct Violet are derivatives of the anti-trypanosomal Suramin which was originally derived from the dye Trypan Red. According to the literature, Direct Red and Direct Violet inhibit trypanosomal TIM approximately 100x more potently than Suramin.

A common structural region was identified in Direct Red, Direct Violet

and a series of other Suramin-related compounds, and this moiety may be used as a basis for the design of anti-TIM lead drug compounds.

## Opsomming

Weerstandbiedenheid teen anti-malaria middels neem toe op 'n globale vlak. Met die vordering in rekenaargebaseerde ontwerp sal die ontwikkeling van terapeutiese agente teen malaria binnekort vinnig toeneem. In hierdie studie is twee metaboliese ensieme van die malaria parasiet geïdentifiseer as potensiële teikens vir rekenaargebaseerde ontwerp van geneesmiddels. Die ensieme dihidrofolaat reduktase (DHFR) en triosefosfaatisomerase (TIM) is gekloneer en uitgedruk in bakteriële uitdrukkingstelsels. Groot hoeveelhede DHFR kon uitgedruk word, maar nie in 'n oplosbare vorm nie. Daar is egter gewys dat natiewe malaria gene in *E. coli* uitgedruk kon word onder die beheer van spesifieke promoters. TIM is uitgedruk in 'n aktiewe vorm en daarna gesuiwer tot homogeniteit deur middel van geïmmobiliseerde metaal affiniteitschromatografie (IMAC).

Die strukture van beide DHFR en TIM is gemodelleer deur middel van homologie metodes en geanaliseer in terme van kwaliteit. DHFR was besonder ingewikkeld om te modelleer as gevolg van die teenwoordigheid van verskeie groot ingevoegde segmente, en het nie 'n model van hoë kwaliteit gelewer nie. Die model van TIM was van voldoende kwaliteit om vir ligandbindingstudies gebruik te word.

Aangesien die X-straalstruktuur van malaria TIM onlangs beskikbaar geword het, is dit vergelyk met die homologie model en min verskille is gevind. Die enigste belangrike verskil was die rotasiehoek van die Phe96 syketting by die aktiewe sentrum. Die X-straalstruktuur is daarna gebruik vir ligandsoektogte teen die NCI-3D kleinmolekuul databasis. Twee van die 7 *in vitro* getoetste verbindings het inhibisie in die  $<100\mu\text{M}$  konsentrasiegebied gewys. Die twee verbindings, Direct Red en Direct Violet is derivate van die anti-trypanosomale middel Suramin, wat oorspronklik op die kleurstof Trypan Rooi gebaseer is. In die literatuur is gevind dat Direct Red en Direct Violet trypanosomale

TIM ongeveer 100x sterker inhibeer as Suramin.

'n Gedeelte strukturele motief is geïdentifiseer in Direct Red, Direct Violet en 'n reeks ander Suramin-verwante verbindings. Die motief mag gebruik word as basis vir die ontwerp van anti-TIM voorloper geneesmiddels.