Structural studies of enoyl acyl carrier protein reductase from *Plasmodium falciparum* and *Toxoplasma gondii*

A thesis submitted in part fulfilment of the requirements for the degree of Doctor of Philosophy

By

Stephen Muench

B.Sc. (Hons) Biochemistry/Microbiology, The University of Sheffield

Department of Molecular Biology and Biotechnology

The University of Sheffield

(September 2004)

Summary

Structural studies of enoyl acyl carrier protein reductase from *Plasmodium falciparum* and *Toxoplasma gondii*

Stephen Muench
Ph.D. Thesis

Enoyl acyl carrier protein reductase enzyme (ENR) catalyses one of the two reductive steps in fatty acid elongation within the fatty acid synthesis type II cycle that is common to plants and prokaryotes. Since enzymes of this pathway are absent in humans they have become the target for several potent antibacterial compounds including triclosan which inhibits ENR in the picomolar range.

As part of this thesis the gene for a type II ENR was located in the genomes of the apicomplexan parasites *Plasmodium falciparum* and *Toxoplasma gondii*. Analysis of the derived protein sequences suggested that these enzyme reside in the apicoplast. X-ray crystallographic techniques have been used to solve the structure for *Plasmodium falciparum* (Pf) and *Toxoplasma gondii* (Tg) ENR in complex with the NAD^+ cofactor and triclosan by molecular replacement to 2.2Å and 2.6Å, respectively. Both enzymes are tetrameric with the approximate dimensions of 90Å x 90Å x 50Å. Each subunit is formed by a 7 stranded parallel β-sheet flanked by 9α helices, reminiscent of a Rossmann nucleotide binding fold common to several NAD^+ binding enzymes. Analysis of the ENR family reveals that a characteristic of apicomplexan ENRs is an insert which varies in size from 42 residues in the *P.falciparum* enzyme to 6 residues in *T.gondii* ENR and which flanks the inhibitor/substrate binding site. In PfENR this loop is disordered but in the structure of TgENR the loop can be clearly seen and the structure shows that the loop lies close to the bound inhibitor but makes no direct contacts.

Comparisons of the binding sites of a range of different ENR inhibitor complexes has led to a better understanding of the plasticity of the enzyme in response to inhibitor (and possibly substrate) binding. Moreover analysis of the substrate/inhibitor binding pocket in *P.falciparum* and *T.gondii* ENR shows that whilst they are similar to the bacterial enzymes there are distinct differences which could be exploited for the development of novel antiparasitic agents. A major hurdle in the delivery of inhibitors targeted towards the apicoplast organelle is the need to cross several barriers including the parasite membranes and host cell walls. However the addition of a releasable eight arginine linker to the phenolic OH group of triclosan significantly improved the speed of delivery and enabled triclosan to enter both the extracellular and intracellular *T.gondii* tachyzoites and bradyzoites. The identification of both a novel inhibitor for the apicomplexan family and a possible general delivery mechanism may provide a foundation for the development of ENR inhibitors that will efficiently treat several key parasitic diseases.

Acknowledgements

Firstly I would like to thank my supervisor, Professor David Rice for offering me invaluable advice and support over the last few years. I would also like to thank Dr John Rafferty and Dr Patrick Baker for there help and advice whenever asked which has been a lot throughout the course of this project. I would also like to thank all members of the crystallography group for their help over the last four years.

I would also like to thank Professor Rima Mcleod and Professor Sean Prigge and everyone involved in their laboratories for their support and collaborative work on the overexpression and characterisation of the *P.falciparum* and *T.gondii* ENR without which this project would not have been possible.

On a personal note a huge thanks to my family especially for mum and dad whose continuing support has been invaluable of the seven years I have been at University and don't worry I will get a job soon. Thanks also to all my friends at Sheffield whose joint appreciation for fermented fruit drinks has helped me through many a rainy Sheffield day. Finally, I would like to thank Nicky for your continuing love and moral support.

Abbreviations and Symbols

Crystallographic Terms

Å	Ångstrom
a, b, c	Real space unit cell dimensions
a*,b*,c*	Reciprocal space unit cell dimensions
α, β, γ	Real space unit cell angles
$\alpha^*, \beta^*, \gamma^*$	Reciprocal space unit cell angles
B	Crystallographic temperature factor
Cc	Correlation Coefficient
CCD	Charged coupled device
ESRF	European Synchrotron Radiation Facility
f	Atomic scattering factor
F(S)	Structure factor
$\mathbf{F_{calc}}$	Calculated structure factor
$\mathbf{\lvert F_{calc} \rvert}$	Calculated structure factor amplitude
$\mathbf{F_{obs}}$	Observed structure factor
$\mathbf{\lvert F_{obs} \rvert}$	Observed structure factor amplitude
$\mathbf{F(}\boldsymbol{hkl}\mathbf{)}$	Structure factor for reflection *hkl*
$\mathbf{\lvert F(}\boldsymbol{hkl}\mathbf{) \rvert}$	Structure factor amplitude for reflection *hkl*
h, k, l	Reciprocal lattice indices
I	Diffraction intensity
λ	Wavelength
MR	Molecular Replacement
NCS	Non-Crystallographic Symmetry
RF	Rotation function
R_{factor}	Crystallographic refinement R-factor
R_{free}	Free R-factor
R_{merge}	R-factor of agreement between symmetry related terms
ρ	Electron density
$\mathbf{P(}\boldsymbol{uvw}\mathbf{)}$	Patterson function
r	Position of atom
S	Scattering vector
s	Vector describing direction of scattered beam

$\mathbf{s}_0$	Vector describing direction of incident beam
SRS	Synchrotron radiation source
TF	Translation function
TLS	Translation, Libration and Screw-rotation displacements
u, v, w	Patterson space coordinates
V	Volume of unit cell
V_m	Mathews number
x, y, z	Real space coordinates
ω, ϕ, κ	Polar angles

Chemical terms

DMSO	Dimethyl Sulfoxide
NAD^+	Nicotinamide adenine dinucleotide- oxidised form
NADH	Nicotinamide adenine dinucleotide- reduced form
NADPH	Nicotinamide adenine dinucleotide phosphate- reduced form
PEG	Polyethylene glycol
SDS	Sodium Dodecyl Sulphate
Tric-HCL	Tris(Hydroxymethyl)methylamine-HCL

Enzyme Abbreviations

ACCase	Acetyl-CoA Carboxylase
ACP	Acyl carrier protein
BKR	β-Keto ACP reductase
BnENR	*Brassica napus* Enoyl ACP reductase
EcENR	*Escherichia coli* Enoyl ACP reductase
ENR	Enoyl ACP reductase
MtENR	*Mycobacterium tuberculosis* Enoyl ACP reductase
PfENR	*Plasmodium falciparum* Enoyl ACP reductase
SDR	Short chain dehydrogenase, oxidoreductase and epimerase
TgENR	*Toxoplasma gondii* Enoyl ACP reductase

Species Abbreviations

A. thialana	*Arabidopsis thialana*
B. napus	*Brassica napus*

B. subtilis	*Bacillus subtilis*
C. testosteroni	*Comamonas testosteroni*
E. coli	*Escherichia coli*
H. pylori	*Helicobacter pylori*
M. tuberculosis	*Mycobacterium tuberculosis*
T. gondii	*Toxoplasma gondii*
P. berghei	*Plasmodium berghei*
P. chabaudi	*Plasmodium chabaudi*
P. knowlesi	*Plasmodium knowlesi*
P. falciparum	*Plasmodium falciparum*
P. reichenowi	*Plasmodium reichenowi*
P. vivax	*Plasmodium vivax*
P. yoelii	*Plasmodium yoelii*

Miscellaneous

Da	Dalton
FabB	β-Ketoacyl ACP synthase I
FabF	β-Ketoacyl ACP synthase II
FabG	β-Ketoacyl ACP reductase
FabH	β-Ketoacyl ACP synthase III
FabI	Enoyl ACP reductase
FabZ	β-Hydroxy ACP dehydratase
FAS	Fatty acid synthetase
IC50	Median effective dose
MIC	Minimum inhibitory concentration
Mr	Molecular weight
PDB	Protein Data Bank
RBC	Red blood cell
r.m.s.d.	Root mean square deviation
TCA	Tricarboxylic acid cycle

All amino acids are abbreviated to their standard one or three letter codes

All other units are S.I.

Table of contents

Chapter 1: Introduction

Chapter 2: X-ray Diffraction Theory and Crystallographic Principles

Chapter 5: Initial attempts to solve the structure of the *Toxoplasma gondii* enoyl acyl carrier protein reductase

Chapter 6: Structure determination and refinement of the TgENR/NAD^+/triclosan complex

List of Figures

All figures follow the given page numbers

Chapter 1: Introduction

Chapter 2: X-ray Diffraction Theory and Crystallographic Principles

Chapter 3: Structure determination of the *Plamodium falciparum* ENR/NAD^+/triclosan complex

Chapter 4: Analysis of the *P.falciparum*/NAD^+/triclosan complex structure

Chapter 9: Inhibitor binding and modelling studies of the apicomplexan *T.gondii* and *P.falciparum* ENR enzymes.

List of Tables

All tables follow the given page numbers

Chapter 1: Introduction

Chapter 3: Structure determination of the *P.falciparum* ENR/NAD^+/triclosan complex

Chapter 4: Analysis of the *P.falciparum*/NAD^+/triclosan complex structure

Chapter 5: Initial attempts to solve the structure of *Toxoplasma gondii* Enoyl ACP reductase

Chapter 6: Structure determination and refinement of *Toxoplasma gondii* Enoyl ACP reductase

Chapter 7: Analysis of the *T.gondii* ENR/NAD^+/triclosan complex structure

Chapter 8: Sequence and structural comparisons between the ENR family

Chapter 9: Inhibitor binding and modelling studies of the apicomplexan *T.gondii* and *P.falciparum* ENR enzymes.

CHAPTER 1

General Introduction

1.1 Fatty acid biosynthesis

1.1.1 Introduction

The use of fatty acids for membrane growth, maintenance and for the storage of energy in the form of triacyglcerols, makes fatty acid production essential to all cellular life. Fatty acid biosynthesis is catalysed by the Fatty Acid Synthase Complex (FAS), which exists in two forms. The type I FAS is found within yeast and most eukaryotic cells and has all of its catalytic domains on one or two polypeptides (Smith, 2003). In contrast the type II FAS is found in bacteria and plants and has each catalytic step on a separate polypeptide (Magnuson *et al.*, 1993). Although there is a clear difference in the two types of pathway they both carry out fatty acid synthesis through the same set of reactions and require an acyl-carrier protein (ACP) to carry the fatty acid substrate through each stage of the cycle. The overall reaction involves a growing fatty acid chain being elongated by the sequential addition of two carbon units derived from malonyl-CoA in each reaction cycle. The difference between the type I FAS pathway and the type II FAS pathway has led to it becoming a prime target for drug design where the toxicity to humans could be potentially low.

1.1.2 Fatty acid synthesis type I pathway

The type I FAS has all of the catalytic domains on one or two multifunctional polypeptide chains, with the yeast type 1 FAS pathway being encoded by two multifunctional domains, whereas the vertebrate and some bacterial type I FAS are encoded by just one multifunctional protein (Smith, 1994). The type I FAS is composed of eight distinct domains, seven of which have a discrete catalytic function. It has been proposed that the genes encoding the type I FAS pathway have "fused" over time, producing the large multidomain complex (Wakil *et al.*, 1989). One proposed advantage of this multidomain

system is the ability to move the substrate from one active site to another, removing the limitation of random diffusion of the reactants through the cytosol to successive enzymes in the pathway (McCarthy & Hardy, 1984).

1.1.3 Fatty acid synthesis type II pathway

The type II FAS pathway is composed of seven separate enzymes all encoded on different polypeptides and is responsible for carrying out fatty acid synthesis in Prokaryotes and plants (**Figure 1.1**). Each of these enzymes can be purified and catalyse their respective stage of the fatty acid synthesis pathway independently (Magnuson *et al.*, 1993). In plants and bacteria, the use of the type II FAS pathway may be advantageous to the host in that each stage can be independently regulated and the consequences of misincorporation of residues during the synthesis of the components may be reduced. The ability to over-express, purify and characterise each of the separate enzymes of the type II FAS independently has led to a detailed picture of each stage of the cycle to be developed (Rock & Cronan. 1996).

1.1.4 Enzymes of the type II FAS pathway

1.1.4.1 Acyl Carrier protein

The Acyl Carrier protein (ACP) is a small, charged, highly soluble protein and is one of the most abundant proteins within the *E.coli* cell (Magnuson *et al.*, 1993). The role of ACP is to covalently bind the acyl intermediates of fatty acid biosynthesis via a phosphopantetheine arm which is covalently linked to a fully conserved serine residue (Worsham *et al.*, 2003). The flexibility and 20Å length of the prosthetic group is essential in allowing the substrate into close contact with each enzyme active site without any unwanted protein-protein interactions. To date the structure of ACP has been solved

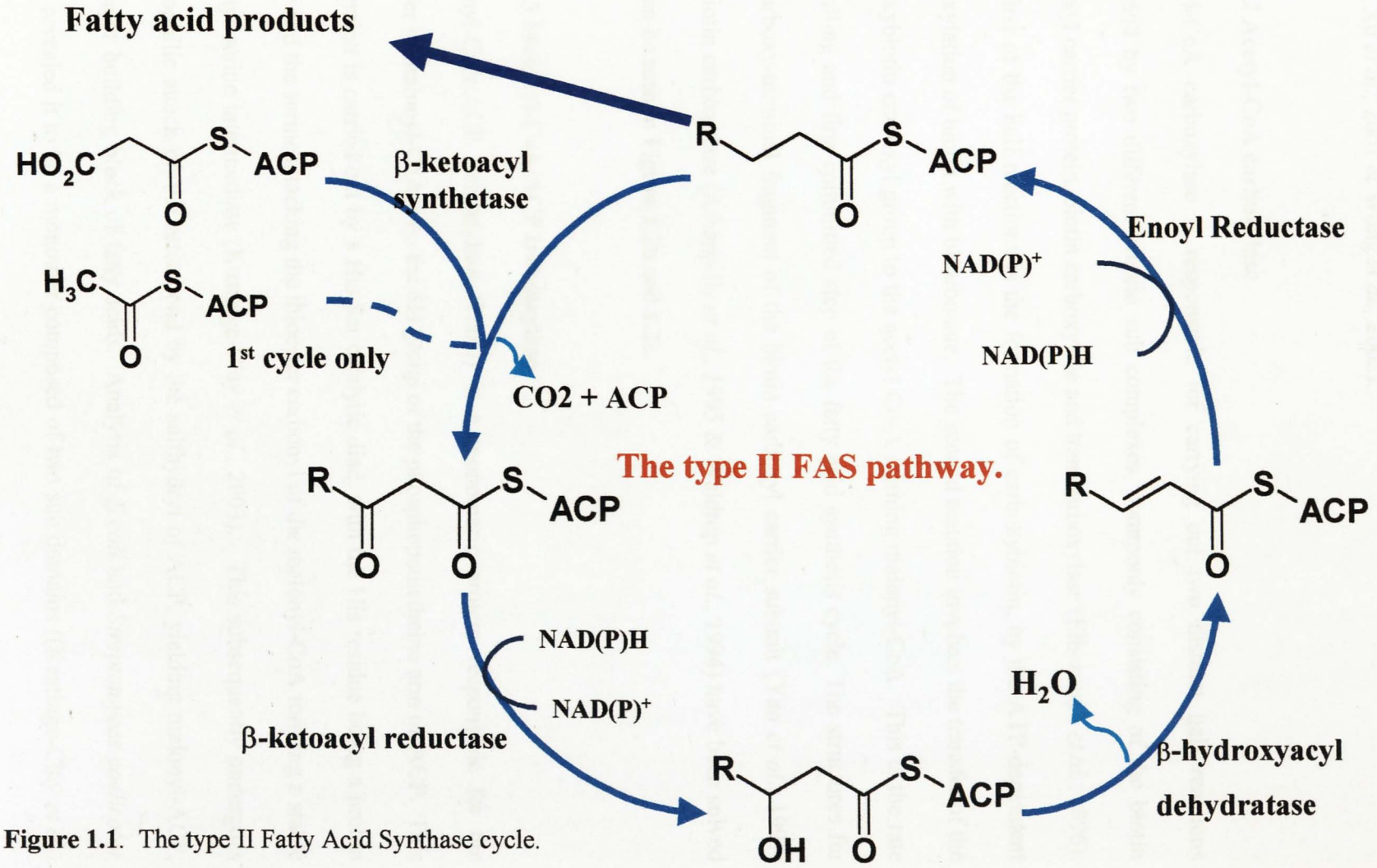


Figure 1.1. The type II Fatty Acid Synthase cycle.

in several species including *E.coli*, *B.subtilis* and *M.tuberculosis* and it is based on a 4 helix bundle (**Figure 1.2a**), packed through hydrophobic interactions (Roujeinikova *et al.*, 2002, Xu *et al.*, 2001 & Wong *et al.*, 2002).

1.1.4.2 Acetyl-CoA carboxylase

Acetyl-CoA carboxylase is responsible for carrying out two distinct half reactions catalysed by two different protein sub complexes, commonly consisting of the biotin carboxyl carrier protein, biotin carboxylase and transcarboxylase (Elborough *et al.*, 1996). The first of the half reactions is the formation of carboxybiotin, by the ATP-dependent carboxylation of biotin with bicarbonate. The second reaction involves the transfer of the carboxybiotin carboxyl group to the acetyl Co-A forming malonyl-CoA. This is the rate controlling and first committed step of the fatty acid synthesis cycle. The structures for the carboxy-terminal fragment of the biotin carboxyl carrier subunit (Yao *et al.*, 1997) and biotin carboxylase (Athappilly *et al.*, 1995 & Waldrop *et al.*, 1994) have been solved and can be seen in **Figure 1.2b** and **1.2c**.

1.1.4.3 Malonyl-CoA:ACP transacylase

Malonyl-CoA:ACP transacylase (FabD) is a monomeric protein responsible for the transfer of malonyl-CoA onto the SH group of the phosphopantetheine arm on ACP. This conversion is carried out by a His-Ser catalytic diad, with the His residue being a proton donor and the serine attacking the thioester carbonyl of the malonyl-CoA forming a stable malonyl-serine intermediate (Keatinge-Clay *et al.*, 2003). This subsequently undergoes nucleophilic attack on the ester bond by the sulfhydryl of ACP, yielding malonyl-ACP, the major building block of fatty acids. Analysis of *E.coli* and *Streptomyces coelicolor* FabD revealed it to be a monomer composed of two sub domains ((Keatinge-Clay *et al.*,

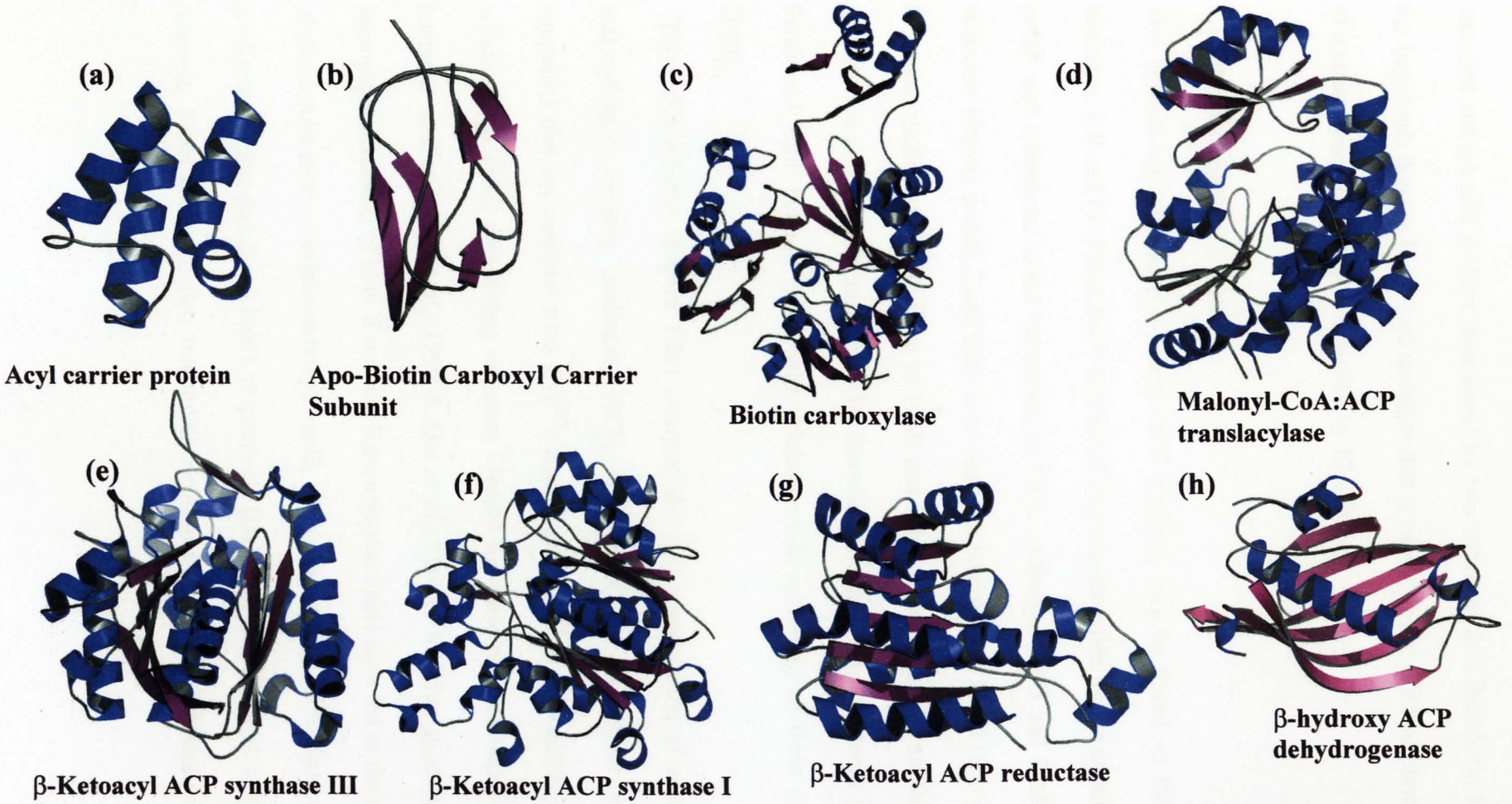


Figure 1.2. Representative structures for the enzymes of the type II FAS pathway. All figures are coloured with blue for α-helices and purple for β-strands. The structures are from *E.coli* except for e and f which are from *B. subtilis* and *S. coelicolor*, respectively. All structures were obtained from the rcsb protein databank and were determined using X-ray crystallographic techniques except for structures a and h which were determined using NMR techniques. The figure was produced using the graphics program Pymol (Delano, 2002).

2003 and Serre *et al.,* 1995)(**Figure 1.2d**)). The small domain is composed of a four stranded anti-parallel β-sheet surrounded by two α-helices in a ferredoxin like fold and the larger sub domain has a core structure that is similar to the α/β hydrolases composed of a central parallel β-sheet surrounded by 12 α-helices.

1.1.4.4 β-Ketoacyl ACP synthase

The condensing enzyme β-Ketoacyl ACP synthase can be found as three different isoforms, I, II and III (FabB, FabF and FabH) and is responsible for the addition of a CH_2 group and subsequent chain elongation in FAS. Although FabB and FabF share 38% sequence identity within *E.coli* there is no apparent homology to FabH (Qiu *et al.*, 2001). However studies of the structure of FabB and FabH have revealed a similar active site architecture and dimeric structure. Furthermore, the antibiotic cerulenin inhibits both FabB and FabF but not FabH, with thiolactomycin inhibiting all three (Price[a] *et al.,* 2001).

The FabH enzyme (**Figure 1.2e**) catalyses the first condensation of acetyl-CoA and malonyl-ACP yielding acetoacetyl-ACP, releasing CO_2. Structural studies have suggested that the catalytic triad Cys^{112}, Asn^{274} and His^{244} is responsible for catalysis within *E.coli* FabH with strong sequence identity for these residues being found in all bacterial species (Qiu[b] *et al.,* 1999 & Qui *et al.*, 2001). Inhibitor modeling and high throughput screening of both *E.coli* and *S.pneumoniae* FabH have led to the development of dichlorobenzyloxy-indole-carboxylic acid, which has an IC_{50} of 0.04μM and 0.83μM for *S.pneumoniae* and *E.coli* FabH, respectively (Daines *et al.,* 2003). A second study on *S.aureus* FabH produced two inhibitors, 5-chloro-4-phenyl-[1,2]-dithiol-3-one, and 4-

phenyl-5-phenylimino-[1,2,4]di-thiazolidin-3-one, which have an IC_{50} for FabH of 1.87μM and 0.775μM, respectively (He *et al.*, 2002).

The FabB and FabF enzymes differ from FabH in that an acyl-ACP primer is used rather than acyl-CoA. Furthermore, the FabB and FabF active sites have a conserved His-His-Cys triad in contrast to the FabH His-Asn-Cys triad (Price[a] *et al.*, 2001). The FabB enzyme (**Figure 1.2f**) is critical for the condensation of fatty acids within the type II FAS pathway and gene knockout mutants require fatty acid supplements for growth. The FabF enzyme controls the temperature dependent regulation of fatty acid conditions allowing the elongation of palmitoleate to cis-vaccenate (Magnuson *et al.*, 1993), with mutants lacking this gene growing under normal standard growth conditions. The overall reaction can be written as:

R-CO-S-ACP + HOOC-CH2-CO-S-ACP ⇔ R-CO-CH2-CO-S-ACP + CO2 + ACP-SH

1.1.4.5 β-Keto ACP reductase

The first reductive step in the fatty acid synthesis cycle is the formation of β-hydroxybutyryl-ACP, catalysed by β-Keto ACP reductase (FabG). The structure of FabG from *B.napus* has been solved to 2.3Å and revealed each subunit to be comprised of a central parallel seven stranded β-sheet flanked by eight α-helices in a manner similar to the enoyl reductase enzyme ((**Figure 1.2g**)(Fisher *et al.*, 2000)). The FabG active site has been shown within *E.coli* to contain a Ser, Tyr and Lys catalytic triad which through a Lys residue is linked to four water molecules (Price *et al.*, 2004). These water molecules are predicted to be responsible for replenishing the proton donated to the substrate during catalysis by the catalytic Tyr residue. Furthermore the binding of the NAD(P)H cofactor has been shown to cause a substantial structural change, positioning the catalytic triad

(Ser[138], Tyr[151] and Lys[155]) into the active conformation (Price[b] *et al.*, 2001). FabG shows negative cooperativity with the binding of ACP in one active site of the tetramer decreasing the affinity of the other sites to ACP allowing the active site to be cleared of product and promote NAD(P)H binding. Sequence alignments have shown a highly conserved YX_3K motif within the FabG family, common to the short-chain dehydrogenase/reductase (SDR) family (Filling *et al.*, 2002). The overall reaction can be shown as:

R-CO-CH2-CO-S-ACP + NAD(P)H ⇔ R-CH(OH)-CH2-CO-S-ACP + NAD(P)$^+$

1.1.4.6 β-Hydroxy ACP dehydratase

β-Hydroxy ACP dehydratase is responsible for catalysing the reversible dehydration of the hydroxyl group at C3 to produce trans-2-enoyl-ACP. Within *E.coli* two genes have been shown to encode a β-Hydroxy ACP dehydratase, FabA and FabZ, with FabA being more efficient at catalysing medium length substrates and FabZ efficiently catalysing short and long chain substrates (Heath & Rock. 1996). The FabA structure revealed a homodimer with each subunit containing a seven stranded antiparallel β-sheet wrapped around a five-turn α helix, reminiscent of a "hot dog fold" ((**Figure 1.2h**)(Leesong *et al.*, 1996)). The overall reaction can be shown as:

R-CH(OH)-CH2-CO-S-ACP ⇔ RHC=CH(CO)-S-ACP + H_20

1.1.4.7 Enoyl ACP reductase

Enoyl ACP reductase (FabI) is responsible for the last stage of the FAS pathway and involves the reduction of the trans-2-enoyl-ACP by the reduction of the double bond between positions C2 and C3 by the following reaction:

RHC=CH(CO)-S-ACP + NAD(P)H + H$^+$ ⇔ R-CH2-CH2-CO-S-ACP + NAD(P)$^+$

The structure of this enzyme has previously been solved for *B.napus*, *E.coli*, *H.pylori* and *M.tuberculosis* (Baldock *et al.*, 1996, Rafferty *et al.*, 1995, Moon *et al.*, 2001 and Dessen *et al.*, 1995) and will be dealt with in detail in section 1.2.

1.1.4.8 Successive Cycles

For every cycle of the type II FAS pathway two carbon atoms are added through the addition of a malonyl-ACP molecule (Magnuson *et al.*, 1993). The lipid chain is transferred from ACP to the cysteine side chain of the condensing enzyme (Wakil, 1989), after which a malonyl group is added to the ACP by the malonyl CoA:ACP transacylase. The lipid chain is then transferred back to the malonyl-ACP with the reactions catalysed by FabG, FabZ and FabI being repeated, resulting in the addition of two carbon units. Termination of this cycle occurs once the chain has reached the required length, typically 16 carbons (palmitoyl-ACP or C:16-ACP).

1.1.4.9 Chain termination

Palmitoyl thioesterase catalyses the hydrolysis of palmitoyl-ACP to yield one unit of palmitate and ACP resulting in chain termination (Vance *et al.*, 1973). It is essential that chain termination is carried out with long chain acyl-ACP's inhibiting the continuation of the cycle at the condensation stage, affecting the β-ketoacyl ACP synthase by feedback inhibition (Heath & Rock, 1996).

1.2 Enoyl ACP Reductase

1.2.1 Structural studies of enoyl reductase (ENR)

At the outset of work described in this thesis the structures of *E.coli*, *B.napus*, *H.pylori* and *M.tuberculosis* ENR had been determined (Baldock *et al.*, 1996, Rafferty *et al.*,

1995., Moon *et al.*, 2001 & Dessen *et al.*, 1995) The family shares both a common fold similar to the Rossmann dinucleotide binding fold (Rossmann *et al.*, 1974) and sequence similarities (**Figure 1.3**), especially for those residues situated close to the active site.

1.2.1.1 *B.napus* ENR

The crystal structure for *B.napus* (oil seed rape) ENR (BnENR), a 30 KDa protein organized into a homotetramer with four independent active sites, was determined to 1.9Å by X-ray crystallography (Rafferty *et al.*, 1995, **Figure 1.4**). The overall fold is reminiscent of a Rossmann nucleotide binding fold (Rossmann *et al.*, 1974) consisting of a central parallel β-sheet flanked on either side by α-helices which together form a hydrophobic core. The plant ENR enzymes contain two large inserted loops, which are distant to both the active and NAD^+ binding sites and appear to play no role in enzyme catalysis.

A striking similarity was noted between the BnENR structure and members of the short-chain alcohol dehydrogenase family as exemplified by the similarity to 3α, 20β-hydroxysteroid dehydrogenase (HSD)(Rafferty *et al.*, 1995). Both enzymes utilize NADH as a cofactor and have a conserved active site motif consisting of a catalytic tyrosine and lysine residue.

1.2.1.2 *M.tuberculosis* ENR

Mycobacterium tuberculosis ENR, (MtENR), is the target of the major anti tuberculosis drug isoniazid, the front line chemotherapeutic treatment for tuberculosis (Banerjee *et al.*, 1994) and differs from the ENR family as it prefers longer chain fatty acyl substrates for the purpose of synthesizing mycolic acids (Rozwarski *et al.*, 1998). In order to

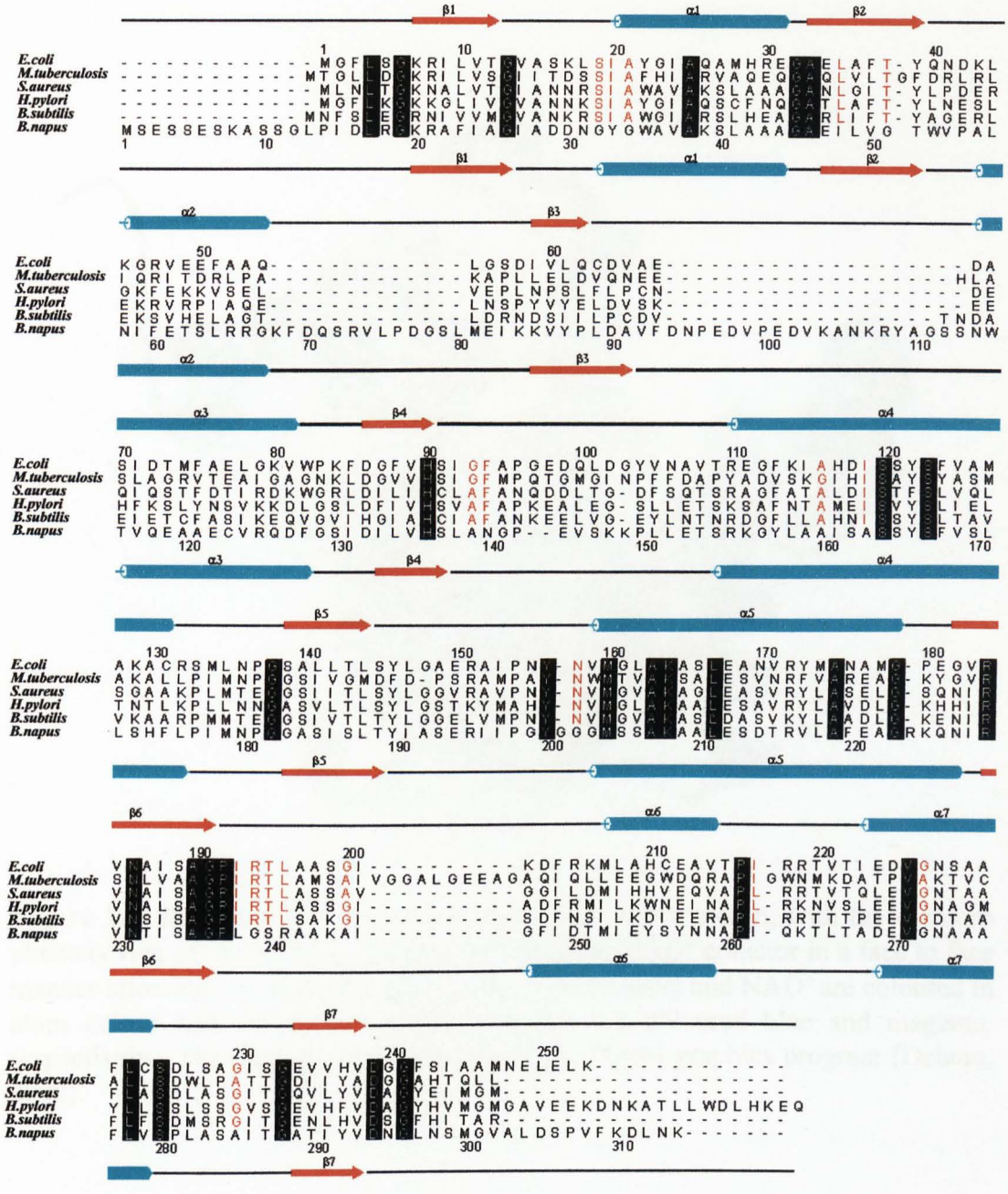


Figure 1.3. A structure based sequence alignment of the ENR enzymes from *E.coli*, *M.tuberculosis*, *S.aureus*, *H.pylori*, *B.subtilis* and *B.napus*. The structural elements and sequence numbering for the *E.coli* and *B.napus* enzymes are shown above and below the alignment, respectively. The blue cylinders represent α-helices and red arrows β-strands. Those residues fully conserved between all sequences are in black boxes and those residues with strong sequence similarity or sequence identity within the bacterial sequences are highlighted in red. The sequences were aligned using ClustalW (Higgins & Thompson 1994) and the figure was produced using the program Alscript (Barton, G.J. 1993).

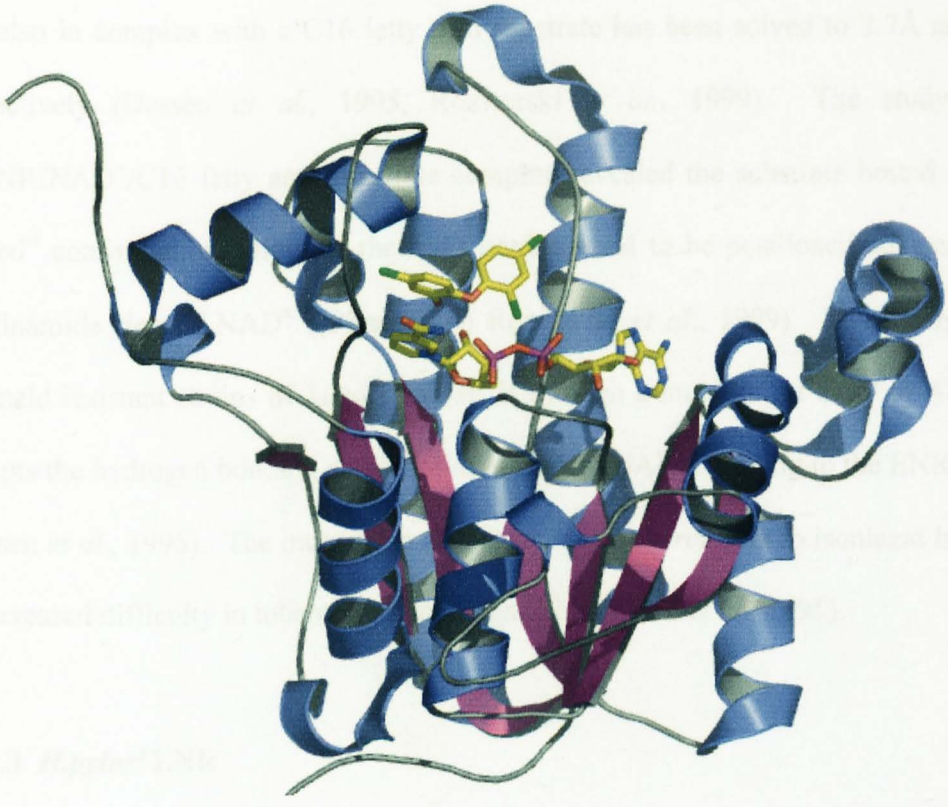


Figure 1.4. *B.napus* ENR complexed with NAD^+ and triclosan, with the 4-chloro phenoxy ring of the triclosan stacking on top of the NAD^+ cofactor in a face to face manner allowing π-π stacking interactions. The triclosan and NAD^+ are coloured in atom colour and the α-helices and β-strands are coloured blue and magenta, respectively. The picture was created using the Pymol graphics program (Delano, 2002).

accommodate a larger substrate a predominantly hydrophobic, glycine rich 10 amino acid insert, not present within other sequenced bacterial or plant ENR enzymes, increases the size of the substrate binding pocket. The structure of MtENR bound to its NAD^+ cofactor and also in complex with a C16 fatty acid substrate has been solved to 2.7Å and 2.8Å, respectively (Dessen *et al.*, 1995, Rozwarski *et al.*, 1999). The study of the MtENR/NAD^+/C16 fatty acyl substrate complex revealed the substrate bound in a "U-shaped" conformation, allowing the trans double bond to be positioned adjacent to the nicotinamide ring of NAD^+ ((**Figure 1.5**) Rozwarski *et al.*, 1999). The emergence of isoniazid resistant strains of *M.tuberculosis* is through a mutation of Ser^{94} to Ala which disrupts the hydrogen bonding network that stabilises NADH binding in the ENR enzyme (Dessen *et al.*, 1995). The increase in resistance of *M.tuberculosis* to isoniazid has led to an increased difficulty in tuberculosis treatment (Raviglione *et al.*, 1995).

1.2.1.3 *H.pylori* ENR

The structure of *H.pylori* ENR in complex with NAD^+ and diclosan has been deposited within the rscb PDB (1JW7) and has a fold similar to that described for other members of the ENR family with a parallel β-sheet central core flanked by several α-helices (**Figure 1.6 a**). The binding of the diclosan inhibitor is in a manner similar to triclosan with the 4-chloro phenoxy ring of diclosan stacking on top of the NAD^+ inhibitor making extensive π-π stacking interactions (**Figure 1.6 b**). There are currently no publications which describe the structure of *H.pylori* ENR.

Figure 1.5. *M.tuberculosis* ENR with bound NADH and C16 fatty acyl substrate. which can be seen in a "U-shaped" conformation, with the trans double bond placed above the NADH unit. The α-helices are in blue and the β-strands are in green. The figure was created using the graphics program Pymol (Delano, 2002).

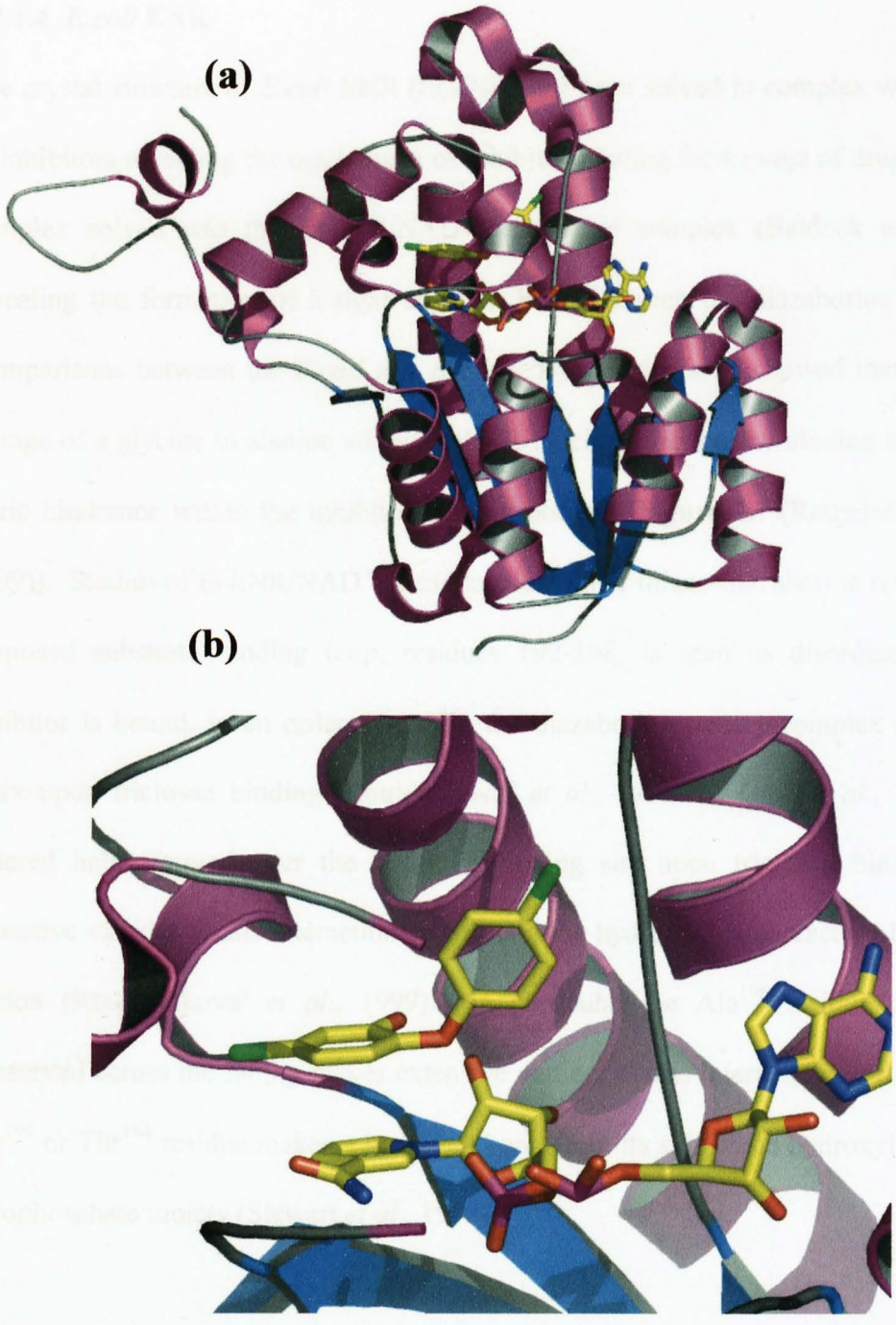


Figure 1.6. (a) *H.pylori* ENR complexed with NAD^+ and diclosan. (b) A close up view of the diclosan binding site showing the stacking interactions formed between the diclosan inhibitor and NAD^+ cofactor. All figures are coloured with yellow for carbon, blue for nitrogen, red for oxygen, green for chlorine and purple for phosphorous. Figures were produced using the graphics program Pymol (Delano, 2002).

1.2.1.4 *E.coli* ENR

The crystal structure of *E.coli* ENR (EcENR) has been solved in complex with a variety of inhibitors revealing the mechanism of inhibitor binding for a range of drugs. The first complex solved was the EcENR/NAD^+/diazaborine complex (Baldock *et al.*, 1996) revealing the formation of a tight covalent bond between the diazaborine and NAD^+. Comparisons between the *E.coli* and *B.napus* ENR active sites showed that a sequence change of a glycine to alanine within BnENR blocked diazaborine binding by producing steric hindrance within the inhibitor binding pocket ((**Figure 1.7**)(Roujeinikova[b] *et al.*, 1999)). Studies of EcENR/NAD^+ bound to triclosan or thieno-diazaborine revealed that a proposed substrate binding loop, residues 192-198, is seen as disordered when no inhibitor is bound, in an ordered loop in the diazaborine ternary complex and forms a helix upon triclosan binding (Roujeinikova[b] *et al.*, 1999 & Qiu[a] *et al.*, 1999). The ordered helix formed over the inhibitor binding site upon triclosan binding, makes extensive van der Waals interactions increasing the hydrophobic character of the binding pocket (Roujeinokova[a] *et al.*, 1999). In particular the Ala^{196} residue that is fully conserved across the family makes extensive van der Waals interactions and a conserved Ser^{194} or Thr^{194} residue makes a hydrogen bond from its side chain hydroxyl atom to the pyrophosphate moiety (Stewart *et al.*, 1999).

1.2.2 Inhibitors of the Enoyl reductase enzyme

1.2.2.1 Triclosan

ENR is the target for the broad spectrum anti microbial agent triclosan (McMurry *et al.*, 1998). Triclosan has been demonstrated to be a very potent inhibitor of ENR and is found in many household formulations including toothpastes, mouthwashes and chopping boards ((**Table 1.1**)(Bhargava & Leonard 1996)). Its success arises from its ability to

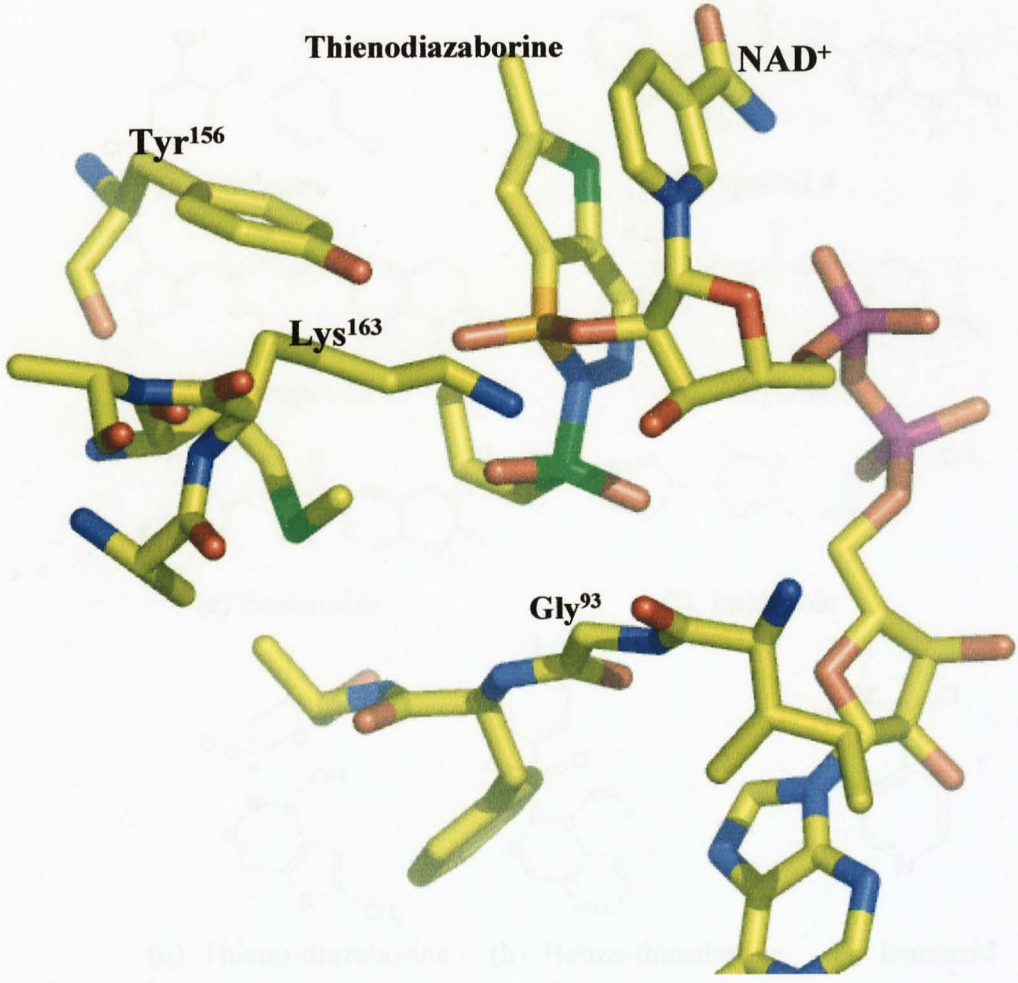


Figure 1.7. The active site of *E.coli* ENR with NAD^+ and thienodiazaborine bound. The boron atom of the inhibitor (orange) is approximately 1.5 Å away from the NAD^+ cofactor allowing for the formation of a covalent bond. The two fully conserved catalytic residues, Tyr and Lys are labelled and are in close proximity to the inhibitor. The residue Gly^{93} is within 3.5 Å of the inhibitor, however if replaced by an alanine residue, a feature common to the plant ENR enzymes this distance is reduced to 2.3 Å, causing steric hindrance and subsequently inhibiting drug binding. The figure is coloured in atom colour with yellow for carbon, red for oxygen, green for sulphur, blue for nitrogen, purple for phosphorous and orange for the boron atom in the diazaborine inhibitor. The figure was produced using the graphics program Pymol (Delano, 2002).

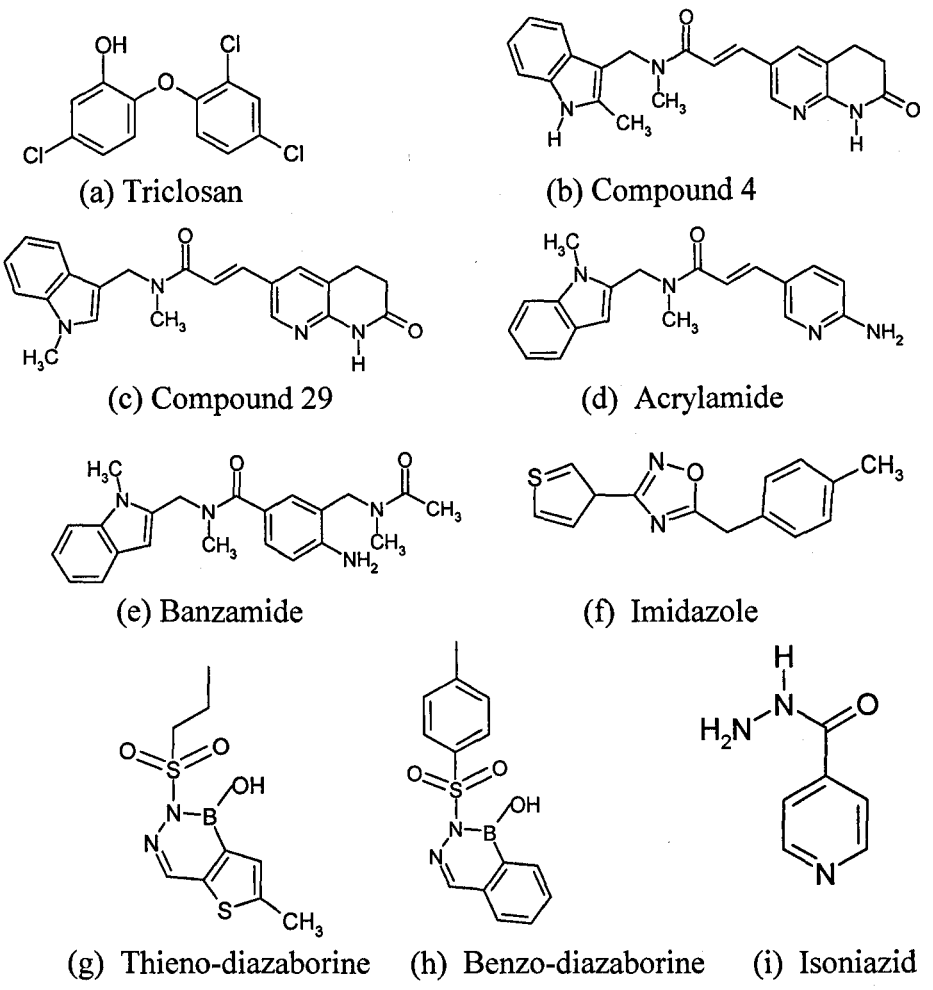


Figure 1.8 The structural formulae of nine ENR enzyme inhibitors, produced using the program Isis draw.

	FabI IC50 (μM)			MIC (μg/mL)		
Compound	***S.aureus***	***H.influenzae***	***E.coli***	***S.aureus***	***H.influenzae***	***E.coli***
Triclosan	0.07±0.05[a]	—	0.43±0.036[c]	≤0.06[a]	≤0.06[a]	—
Compound 4	0.047[b]	0.13[b]	—	0.06[b]	2.0[b]	8.0[b]
Compound 29	0.13±0.03[c]	0.39±0.06[c]	0.07[c]	0.06[c]	16.0[c]	8.0[c]
Acrylamide	2.4±1.2[a]	4.2±1.1[a]	—	0.5[a]	>64[a]	—
Benzamide	6.7±0.5[a]	4.7±0.3[a]	—	16[a]	>64[a]	—
Imidazole	50.7±16.9[d]	—	>100[d]	—	—	—
Thieno-diazaborine	—	—	—	—	—	1.6[e]
Benzodiazaborine	—	—	—	—	—	12.5[e]
Isoniazid	—	—	—	—	—	—

Table 1.1 The inhibitory properties of several inhibitors of the ENR enzyme from *S.aureus*, *H.influenza* and *E.coli*. [a]values taken from Miller[b] *et al.*, 2002, [b] values taken Payne *et al.*, 2002, [c] values taken from Seefeld *et al.*, 2003, [d] values taken from Heerding *et al.*, 2001, [e] values taken from Grassberger *et al.*, 1984.

inhibit the ENR enzyme at sub μm concentrations with inhibitor binding being shown within EcENR to be stoichiometric (Ward *et al.*, 1999). Furthermore, it has been shown within EcENR that triclosan binding is reversible with neither the NAD^+ or triclosan being modified during inhibition. Triclosan is however highly insoluble and must be dissolved in a solvent such as DMSO, limiting its use as a microbial inhibitor. Its mode of binding is through the mimicking of the fatty acid substrate with the 4-chloro phenoxy ring of triclosan and in particular the alcohol group mimicking the enolate anion intermediate ((**Figure 1.8**)(Levy *et al.*, 1999)). The 4-chloro phenoxy ring of triclosan is responsible for stacking on top of the NAD^+ nicotinamide unit forming π-π stacking interactions and extensive van der Waals interactions. Furthermore a hydrogen bond is formed between the triclosan alcohol group and the catalytic tyrosine of ENR (Baldock *et al.*, 1996). The formation of the ENR/NAD^+ complex has been shown to be critical for triclosan binding (Heath *et al.*, 2000). This tight binding of the triclosan blocks fatty acid binding, subsequently inhibiting the enzyme.

1.2.2.2 Isoniazid

Isoniazid (**Figure 1.8**), is currently a first line drug for tuberculosis and works through the inhibition of MtENR. Isoniazid is a prodrug and requires a mycobacterial catalase peroxidase (KatG) to convert it into the active form of the drug that covalently binds to the NAD^+ cofactor causing inhibition (Rozwarski *et al.*, 1998). It has been shown that isoniazid binding is a two step enzyme inhibition mechanism, with an initial weak binding complex followed by a slow conversion to the tight binding final complex (Rawat *et al.*, 2003). The inhibitory effects of isoniazid are limited to the *Tuberculosis* family

due to the requirement of drug conversion and as such have a limited use as a broad spectrum antibmicrobial.

1.2.2.3 Diazaborines

Inhibition of the ENR enzyme by members of the diazaborine family involves the formation of a covalent bond from the diazaborine boron atom to the 2'-hydroxyl group of the nicotinamide ribose moiety of the NAD^+ cofactor ((Baldock *et al.*, 1996)(**Figure 1.7**)). Although effective inhibitors, the diazaborine family are potentially toxic due to the structural element of arenoboronic acid amide ((**Figure 1.8**)(Grassberger *et al.*, 1984)). Furthermore a point mutation of Gly^{93} to a bulkier Ala residue, which is common in plant enzymes, leads to diazaborine resistance through steric hindrance of the inhibitor (Roujeinikova[b] *et al.*, 1999).

1.2.2.4 New and Developing drugs

Initial studies carried out by GlaxoSmithKline Pharmaceuticals revealed that the 1,4-disubstituted imidazoles (**Figure 1.8**) target EcENR with modest success ((**Table 1.1**) (Heerding *et al.*, 2001)) with the thiophene ring of the inhibitor forming π-stacking interactions with the nicotinamide ring of the NAD^+ cofactor (**Figure 1.9**). A second class of inhibitors reported by GlaxoSmithKline Pharmaceuticals is based on aminopyridine derivatives ((Miller[b] *et al.*, 2002)(**Figure 1.8**)) whose inhibitory properties were better than those previously reported for the imidazole based drugs (**Table 1.1**). The structure of EcENR in complex with NAD^+ and the aminopyridine based inhibitors ((1LX6 benzamide, 1LXC acrylamide) has shown that these inhibitors interact with the fully conserved Tyr^{146} and Tyr^{156} residues and also Phe^{203}, with the linking amide π-

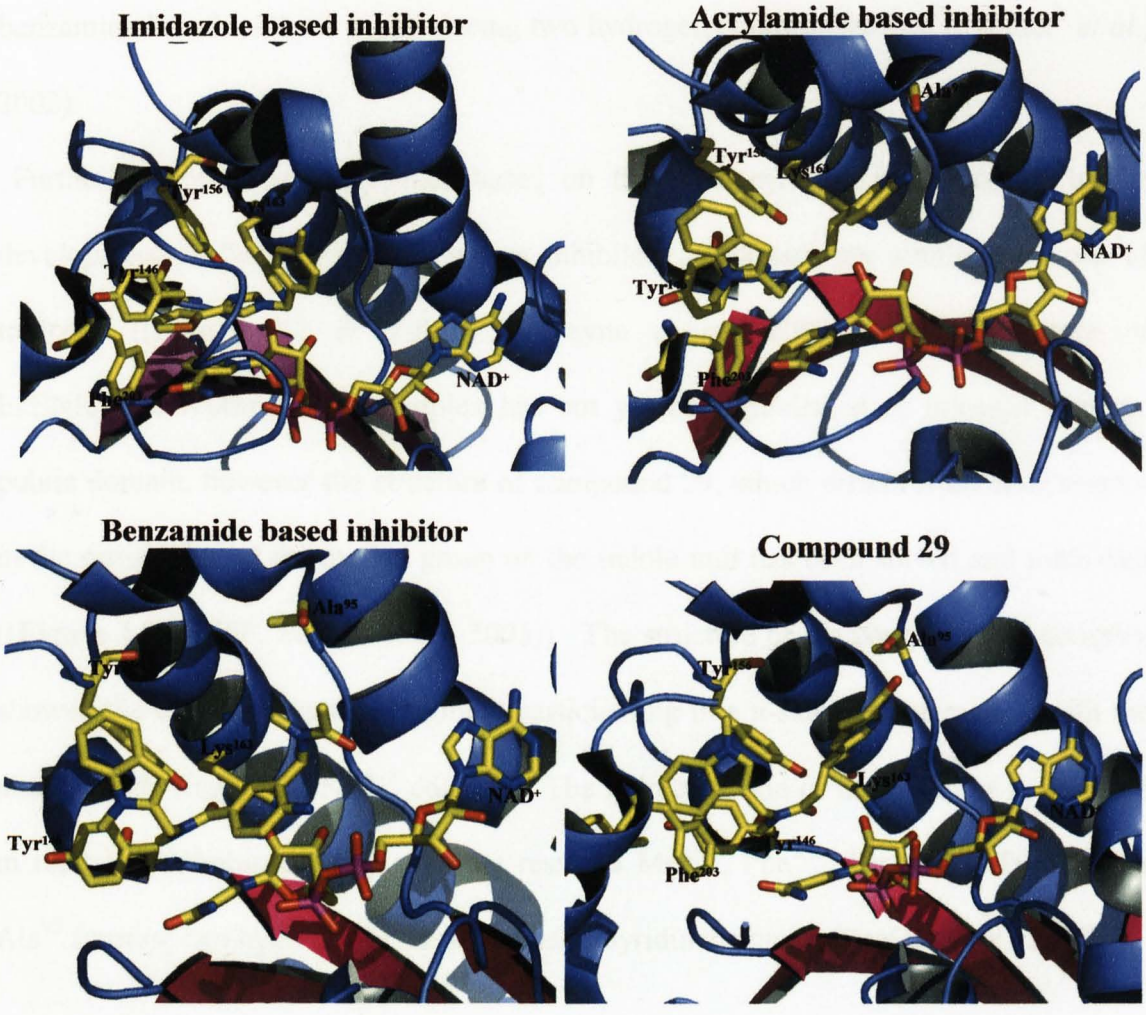


Figure 1.9. A diagrammatic representation of several inhibitors bound within the active site of EcENR. For all figures the atoms are coloured with yellow for carbon, blue for nitrogen, red for oxygen, purple for phosphorous and orange for sulphur. The figure was produced in the graphics program Pymol (Delano, 2002).

stacking over the nicotinamide ring of the NAD^+ cofactor (**Figure 1.9**). However the aminopyridine subunit of the acrylamide based inhibitor binds in a different mode to the benzamide based inhibitor by producing two hydrogen bonds to the Ala^{95} (Miller[b] *et al.*, 2002).

Further screening of compounds based on the aminopyridine inhibitors led to the development of "compound 4", whose inhibitory properties are similar to those of triclosan ((**Figure 1.8 & Table 1.1**)(Payne *et al.*, 2002)). The structure of EcENR/NAD^+/compound 4 complex has not yet been published or released into the public domain, however the structure of compound 29, which differs from compound 4 in the positioning of the methyl group on the indole unit has been solved and published ((**Figure 1.9**)(1MFP, Seefeld *et al.*, 2003)). The structure of the compound 29 complex showed the central cis-amide fragment participating in a π-stacking interaction with the nicotinamide ring of the NAD^+ cofactor. The indole portion of the inhibitor is involved in tight hydrophobic packing with the residues Met^{206}, Phe^{203}, Tyr^{156} and Tyr^{146}, with Ala^{95} forming two hydrogen bonds to the naphthyridinone group (Seefeld *et al.*, 2003).

1.2.3. Enzyme mechanism

The proposed ENR catalytic mechanism, involves the hydride transfer from the C4 position of NADH to the C3 position at the double bond of the enoyl substrate ((Rafferty *et al.*, 1995)(**Figure 1.10a** and **b**)). This leads to the formation of an enolate anion intermediate, which can be protonated on the oxygen atom to form an enol, which subsequently goes through tautomerization forming the reduced acyl product. There are two residues which are fully conserved between all ENR sequences implicated in this mechanism: a Tyr residue is the proton donor to the enolate anion and a Lys residue

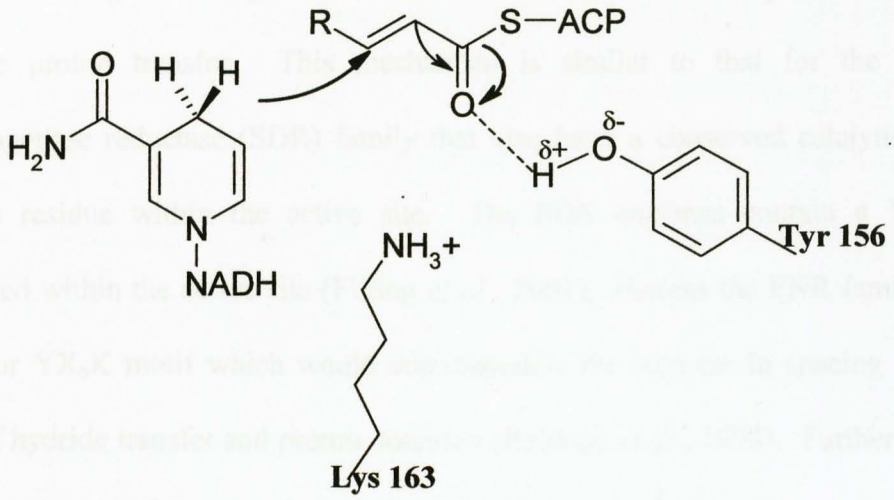


Figure 1.10a. The proposed catalytic mechanism for the ENR enzyme, numbered with respect to the *E.coli* structure. Figure adapted from Bladock *et al.*, 1998.

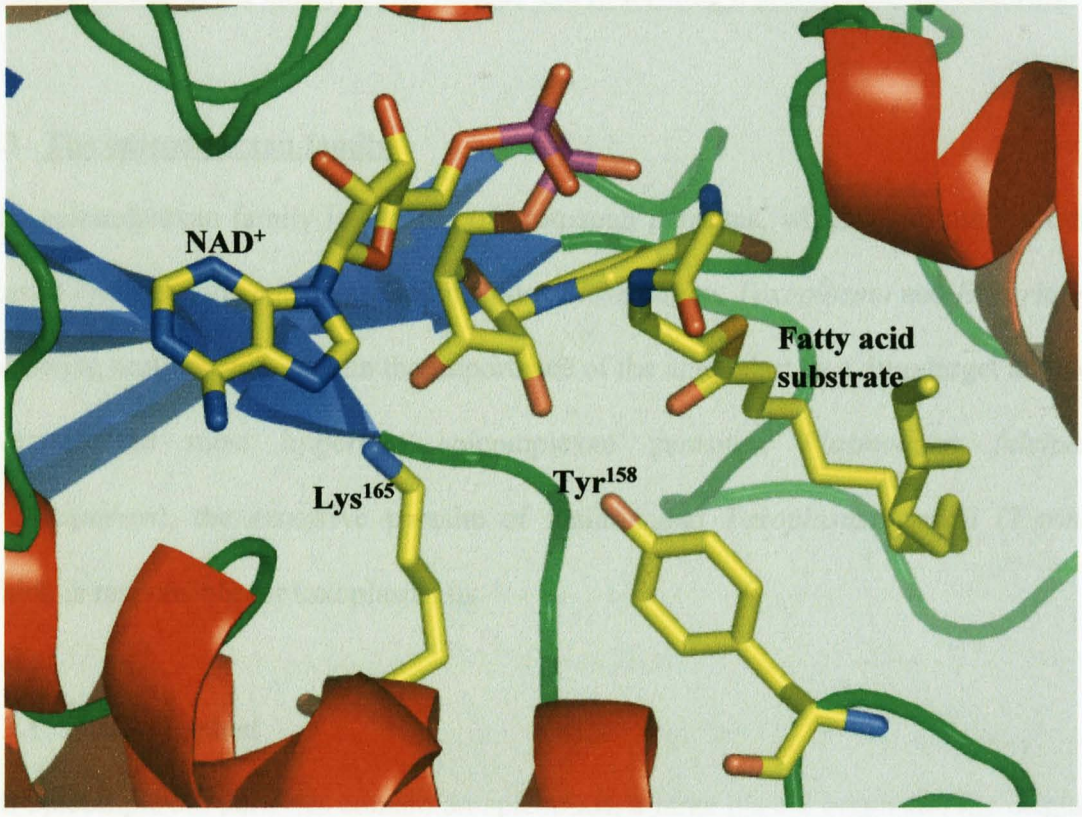


Figure 1.10b. The active site of *M.tuberculosis* ENR, with bound NAD^+ and fatty acid substrate. The ENR secondary structure is shown with α-helices coloured in red, β-strands in blue and loop regions in green. The substrates and residues are coloured with yellow for carbon, red for oxygen, blue for nitrogen, purple for phosphorous and orange for sulphur. The residues are numbered with respect to *M.tuberculosis* ENR. The figure was produced using the graphics program Pymol (Delano, 2002).

stabilises the negative charge of the transition state and lowers the pKα of the Tyr-OH to promote proton transfer. This mechanism is similar to that for the short chain dehydrogenase reductase (SDR) family that also have a conserved catalytic lysine and tyrosine residue within the active site. The SDR enzymes contain a YX_3K motif conserved within the active site (Filling *et al.*, 2001), whereas the ENR family contain a YX_7K or YX_6K motif which would accommodate the increase in spacing between the point of hydride transfer and proton donation (Baldock *et al.*, 1998). Furthermore two or three water molecules are present within the active site of the SDR enzymes, (Filling *et al.*, 2002, Benach *et al.*, 1998) the role of which is predicted to be in stabilising the positive charge of the catalytic Lys residue and in forming a proton relay mechanism.

1.3 The apicomplexan family

The apicomplexan family is a group of protozoan parasites, which all contain a unique plastid known as the apicoplast and includes *Plasmodium*, *Toxoplasma* and *Eimeria*. The following sections will explain the importance of the apicoplast as a drug target and detail two of the most important apicomplexan parasites, *Plasmodium falciparum* (*P.falciparum*), the causative parasite of malaria and *Toxoplasma gondii* (*T.gondii*), which is responsible for toxoplasmosis.

1.3.1 The apicoplast

All apicomplexan parasites contain an apicoplast, a large plastid organelle, the origins of which arose through the endosymbiosis of a cyanobacterial like prokaryotic cell (McFadden *et al.*, 1999). The presence of four membranes surrounding the apicoplast demonstrates secondary symbiosis in which a plastid containing the primary endosymbiont has been engulfed and retained by the apicomplexan parasites (Waller *et*

al., 1998). The initial algae origins of this organelle are unclear due to the lack of any photosynthetic pigment, although phylogenic trees based on elongation factor Tu, have suggested a green-algal ancestry (Kohler *et al.*, 1997). This has resulted in a eukaryotic parasite family containing several prokaryotic enzymatic pathways, which may prove important in the design of inhibitors with potentially low toxicity in man.

1.3.2 The apicoplast as a drug target

The apicoplast genome is the smallest plastid genome currently reported at 35 Kb and is composed of a supercoiled circular architecture (Gleeson *et al.*, 2000). It is thought that the majority of genes, which were once encoded for by the apicoplast genome have, over time been incorporated into the apicomplexan parasites genome and are subsequently, targeted back. Importantly ciproflaxin and other flouroquinolones, which have been shown to have an inhibitory effect on prokaryotic DNA gyrases, inhibited the replication of the apicoplast and resulted in cell death (Fichera & Roos, 1997). Furthermore it is completely devoid of any internal membranes and contains little or no photosynthetic pigment. Hence its maintenance was hypothesised to be based on other essential metabolic pathways or functions which it carries out and not through any photosynthetic role (McFadden & Roos 1999). One such pathway is the shikimate pathway common to plant plastids, which is inhibited by the herbicide glyphosate (Keeling *et al.*, 1999 & Roberts *et al.*, 1998). The apicomplexan parasite *T.gondii* was shown to be sensitive to glyphosate, with the genes encoding the shikamate pathway being encoded on the *T.gondii* genome but targeted towards the apicoplast. The *P.falciparum* genome project has revealed both a non-mevalonate pathway that leads to isopentenyl diphosphate subunit synthesis (Jomaa *et al.*, 1999) but also several key enzymes in the fatty acid synthesis type II pathway, (FabH, FabZ and ACP), (Waller *et al.*, 1998), which has been

shown to be the target of several antimicrobial drugs. It is hoped that the localization and characterization of a possible ENR enzyme, which may also be targeted towards the apicoplast, could lead to the development of novel inhibitors against the apicomplexan family.

1.3.3 Targeting to the Apicoplast

The apicoplast is predicted to be targeted by approximately 500 proteins (Foth *et al.*, 2003), all of which must be able to target and cross the 4 membrane bound plastid organelle. The minimum time for targeting proteins to the apicoplast was shown to be less than 45 minutes (Dooren *et al.*, 2002) and is achieved by using a bipartite N-terminal signal sequence. The signal sequence which targets the proteins to the secretary pathway has been shown to contain a von Heijne-type cleavage motif which is composed of a positively charged region followed by a short hydrophobic region which precedes the neutral but polar end region (Nielson *et al.*, 1997). The second part of the signal sequence, the transit peptide, has a net positive charge which is required to transport the proteins across the multiple membranes and divert the protein into the apicoplast (Waller *et al.*, 2000). Although the sequence homology for the transit sequence is very low with no known secondary structure, there are a set of commonalities. The transit peptides are low in acidic residues but highly enriched in lysine and asparagine residues, the former resulting in an overall positive charge. Furthermore >90% of apicoplast transit peptides contain an Hsp70 binding site, suggesting an importance in apicoplast targeting (Foth *et al.*, 2003).

1.4 *Plasmodium falciparum*

The *Plasmodium* species is a large family of intracellular parasites responsible for the development of malaria within specific mammalian hosts. There are four species responsible for infecting Homo *sapians*, they are *P.vivax*, *P.malariae*, *P.ovale* and *P.falciparum*, the most important of which is *P. falciparum*, which is responsible for over 95% of severe cases of malaria. The global impact of malaria is extensive with over 500 million people infected a year of whom an estimated 2 to 3 million die (Breman *et al.*, 2001). Recently the genome sequence was completed for *P.falciparum*, with work continuing for several other *Plasmodium* species.

1.4.1 Pathology of the *Plasmodium falciparum* parasite

The factors affecting the onset of disease after a malarial infection are complex and involve a variety of organs and tissues within the body, with the parasite invading two cell types within the body, the red blood cells (RBC) and liver hepatocytes. The infection and subsequent destruction of a large proportion of RBC results in various degrees of anaemia. This can be compounded by the *P.falciparum* parasite, which causes rosseting of red blood cells, which in turn attach to blood vessel walls causing an obstruction in blood flow (Dondorp *et al.*, 2000). Importantly the lack of a fully functional tricarboxylic acid cycle within the parasite results in the majority of its energy being produced through anaerobic respiration, resulting in the mass production of lactic acid, which can build up in the host's blood. Furthermore damage caused to the liver through hepatocyte invasion, results in a decrease in efficiency of lactic acid removal (English *et al.*, 1997). The increase in metabolic stress caused in the main by severe anaemia, liver damage and blood acidosis results in a decrease in the oxygen delivery to tissues, including the brain and may result in a coma (English[a] *et al.*, 1996). There is a need for significant amounts

of clean water to re-hydrate the body during infection, where clean water is unavailable severe dehydration can occur and add to the stresses already placed on the body (English[b] *et al.*, 1996).

1.4.2 Life cycle of the *Plasmodium falciparum* parasite within a human host

The initiation of the *P.falciparum* lifecycle within the human host (**Figure 1.11**) is through the introduction of sporozoites into the body either directly into the blood stream or into the subcutaneous tissue by a mosquito bite. Once inside the blood stream the sporozoites quickly migrate over 30-60 minutes to the liver passing through several different hepatocyte cells before going through invasion and the subsequent parasite development within the selected hepatocyte cell (Mota *et al.*, 2001). Within the hepatocytes the sporozoites develop into tens of thousands of merozoites, each merozoite capable of invading a RBC upon release from the liver and it is this invasion of the RBC, which causes the onset of the clinical symptoms of malaria (Miller[a] *et al.*, 2002). After the red blood cells are invaded by merozoites they continue to mature and divide asexually, rupturing 48 hours later. It is at this point that a small proportion of the merozoites develop into gametocytes, essential for the transmission to others through the mosquito vector. These gametocytes are taken up by the mosquito when it takes a blood meal allowing for parasite development within the mosquito. The complexity of the *Plasmodium* lifecycle and the intracellular nature of all but the short lived merozoite stage makes it difficult for a host to mount an efficient immune response and leads to difficulties in the production of an effective vaccine against the *Plasmodium* parasite.

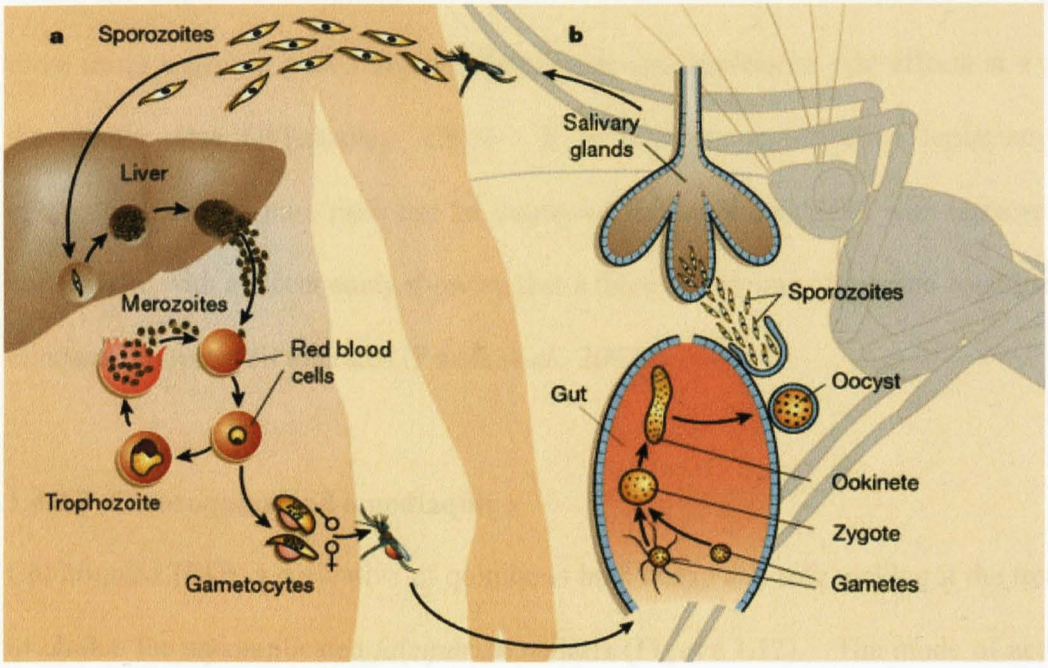


Figure 1.11. A diagrammatic representation of the *Plasmodium* lifecycle within both a human host and the definitive Anophelles mosquito host. The figure is taken from the Sanger Centre (www.sanger.ac.uk).

1.4.3 Currently available antimalarial drugs

1.4.3.1 Quinine

The first antimalarial drug to be used on a global scale was Quinine (**Figure 1.12**), a relatively inexpensive compound (**Table 1.2**), which is extracted from the bark of the Cinchoma tree. It was used as a first line drug in the U.K, but this practice is unusual within the tropics as compliance is a major problem caused by the need to take Quinine three times a day for seven days, its bitter taste and unpleasant side effects at a normal therapeutic dose (Winstanley, 2003). It is however an effective replacement for chloroquine and its cure rates can be improved further if combined with tetracycline or clidamycin, with a recent study showing that a three day regime of quinine combined with clindamycin was well tolerated (Parola *et al.*, 2002).

1.4.3.2 Chloroquine and amodiaquine

Chloroquine (CQ), a derivative of quinine is both cheap and safe making it the treatment of choice for uncomplicated *falciparum* malaria (**Figure 1.12**). The mode of action for CQ and its related analogues is thought to be through inhibiting the removal of the toxic product of hemoglobin degradation, ferriproporphyrin IX (FP) (Ginsburg *et al.*, 1999 & Sanchez *et al.*, 2000). It is suggested that CQ forms a complex with FP, in so doing the parasite is unable to polymerize FP into harmless hemozoin. This subsequently leads to an increase in cellular FP and an increase in membrane permeability resulting in parasite cell lysis. In addition to this chloroquine has been shown to bind within the cofactor binding site of the *P.falciparum* lactate dehydrogenase, acting as a competitive inhibitor in blocking NADH binding by occupying a position similar to that of the adenine ring of the NADH and forming interactions with several amino acids unique to the malarial form of the enzyme (Read *et al.*, 1999). The more expensive amodiaquine is thought to

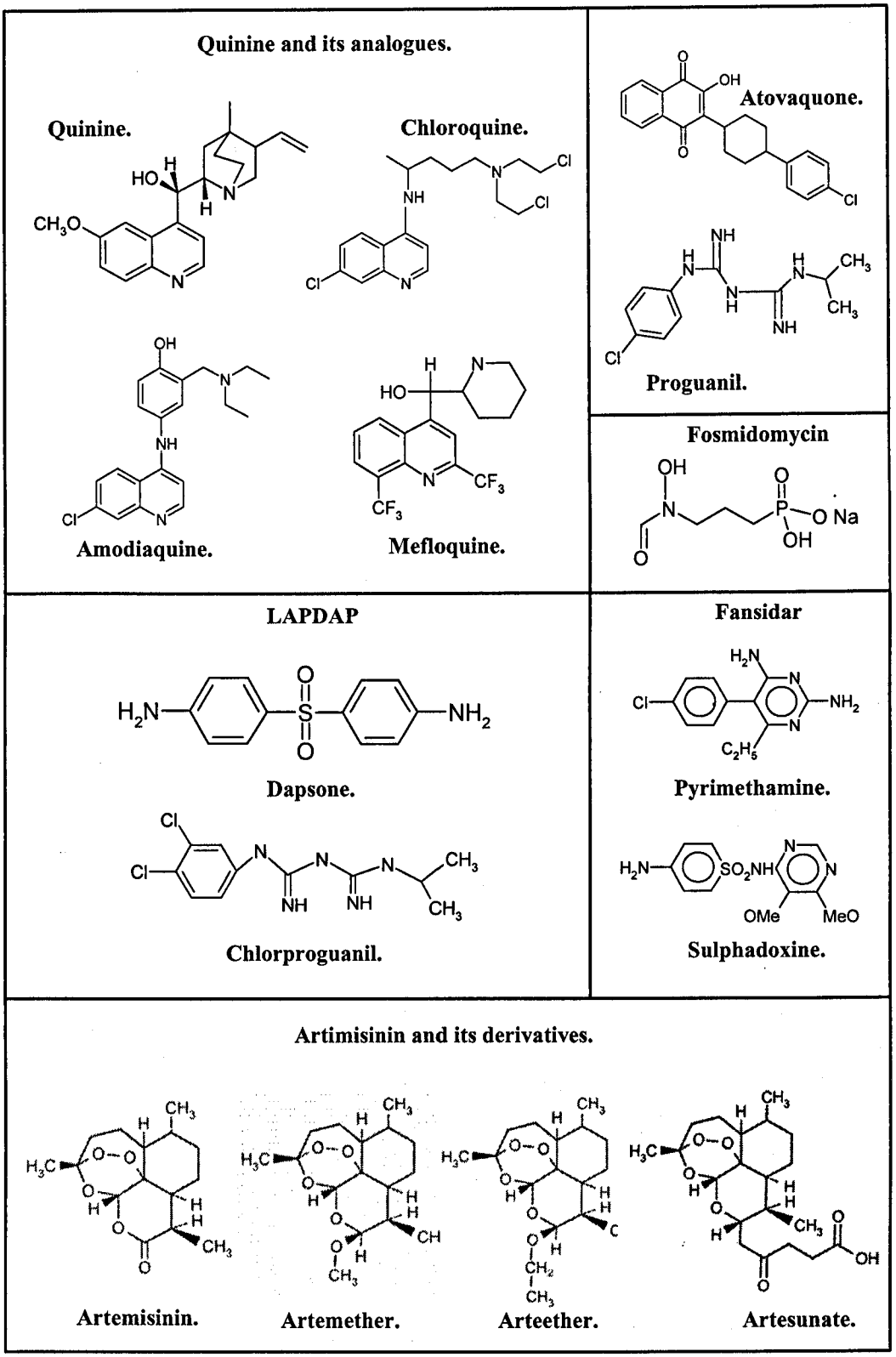


Figure 1.12. The structural formulae for the main antimalarial drugs currently in use. The figures for artimisinin and its derivatives were taken from www.dafra.be/english/artemether . All other figures were created in the program IDraw32.

Drug	Cost of adult dose US$ 1990	Cost of adult dose US$ 2003	Introduced
Chloroquine	**0.12**	**0.1**	**1945**
Sulphadoxine-pyrimethamine	**0.2**	**0.13**	**1967**
Amodiaquine	**-**	**0.3**	**1945**
Quinine	**1.8**	**1.5**	**1632**
Mefloquine	**2.4**	**2.0**	**1977**
Artemisinin	**6.0**	**1.0-3.0**	**1984**
LapDap	**-**	**5.0[a]**	**TBR[b]**
Halofantrine	**7.0**	**5.0**	**1977**
Atovaquone proguanil	**-**	**30.00**	**1996**

Table 1.2. The relative cost of antimalarial treatments over a 12 year period, values from 1990 are taken from Molyneux *et al.*, 1999, data for 2003 is taken from Pharmacopeia 2003. (a) recent publications have suggested that LapDap may become available at non profit prices within Africa, costing approximately $0.29 for a three day adult treatment. (Pincock *et al.*, 2003). (b) The general release date is yet to be established for LapDap

produce growth inhibition in a manner similar to that of chloroquine but with a higher efficacy and as such is a common second line drug in many areas of Africa.

1.4.3.3 Mefloquine (Larium)

Mefloquine is a more expensive member of the quinolone family and (**Figure 1.12**) although too expensive for general use throughout tropical Africa is used in wealthier countries with multidrug resistance such as Thailand (Winstanley, 2003). An increase in drug resistance has led to its combination with artimisinin derivatives, increasing the overall effectiveness of the drug.

1.4.3.4 Artemisinins

Acting against the sexual stage gametocytes within the blood, artemisinins are potent and fast acting but expensive in comparison to the quinoline drug family (**Table 1.2**). Several semi-synthetic derivatives of artemisinin are in use and include artemether, artesunate and arteether, all of which are metabolised to the main active agent dihydroartemisinin ((**Figure 1.12**)(Meshnick *et al.*, 2002)). The mechanism of action of the artimisinin based drugs is yet to be established with evidence supporting two different modes of action being published (Ridley, 2003). It has been hypothesized that the endoperoxide bridge, necessary for antimalarial activity, breaks open in the presence of intra parasitic heme and iron in turn releasing toxic free radicals, which inhibit several parasitic pathways (Olliaro *et al.*, 2001). More recently it has been suggested that these drugs target and irreversibly inhibit the malarial calcium-dependant ATPase (Eckstein-Ludwig *et al.*, 2003).

1.4.3.5 Atovaquone/proguanil (Malarone)

The antimalarial atovaquone (**Figure 1.12**) targets the cytochrome c reductase, which is part of the mitochondrial electron transport system (Fry *et al.*, 1992), however a point mutation has lead to the onset of drug resistance (Srivastava *et al.*, 1999). The rapid development of resistance after only several months of general release has led to the drug being used in conjunction with proguanil in a synergistic manner. Proguanil is metabolized to the active reagent cycloguaunil that primarily exerts its effects by inhibiting the dihydrofolate reductase-thymidylate synthase (DHFR-TS) and disrupting deoxythymidylate synthesis (Winstanley *et al.*, 2001). The difficulty in synthesising atovaquone has led to a relatively high cost treatment and it is consequently rarely used in developing countries.

1.4.3.6 Sulphadoxine-Pyrimethamine antifolates (Fansidar)

The most commonly used antifolate drugs are sulphadoxine and pyrimethamine, which are given together as a cheap combination therapy (**Table 1.2, Figure 1.12**). The 2,4-diaminopyrimidine pyrimethamine, targets the dihydrofolate reductase enzyme (DHFR-TS), in a manner similar to that of proguanil (Ridley *et al.*, 2002 & Yuvaniyama *et al.*, 2003). Sulphadoxine inhibits the dihydroopteroate synthase (DHPS), a member of the folate pathway, in an as yet undetermined manner. Both of these are used together in a synergistic manner enhancing their activity and reducing the onset of resistance (Newton & White, 1999).

1.4.3.7 Antibiotics

Used in combination with antimalarials, antibiotics have been shown to help both inhibit parasite growth and slow down the onset of resistance. It is thought that their mode of

action is through inhibiting prokaryotic like pathways within the apicoplast, in particular self replication and protein synthesis (Ralph *et al.*, 2001). The combinations of quinine with tetracycline or doxycycline are commonly found in parts of south east Asia (Ridley, 2002).

1.4.3.8 Drug resistance

There are several factors which can influence the development of resistance including the drugs half life, patient compliance, administration, non-target drug pressure and cross resistance. This is coupled with the ability of both the *P.falciparum* and the *Anophelles* mosquito to undergo positive genetic mutations within drug pressure environments. Resistance to CQ is through selective mutations of two genes *pfcrt*, a membrane transporter and *pfmdr*1 a P-glycoprotein homologue (Wongsrichanalai *et al.*, 2002). Furthermore recent publications may suggest that CQ resistance is achieved or compounded through the acquisition of a CQ drug efflux pump (Sanchez *et al.*, 2003).

Point mutations of the DHFR have been shown to confer resistance to pyrimethamine, however the crystal structure of DHFR may offer insights into the development of new inhibitor analogues capable of overcoming these mutations (Yuvaniyama *et al.*, 2003). The time period between drug introduction and the first reported resistance can be very short (**Table 1.3**), emphasizing the need for drug combination therapies to lengthen the lifespan of new antimalarials. Although drug resistance can be found world wide (**Figure 1.13**), an area of great concern is Africa which has resistance to CQ and a loss of sulphadoxine-pyrimethamine sensitivity. This is compounded by a lack of economic wealth to afford the more expensive antimalarials (Wongsrichanalai *et al.*, 2002).

Antimalarial drug	Introduced	1st reported resistance	Difference (years)
Quinine	1632	1910	278
Chloroquine	1945	1957	12
Proguanil	1948	1949	1
Sulphadoxine-Pyrimethamine	1967	1967	0
Mefloquine	1977	1982	5
Atovaquone	1996	1996	0

Table 1.3. The time taken for resistance to occur for each of the main antimalarial drugs produced. The figure is based on data from Wongsrichanalai *et al.*, 2002.

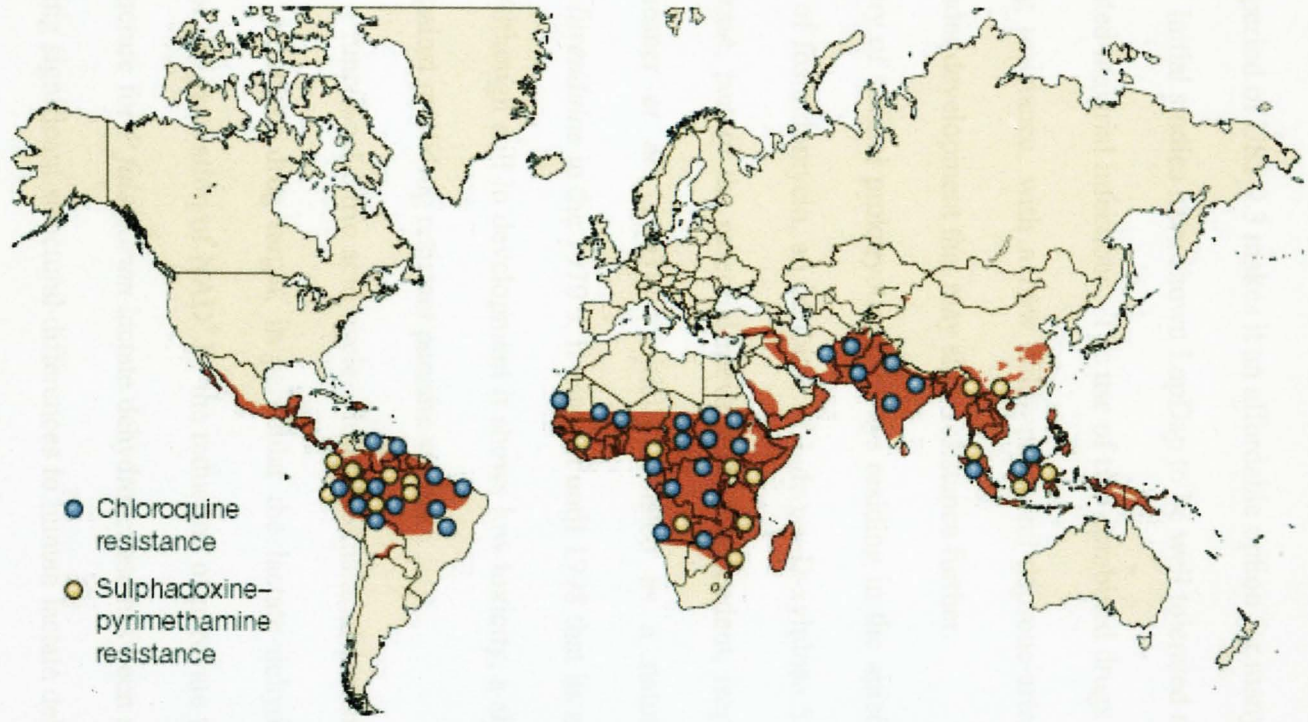


Figure 1.13. The global distribution of malaria in 2002 is represented through the red shading of the map, with red and yellow circles showing areas of parasite resistance to chloroquine and sulphadoxine-pyrimethamine, respectively. Figure is taken from Ridley *et al.,* 2002.

1.4.3.9 New and Developing drugs

Although new antimalarial drugs have been developed over the last two decades both the price of these drugs and the level of drug resistance have rapidly increased ((**Table 1.2**) (Molyneux *et al.*, 1999)). A promising solution may be in the development of Chlorproguanil-dapsone (LapDap) (Lang *et al.*, 2003) which may have an overall cost over a 3-day period of US$ 0.3 makes it an affordable option for many countries (Pincock *et al.*, 2003). Initial studies have shown LapDap to be well tolerated and effective against non complicated malarial infections. The use of two combined drugs may slow down the onset of drug resistance, with a new chloroproguanil-dapsone-artesunate combination drug being under development that may slow resistance further.

The discovery of several prokaryotic pathways residing in the apicoplast has led to the development of fosmidomycin, an inhibitor of 1-deoxy-D-xylulose 5-phosphate (DOXP) reductoisomerase, part of the essential mevalonate-independent, isoprenoid biosynthesis pathway (Wiesner *et al.*, 2003). Initially isolated as a natural antibiotic from *Streptomyces lavendulae* in the 1970's, it was not until 1998 that its molecular target was discovered. Although still in development it shows low toxicity, a short plasma half life and activity against multidrug resistant parasite strains.

The lack of a functional citric acid cycle within the malarial parasite makes anaerobic glycolosis an attractive drug target, in particular the lactate dehydrogenase, which is responsible for the generation of NAD^+ by the reduction of pyruvate to lactate. The high resolution structure for *P.falciparum* lactate dehydrogenase has been solved (Dunn *et al.*, 1996), revealing significant structural differences to human lactate dehydrogenase, which may be targeted as part of a rationally based drug design program. Currently the only promising lead inhibitor of lactate dehydrogenase is gossypol, which is isolated from

cotton seeds and displays a modest in vitro anti malarial activity (IC_{50}=15-29 μM) (Razakantoanina *et al.*, 2000).

1.5 *Toxoplasma gondii* pathology

Toxoplasma gondii is one of the most widespread infections of man and warm blooded animals, infecting an estimated 25% of the worlds human population (Sibley, 2003). The prevalence of the infection varies greatly on a geographic scale with an estimated 16-40% of the UK and US infected, whereas in continental Europe and Central and South America up to 80% may be infected. The main transmission route into humans is through the consumption of undercooked contaminated meat and is the third most common cause of food-borne deaths in the United States (Mead *et al.*, 1999). Although infections in humans are predominantly asymptomatic, the parasite can produce severe complications and may be fatal. There are two main risk groups for infection by *T.gondii*. Firstly, patients suffering from immune suppression, with approximately 30% of European and 10% of American patients suffering from Acquired Immunodeficency Syndrome (AIDS) dying from toxoplasmosis (Hill & Dubey, 2002). Furthermore patients undergoing transplantation surgery can also suffer from toxoplasmosis infections due to the immuno-suppressive drugs administered to reduce the risk of transplantation rejection (De Medeiros *et al.*, 2001). The second major risk area for a *T.gondii* infection is when it is passed from mother to neonate and is the leading cause of congenital neurological birth defects (Boyer & McLeod, 2002) with the health care burden within the USA, reportedly estimated to cost $5.2 billion per year (Roberts *et al.*, 1994).

Consideration must also be given to the bio-terrorism possibilities of this organism. Its genetics are relatively simple, as there are only two stages to its life cycle in which only the initial infection that is caused by rapidly multiplying tachyzoites can be targeted by

drugs. These turn into slower growing bradyzoites which reside in tissue cysts and can not be targeted by any current drug treatment, due to difficulties in crossing the bradyzoite cyst wall. By a theoretically simple process of genetic engineering it would be possible to place a more virulent toxin such as cholera or shigella under the control of a bradyzoite promoter. This might create a pathogen, which may target most mammalian cells and go unnoticed until established as a bradyzoite by which point no drugs would be available to inhibit parasite growth once it starts producing its more virulent toxins.

1.5.1 *Toxoplasma gondii* lifecycle

The *T.gondii parasite* can infect almost all warm blooded animals, however the definitive host is the *Felidae* family. In the intermediate host such as humans or livestock, the *T.gondii* undergoes two phases of asexual development in which tachyzoites develop into bradyzoites (**Figure 1.14**). Infection by the *T.gondii* parasite is commonly initiated through an oral route by the ingestion of undercooked contaminated meat containing tissue cysts or highly infectious oocysts from cat faeces (Carruthers, 2002). These oocysts once within the small intestine liberate sporozoites which quickly differentiate into trachyzoites and rapidly disseminate throughout the body. Ingestion of contaminated meat may result in the release of bradyzoites which like the oocysts quickly differentiate into trachyzoites which then migrate throughout the body. The tachyzoites, over the next 8-12 days multiply rapidly through repeated endodgeny in many different host cell types after which the tacyzoites develop into slower growing bradyzoites (Dubey *et al.*, 1998). The differentiation of these bradyzoite, which is thought to be triggered by the host immune system, marks the start of the chronic phase in which parasite replication is significantly slowed down and the parasite forms a cyst wall protecting itself from the host immune response (Bohne *et al.*, 1994). These tissue cysts have a preference for the

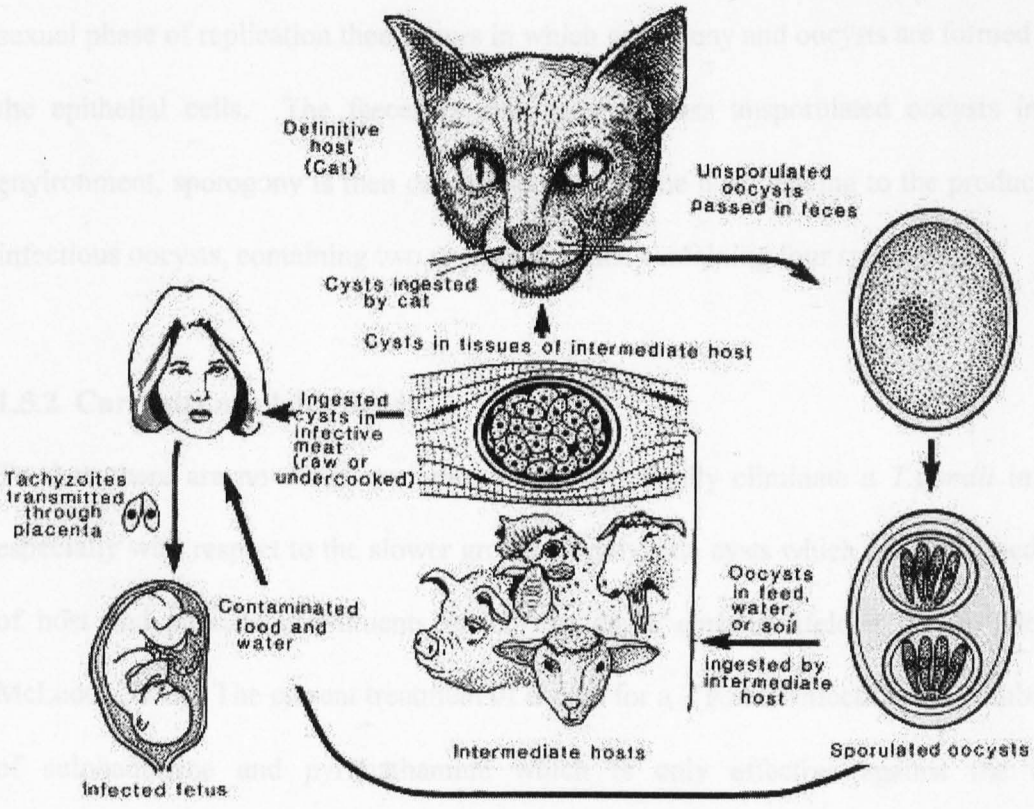


Figure 1.14. A schematic diagram of the *Toxoplasma gondii* life cycle, taken from Dubey *et al.*, 1998. Within this cycle, cats, the definitive hosts of *T.gondii*, can become infected by ingesting sporulated oocysts or (most often) infected animals. The oocysts are infectious to most mammals and birds. Toxoplasma can be transmitted to intermediate hosts through oocysts, by the ingestion of undercooked contaminated meat, or transplacentally.

central nervous system, the eye and skeletal muscle and may last for the life of the host (Tenter *et al.*, 2000). It is thought that these cysts may break down periodically, releasing tachyzoites which can re-infect tissue creating more tissue cysts (Dubey *et al.*, 1998).

If ingested by the definitive host, the *Felidae* family, another stage of asexual proliferation is carried out by the bradyzoite, consisting of initial multiplication followed by repeated endopolygeny in the small intestine epithelial cells (Tenter *et al.*, 2000). A sexual phase of replication then occurs in which gamogony and oocysts are formed within the epithelial cells. The faeces is then used to pass unsporulated oocysts into the environment, sporogony is then carried out outside the host, leading to the production of infectious oocysts, containing two sporocysts, each containing four sporozoites.

1.5.2 Currently available drugs

To date there are no drugs available which will totally eliminate a *T.gondii* infection especially with respect to the slower growing bradyzoite cysts which are composed partly of host and parasite constituents which can cause chronic lifelong illness (Boyer & McLeod, 2002). The current treatment of choice for a *T.gondii* infection is a combination of sulphadiazine and pyrimathamine which is only effective against the rapidly multiplying tachyzoite stage of growth. Furthermore these drugs are limited by their toxicity and allergy (Boyer & Mcleod, 2002). In difficult cases atovaquone, spiramycin, diaminodiphenylsulphone and clindamycin have been used to varying degrees of success.

1.6 Aims and Objectives

Recent studies have suggested that the apicoplast within apicomplexan parasites carries out fatty acid synthesis via the type II FAS system and subsequently could be targeted by key inhibitors of this pathway as described previously. The aim of this project is to

determine through crystallographic techniques the structure of the *P.falciparum* and *T.gondiii* ENR enzymes in complex with NAD^+ and triclosan in the hope of further understanding the ENR family.

CHAPTER 2

X-ray diffraction theory

This chapter describes the techniques used to prepare protein crystals and the principals of protein crystallography used in the structural analysis of the enoyl reductase from *P.falciparum* and *T.gondii*. The mathematical formulae for diffraction theory were obtained from Drenth, 1999 and Blow 2002.

2.1 Macromolecular crystallization

The crystallisation of protein molecules is carried out in a manner similar to that of small molecule crystallisation, with the aim being to remove the solvent molecules from around the protein inducing a low degree of supersaturation in solution. This forces the protein molecules together and encourages intermolecular contacts to form. Unlike small molecule crystals however, protein crystals typically contain large volumes of solvent (30-90%), which reduces the size of protein surface interactions. The majority of intermolecular bonds holding the crystal together tend to be multiple weak forces involving hydrophobic interactions, ion pairs and H-bonds, with the occasional small molecule salt, ligand or metal interaction. There are two stages to the growth of the protein crystal, the first being the initial nucleation stage in which a small number of protein molecules aggregate to form a nucleus whose intermolecular contacts resemble those of the final crystal. This occurs at higher supersaturation levels, whereas at lower supersaturation levels, the second stage, the growth of the crystal is supported with additional molecules joining the stable nuclei. In general slower supersaturation minimises the growth of large numbers of crystal nuclei and as such helps to encourage the growth of fewer larger and better quality crystals (**Figure 2.1**). The growth of protein crystals is halted when either the concentration of the remaining protein in solution is too low or when the surface of the protein crystals become badly formed and no longer allows for the formation of a regular crystal lattice.

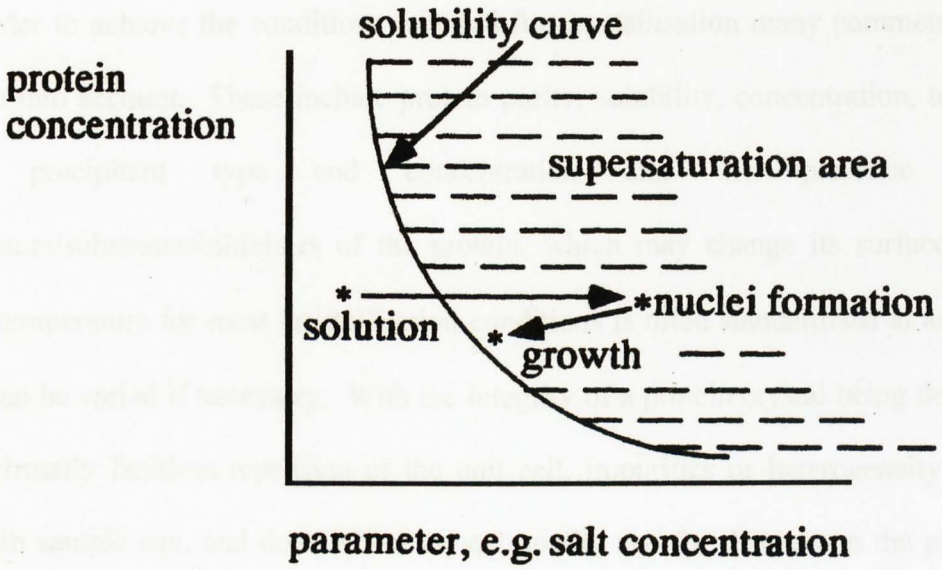


Figure 2.1. The theoretical solubility curve for a protein, as a function of salt concentration, adapted from Drenth, 1994.

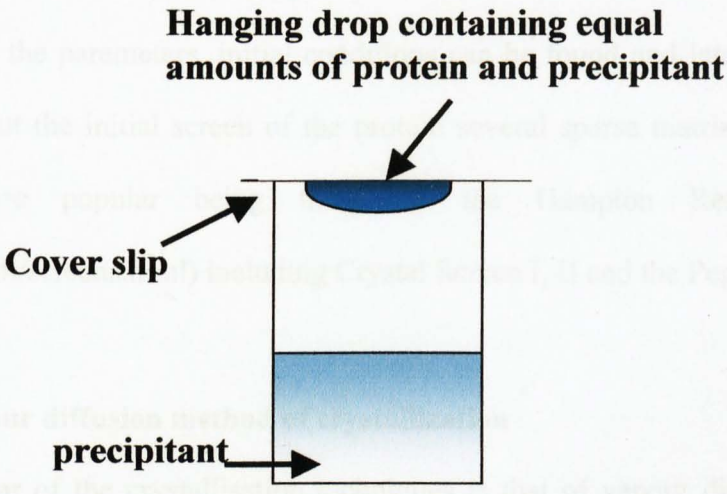


Figure 2.2 The hanging drop method of vapour diffusion crystallisation.

2.1.1 Crystallization experiments

In order to achieve the conditions required for crystallisation many parameters must be taken into account. These include protein purity, solubility, concentration, temperature, pH, precipitant type and concentration and the presence of both cofactors/substrates/inhibitors of the protein, which may change its surface character. The temperature for most crystallisation conditions is often standardised at around 17°C but can be varied if necessary. With the integrity of a protein crystal being dependant on the virtually faultless repetition of the unit cell, impurities or heterogeneity within the protein sample can, and does, affect crystal quality and for this reason the purity of the protein should generally be above 95%. The concentration of the protein generally ranges from 5-20mg/ml with the solubility and stability of a protein affecting the useable concentration range. The number of conditions that may be altered allows for limitless combinations. However by setting up a relatively small number of initial conditions ranging over all the parameters, initial conditions can be found and later optimised. In order to carry out the initial screen of the protein several sparse matrix screens can be used, the more popular being those of the Hampton Research Screens (www.nvxtal.com/screens.html) including Crystal Screen I, II and the PegIon screen.

2.1.2 The vapour diffusion method of crystallization

The most popular of the crystallisation techniques is that of vapour diffusion. In this method a small amount of purified and concentrated protein solution, typically between 2 and 5μl is placed either on a siliconised cover slip (hanging drop method) or on a micro bridge (sitting drop) in a closed environment with the crystallisation solution (**Figure 2.2**). The protein sample is commonly mixed, usually in a 1:1 volume ratio with the reservoir solution prior to sealing. The precipitant and protein concentration increases

gradually over time, through the dehydration driven reduction in droplet volume by equilibration of water vapour with the reservoir or well solution.

2.2 Crystal mounting

Due to the high levels of solvent within a protein crystal lattice any removal of the solvent through evaporation will destabilise the crystal. To overcome this the crystal must be kept within a stable environment either within the mother liquor or in the saturated vapour of the mother liquor. This requires the crystal to be either cryo cooled to 100K in a micro-fibre loop or placed within a sealed glass capillary tube containing mother liquor. This stabilised crystal is then placed within the X-ray beam and aligned using a goniometer head. The goniometer has two perpendicular arcs allowing the crystal rotation about two perpendicular axis permitting the alignment of the crystal within the X-ray beam.

2.2.1 Tube mounting

For tube mounted crystals the crystal is mounted within a glass capillary tube whose diameter is slightly larger than that of the crystal. Once within the tube a small amount of mother liquor is placed either side of the crystal forming a saturated vapour around the crystal, stabilising it. The tube is then sealed using wax. The walls of the capillary are very thin to reduce the amount of background X-ray scattering produced by the glass tube. Once positioned within the tube the protein crystal is stable and can be kept for short periods of time, (weeks) before crystal degradation (**Figure 2.3**).

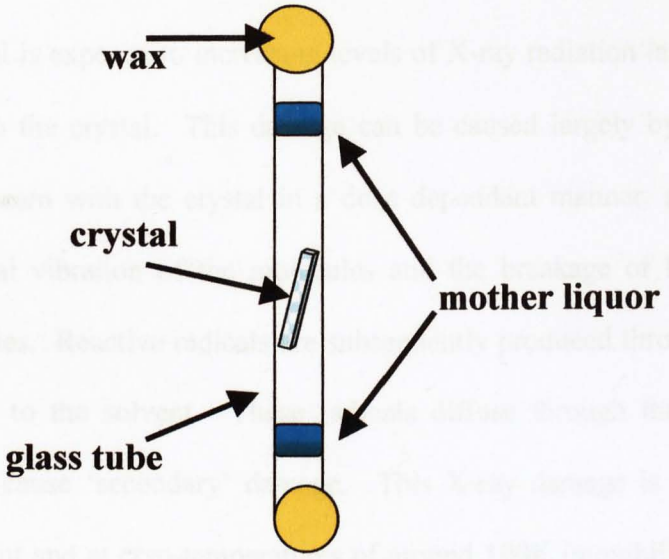


Figure 2.3. A schematic representation of a crystal mounted within a glass capillary tube.

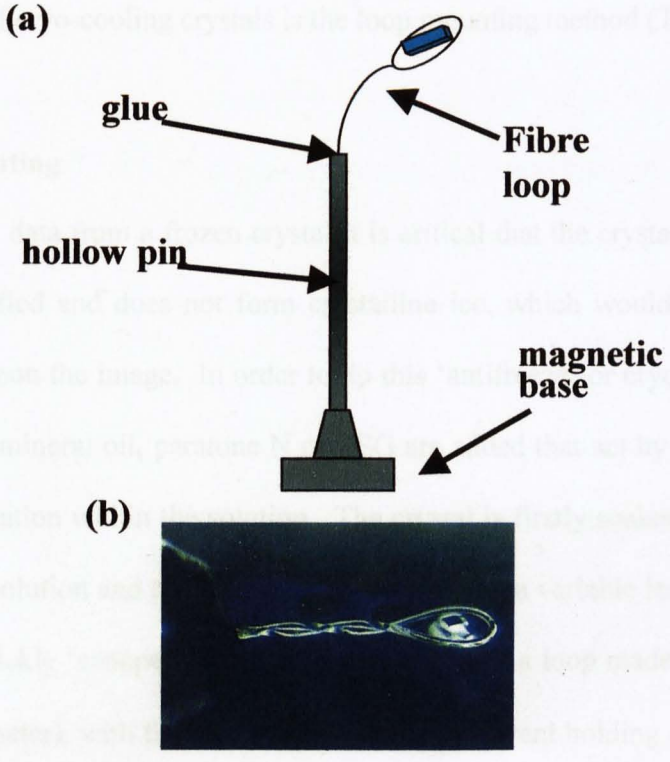


Figure 2.4. (a) A schematic representation of a loop mounted crystal. **(b)** A photograph of a loop mounted crystal taken from Mac Sweeney *et al.*, 2003.

2.2.2 Cryo cooling

As the protein crystal is exposed to increasing levels of X-ray radiation large amounts of damage can occur to the crystal. This damage can be caused largely by the 'primary' interactions of the beam with the crystal in a dose dependant manner, producing both heat, through thermal vibration of the molecules and the breakage of bonds between atoms in the molecules. Reactive radicals are subsequently produced through damage to the polypeptide and to the solvent. These radicals diffuse through the crystal using thermal energy and cause 'secondary' damage. This X-ray damage is both time and temperature dependent and at cryo-temperatures of around 100K immobilisation of these reactive products occurs. At room temperature however this 'secondary' damage diffuses through the crystal and can destroy the crystal. For this reason freezing the crystals at 100K in liquid nitrogen is the preferred method for crystal mounting. The most convenient way of cryo-cooling crystals is the loop mounting method (Teng, 1990).

2.2.3 Loop mounting

In order to obtain data from a frozen crystal it is critical that the crystal cools so fast that the water is vitrified and does not form crystalline ice, which would produce a strong diffraction ring upon the image. In order to do this 'antifreeze' or cryo protecting agents such as glycerol, mineral oil, paratone N or PEG are added that act by slowing down the speed of ice formation within the solution. The crystal is firstly soaked within a mixture of its stabilising solution and the cryo-protecting agent for a variable length of time. This crystal is then quickly 'scooped' out of the solution using a loop made up of a thin fibre (10-30μm in diameter), with the surface tension of the solvent holding it in place (**Figure 2.4 a, b**) and is then immediately plunged into liquid nitrogen. The speed at which this process occurs allows for a vitrified glass to form around the crystal holding it in place

without disrupting the crystal lattice, limiting the disruption to the diffraction properties of the crystal. Several cryo protecting agents are often tested until the optimum conditions can be found. Once frozen the crystals maybe stored in a liquid nitrogen Dewar for long periods of time (month). However frozen crystals can gain an increase in mosaicity due to small changes in the lattice upon freezing. Factors affecting this are, the speed of cooling and transfer to the cryogen and size of the crystal. It is important to limit this through a good technique and choice of cryo protecting agent so as to collect the best set of data possible from your crystal.

2.3 X-ray generation

X-rays are high frequency electromagnetic radiation having wavelengths ranging from 10^{-7} m (1000Å), (soft X-rays), to 10^{-11} m (1Å), (hard penetrating X-rays). For macromolecular crystallography the shorter wavelength X-rays, which are of a similar scale to the distance between atoms are used. X-ray radiation is produced when electrons, from a cathode of high negative potential are accelerated in a vacuum towards a positive anode. The electrons subsequently hit the anode at high speed and are lifted to a higher energy level and as they fall back down to a lower energy level they release energy. This energy is dissipated in the main as heat, however some takes the form of X-rays emitted in two ways, as a sharp peaks of specific wavelengths and as a broad band of wavelengths (**Figure 2.5**). It is the sharp peaks that are used within X-ray studies and are generated by the electron transitions within the metal anode. The usual source of a metal anode is copper and the corresponding CuKα and CuKβ X-ray wavelengths produced are generated by the transition of electrons from the L and M-shell, respectively to the K-shell.

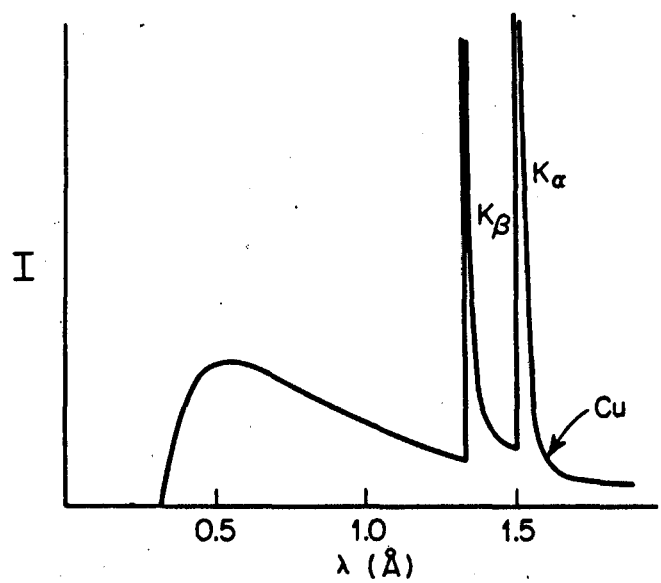


Figure 2.5 The characteristic X-ray emission spectra produced by a copper target, adapted from Drenth, 1994.

2.3.1 The rotating anode source

The most common source of X-rays within protein crystallography is through a rotating anode source where a heated tungsten filament bombards an anode usually constructed from a copper cylinder rotating at high speed with a beam of focussed electrons. The highly negative potential of the filament with respect to the anode and the presence of a vacuum allows the electrons to hit the anode at high speeds and produce X-rays in the manner described above. To select the desired wavelength of X-rays (the CuKα rays), a series of slits and a graphite crystal coated monochromator or nickel or platinum coated focussing mirrors combined with a nickel filter can be used. The anode source must be kept rotating at high speeds as one of the largest constraints upon an X-ray tube is its ability to dissipate heat and if the anode were stationary then the constant bombardment of electrons would concentrate large amounts of heat energy upon a small area of the anode causing it to melt.

2.3.2 Synchrotron radiation

The use of synchrotron radiation is now commonplace within protein crystallography and the ability to both change the wavelength and use a more intense X-ray beam makes it a powerful tool. Synchrotron radiation is produced by a particle accelerator, which has a large storage ring circulating electrons at velocities near to the speed of light. The electrons are circulated using a series of focussing magnets and any loss of energy is compensated by a radio frequency input, the constant change in direction by the electrons causes the emission of energy in the form of electromagnetic radiation. The intensity or flux of the beam on a synchrotron source containing bending magnets is two orders of magnitude greater than that of a conventional X-ray generator and furthermore, the use of a wiggler undulator is another two orders as intense again. This increase in beam

intensity allows for data collection on weakly diffracting or small crystals. For crystals with a large unit cell and subsequently small spacing within the diffraction pattern the low divergence of the synchrotron X-ray beam allows for high resolution data to be collected. The tunability of the wavelength allows multiple wavelength dispersion (MAD) phasing experiments to be carried out by tuning the wavelength to the absorption edge of an anomalous scattering heavy atom such as selenium. The increase in beam intensity also allows for much shorter exposure times on the crystal (seconds rather than minutes) and subsequently can reduce some of the secondary radiation damage and also shorten the time of the experiment.

2.4 Diffraction of X-rays

X-ray beams may be regarded as electromagnetic waves with an oscillating electric field. The interaction between an X-ray photon and an electron may set the electron oscillating about its nuclei with the same frequency as the X-ray. This produces a secondary scattered wave of the same frequency as the incident X-ray. Coherency is achieved because all the scattered waves from a source electron will have the same phase relationship to the incident X-ray beam, known as Rayleigh (elastic) scattering. However there is an incoherent or Compton (inelastic), scattering component, where the wavelength has changed due to exchange of energy and momentum when the X-ray hits the atom. As the coherent scattering component from many atoms in a crystal is significantly greater Compton scattering is generally ignored. Scattering of X-ray radiation by different electrons leads to interference as the relative phases add up, either in phase, resulting in an amplitude equal to the sum of the individual amplitudes or out of phase in which the amplitude is equal to the difference of the two amplitudes, or any amplitude between.

2.4.1 Bragg's law

The simple equation for Bragg's law

$$n\lambda = 2d \sin\theta$$

was derived in 1913 (Bragg, 1913) and can be used to explain why crystals appear to reflect X-ray beams at certain angles of incidence (θ). The wavelength of the incident X-ray beam is given as λ, d is the distance between the atomic layers in a crystal and n is an integer. The equation gives the conditions for which two scattered waves are in phase and interfere constructively, producing diffraction. The two incident beams are always in phase and parallel until the top beam strikes the top layer at atom z (**Figure 2.6a**) whereas the second beam is scattered by atom B in the second layer. In order for the two beams to continue to travel in phase the extra distance traveled (AB + BC) must be an integral (n) multiple of the wavelength (λ) such that

$$n\lambda = AB + BC$$

By taking d as the hypotenuse of the right angled triangle ABz (**Figure 2.6b**), simple trigonometry can be used to relate d and θ to the distance (AB + BC). The distance AB is opposite θ so,

$$AB = d \sin\theta$$

The equation $n\lambda = AB + BC$ can be rewritten as $n\lambda = 2AB$ as AB = BC. Combining this equation to the equation above gives us Bragg's law: $n\lambda = 2d \sin\theta$.

2.4.2 Structure factor for a single electron

The structure factor is a complex number which represents the amplitude and phase of the resulting diffracted wave. An absolute scale is used for defining the structure factor by defining the diffraction from a single electron as having an amplitude of 1e:

$$\mathbf{F(S) = 1}e\ e^{(2\pi i \mathbf{r}\cdot\mathbf{S})}$$

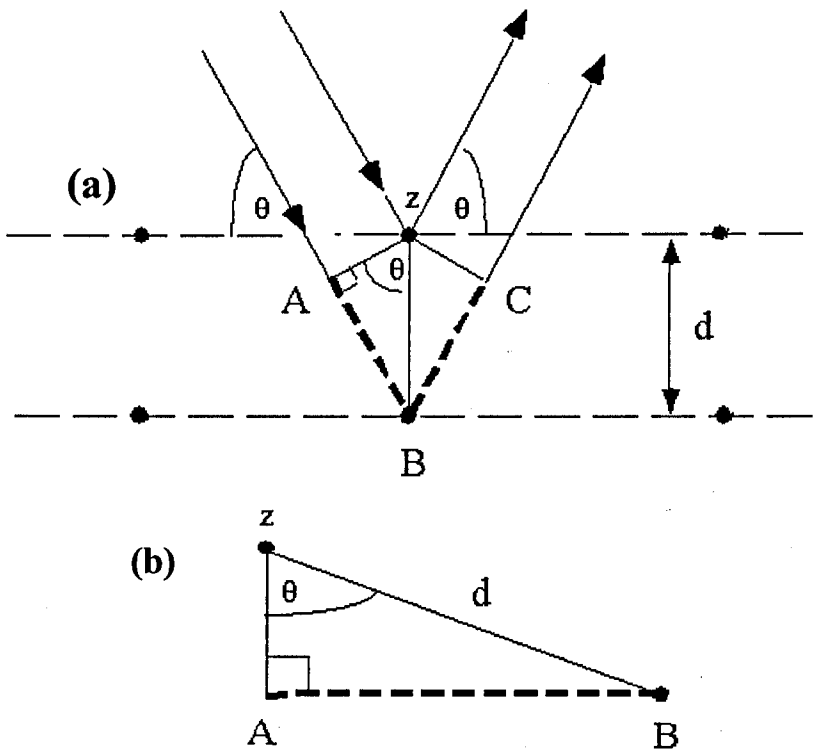


Figure 2.6. (a), (b) Schematic description of diffraction from planes obeying Bragg's law.

where **S** is the vector perpendicular to the reflecting plane and **r** is the vector representing the position of the electron relative to the origin. As such the diffracted wave has a phase of $2\pi i\mathbf{r}\cdot\mathbf{S}$, relative to the origin.

2.4.3 Scattering by an atom

As it is the electron cloud of an atom which scatters X-rays, the scattering is dependent on the nature of the atoms. The atomic scattering factor f is the sum of the entire scattering over the volume of the atom:

$$f = \int_{\mathbf{r}} \rho(\mathbf{r})\, e^{[2\pi i\mathbf{r}\cdot\mathbf{S}]} d\mathbf{r}$$

where integration is over the whole space **r**, the electron density at position **r** is $\rho(\mathbf{r})$, **r** is the position of the electron from the centre of an atom (origin) and **S** is the scattering vector.

2.4.4 Scattering by a unit cell

The unit cell has n atoms at positions $\mathbf{r}_j$ (j = 1, 2, 3, 4, ..., n) with respect to the unit cell origin all of which will make a contribution to the total unit cell scattering. Therefore, the diffraction from the unit cell is the total sum of the scatter from the individual atoms and is defined as:

$$\mathbf{F(S)} = \sum_{j=1}^{n} f_j \exp[2\pi i\mathbf{r}_j \cdot \mathbf{S}]$$

The structure factor **F(S)**, so called because it depends on the arrangement of atoms in the unit cell and describes the distribution of electron density within the unit cell.

2.4.5 Scattering by a Crystal

As a crystal is comprised of many unit cells, the scattering from a crystal can be described as the sum of all the unit cells with respect to a single origin.

$$\mathbf{K(S) = F(S)} \times \sum_{t=0}^{n1} \mathrm{e}^{[2\pi it\mathbf{a}\,.\,\mathbf{S}]} \times \sum_{u=0}^{n2} \mathrm{e}^{[2\pi iu\mathbf{b}\,.\,\mathbf{S}]} \times \sum_{v=0}^{n3} \mathrm{e}^{[2\pi iv\mathbf{c}\,.\,\mathbf{S}]}$$

The crystal contains translation vectors in the direction a, b and c, where n_1, n_2 and n_3 are the number of unit cells in the direction a, b and c, respectively. The indices t, u and v are the translatory steps which relate each of the unit cells along each vector. As there are many unit cells in the crystal, the values of n_1, n_2 and n_3 are very high which means that the summation $\sum_{t=0}^{n1} \mathrm{e}^{[2\pi it\mathbf{a}\,.\,\mathbf{S}]}$ and the other two summations over u and v are almost equal to zero unless $\mathbf{a}\,.\,\mathbf{S} = h$, $\mathbf{b}\,.\,\mathbf{S} = k$ and $\mathbf{c}\,.\,\mathbf{S} = l$, the Laue conditions. Therefore, the total diffraction from a crystal will consist of small well orientated maxima and each maxima is indexed with the hkl indices.

2.4.6 The reciprocal lattice and the Ewald sphere

Protein crystals are comprised of rows of ordered molecules arranged in a lattice with its subsequent diffraction pattern also being seen as a lattice. However the diffraction lattice is the reciprocal of the real lattice and as such has axis dimensions a*, b* and c*, which are inversely proportional to the real crystal lattice dimensions a, b and c. In order for a diffracted beam to be at a maxima it must fulfill the specific conditions defined by the Ewald sphere (Ewald, 1979). The Ewald sphere has a radius of $1/\lambda$, and in order for a reflection to occur the reciprocal lattice point must lie on the surface of the sphere (**Figure 2.7**). By rotating the crystal in the X-ray beam, all lattice points will pass through the surface of the sphere and as such be measured. Some reflections will not be measured due to the occurrence of blind regions, in which, due to the alignment of the crystal, some reflections will lie on the rotation axis and never lie on the Ewald sphere.

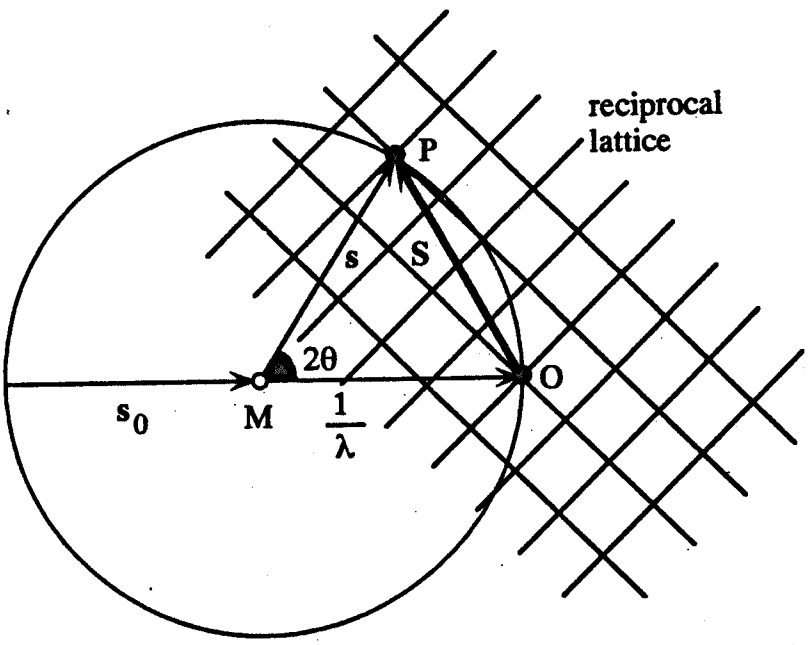


Figure 2.7 The Ewald sphere. With radius $1/\lambda$, the lattice centre is at O. $\mathbf{s}_0$ indicates the direction of the incident beam; **s** indicates the direction of the scattered beam. The figure is taken from Drenth, 1999.

2.5 Calculation of electron density

If x, y and z are fractional coordinates in the unit cell where h, k and l are the indices for a reflection in the reciprocal lattice and V is the volume of the cell, then the structure factor **F (S)** can be rewritten as **F**(hkl).

$$\mathbf{F}(hkl) = V \int_{x=0}^{1} \int_{y=0}^{1} \int_{z=0}^{1} \rho(xyz) \exp\left[2\pi i(hx + ky + lz)\right] dx\, dy\, dz$$

The electron density for every position x, y, z, in the unit cell $\rho(xyz)$, is given by the Fourier transform of **F**(hkl) and can be written as:

$$\rho(xyz) = \frac{1}{V} \sum_{h} \sum_{k} \sum_{l} \mathbf{F}(hkl) \exp\left[-2\pi i(hx + ky + lz)\right]$$

F(hkl) is a wave that has an amplitude component $|\mathbf{F}(hkl)|$ and a phase component $\alpha(hkl)$ such that:

$$\mathbf{F}(hkl) = |\mathbf{F}(hkl)|\, e^{i\alpha(hkl)}$$

The $|\mathbf{F}(hkl)|$ component of this equation can derived experimentally as the intensity of the reflections within the diffraction pattern is proportional to the square of its structure factor amplitude. However as well as the measured amplitudes the phase information must also be calculated and this cannot be gained directly from the experimental data set collected. This is commonly referred to as the phase problem within protein crystallography and can be overcome by a variety of techniques such as isomorphous replacement, anomalous scattering, MAD, direct methods and molecular replacement.

2.6 Data collection

The most widely used devices within protein crystallography for measuring diffraction data are image plate detectors and CCD cameras.

2.6.1 Image plate detectors

A typical image plate is constructed using a thin layer of barium halide phosphor of BaFBr doped with europium, which is excited into a meta-stable electron state by incident X-rays (**Figure 2.8**). The image is subsequently read by a fine He-Ne laser beam (λ=633nm) scanning plate, which through stimulation of the meta-stable europium atoms can read the violet luminescence (λ=390nm) emitted. Importantly the intensity of violet luminescence is proportional to the number of absorbed X-rays, giving a system which offers a uniform sensitivity across the detector face with well characterised spatial distortion and a high signal to noise ratio. The image plate is subsequently erased by the exposure to intense light (optimally yellow), which returns the BaFBr back to its unexposed state.

2.6.2 CCD cameras

A CCD camera (charged coupled device) which utilises a phosphor screen to convert the X-rays into visible light photons is an alternative to an image plate and has the advantage of having a faster readout time. In a CCD the light image is de-magnified down to the size of a CCD chip using a fibre-optic taper converting the light image to an electric charge image which can then be digitised (**Figure 2.9**).

2.7 Data processing with DENZO and SCALEPACK

Having collected diffraction data each image must be correctly indexed, scale and merged which can be done using DENZO and SCALEPACK (Otwinowski *et al.*, 1997). The reduction and analysis of a single crystal diffraction data set can be broken down into the seven major steps laid out in the HKL manual version five (Gewirth, 1999). These are:

1. Visualisation and preliminary analysis of the original, unprocessed diffraction pattern

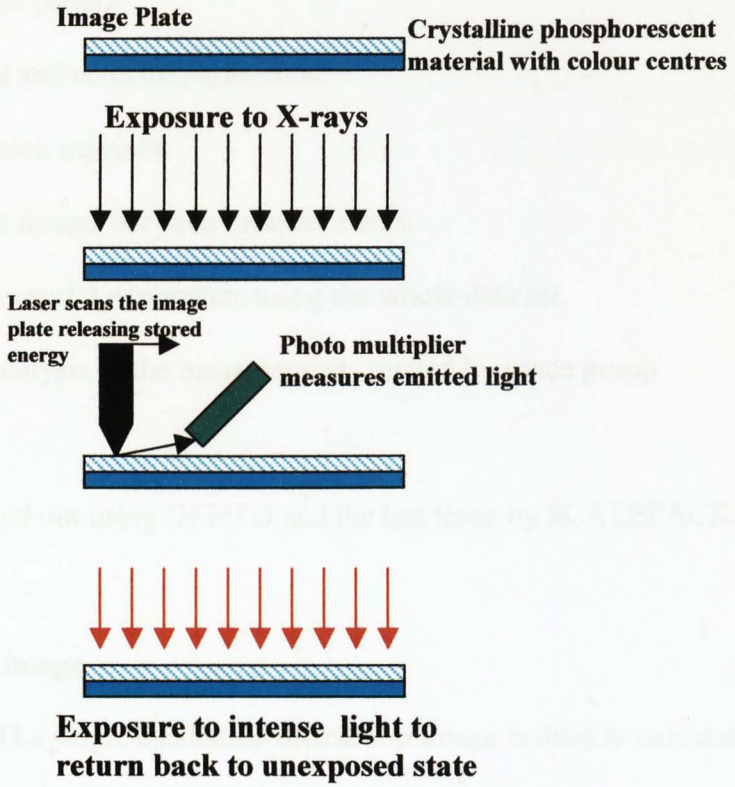


Figure 2.8 Diagrammatic representation to show the principle of a MAR image plate on an X-ray detector. The figure is adapted from Drenth, 1999.

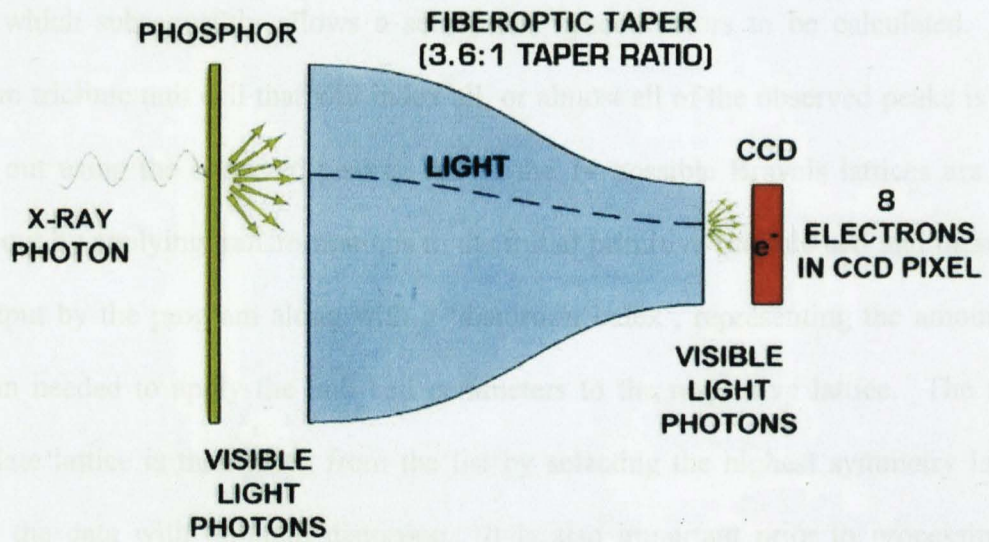


Figure 2.9 A diagrammatic representation of a CCDQ4 camera. Figure is taken from the ADSC home page (www.adsc-x-ray.com/products.html).

2. Indexing of the diffraction pattern.
3. Refinement of the crystal and detector parameters.
4. Integration of the diffraction maxima.
5. Finding the relative scale factors between measurements.
6. Precise refinement of the crystal parameters using the whole data set.
7. Merging and statistical analysis of the measurements related by space group symmetry.

The first four steps are carried out using DENZO and the last three by SCALEPACK.

2.7.1 Auto indexing of an image

For auto indexing in DENZO a single oscillation diffraction image is used to calculate the lattice type, crystal orientation and crystal unit cell parameters. The first step is to carry out a peak search in which reflections above a minimum intensity are located, the default being 10σ. These peaks are then used to carry out a complete search of all possible indices which subsequently allows a set of real space vectors to be calculated. The minimum triclinic unit cell that will index all, or almost all of the observed peaks is then worked out using the observed peaks. All of the 14 possible Bravais lattices are then worked out by applying transformations to the initial primitive triclinic cell and these are then output by the program along with a 'distortion index', representing the amount of distortion needed to apply the unit cell parameters to the respective lattice. The most appropriate lattice is then taken from the list by selecting the highest symmetry lattice that fits the data with minimal distortion. It is also important prior to processing to manually inspect the raw data to look for any abnormalities within the diffraction pattern which would be too complex to be analysed numerically by a computer. These may include the presence of two lattices within the diffraction pattern brought about by the

presence of a satellite crystal. Also spot smearing may result in the incorrect indexing of the image and lead to the incorrect positioning of reflections. Furthermore it is important when carrying out experiment on crystals frozen at 100K, to check that no ice rings are present on the image.

2.7.2 Image refinement and prediction of the diffraction maxima

The precise position and index of each reflection must be known before integration in DENZO, with information obtained from the strong reflections being used to predict the position of weak reflections. High accuracy prediction of the weak reflections depends on the precise knowledge of detector and crystal orientation parameters. A more accurate estimate of these parameters is calculated using a least squares method combined with partiality refinement (the intensity of partially recorded reflections is compared to the predicted partiality multiplied by an average intensity in the same resolution range). Once an accurate index has been established the intensity of the diffraction maximum is integrated by profile fitting. In profile fitting a curve, derived by averaging over the most intense reflections, is fitted to the data and the volume under the curve is taken to be the intensity. For each image a .x file is output that contains a header containing the initial parameters, a list of reflections and finally a list of the final DENZO parameters as refined from the image.

2.8 Data scaling, merging and post refinement

The program SCALEPACK is used to complement DENZO and carries out the scaling and merging of different data sets and also the global refinement of crystal parameters (post refinement). The unit cell parameters refined from a single image may at this stage still be imprecise owing to the correlation between the crystal and detector parameters.

More accurate estimations of unit cell lengths and angles, the orientation of each image and crystal mosaicity may be obtained from global refinement by making use of the information within the whole data set, resulting in a scaled and merged data set. The output statistics from SCALEPACK can be used to evaluate the quality of the resulting data. These statistics include the ratio of the intensity over the standard deviation in intensity ($I/\sigma I$) and on the agreement between symmetry related reflections (Rmerge), which can both be used to estimate the resolution to which the crystal diffracts and the χ^2 (likelihood) which gives a goodness-of-fit of the data to the error model. The resulting output from SCALEPACK is in the form of a .sca file which contains the reflections hkl position and their accompanying I and σI values.

2.9 The Patterson function

In the absence of any phase information it is still possible to obtain some information on the structure using the intensity information alone. The Patterson function is a Fourier summation with intensities as coefficients and with all phase angles equal to zero (Patterson, 1934). It can be expressed by the equation:

$$P(u\ v\ w) = \frac{1}{V} \sum_{hkl} \left| \mathbf{F}(hkl) \right|^2 \cos[2\pi(hu + kv + lw)]$$

which can alternatively be written in a more physically interpretable definition in terms of electron density at real space co-ordinates (x, y, z) as:-

$$\$(u\ v\ w) = \int \rho(x\ y\ z)\, \rho(x + u,\ y + v,\ z + w)\, dv$$

where u, v and w are co-ordinates in the unit cell and are so called to avoid confusion with x, y and z in the real cells. A Patterson function thus represents a density map of the vectors between scattering objects in the cell. The vector between the origin and any non-origin peak in the Patterson map represents the vector between two atoms in the real

cell. The Patterson map for a protein is impossible to solve given that for any N atoms within a real unit cell there will be N^2 peaks, because N vectors can be drawn from each atom. Of the N^2 peaks N will be situated at the origin because they represent vectors between an atom and itself. Therefore the actual number of non-origin peaks is $N(N\text{-}1)$. Important points within a Patterson map are that:

1. A peak is produced within the Patterson map for each pair of atoms.
2. Patterson maps are centro-symmetric.
3. The height of a peak within the Patterson map is directly proportional to the product of the atomic numbers of the two atoms represented by the peak.

2.10 Molecular replacement

One common way to solve the phase problem is to carry out molecular replacement in which a known 3D structure, which is predicted to be similar to the experimental model can be used to give an estimate of the phases. The known structure, or search model must have a reasonable sequence identity of around 25% or greater to the target molecule. It is assumed that a high sequence identity will result in a similar fold and as such a better estimation of the initial phases. In order to obtain the initial phases the search model must be reoriented such that it is in the same position as the target molecule within the unit cell. To do this the search model must be rotated and then translated. Molecular replacement is thus carried out in two steps, first the rotation, then the translation to minimise the computing power required.

2.10.1 The cross rotation function

The cross rotation function allows for the orientation of the search model that produces the maximal overlap onto that of the target structure. To achieve this a Patterson function

of the search model is calculated and compared to that calculated from data collected on crystals of the target molecule. This exploits the fact that the Patterson function is dependent on only the orientation and not its position within a unit cell. It is desirable that when calculating a rotation function, a radius of integration is carefully chosen so that cross-vectors, (intermolecular vectors), arising from neighboring molecules in the Patterson are excluded as the rotation function should be dominated by self vectors (intramolecular vectors). For this reason a sphere of integration of approximately 75% of the radius of the search molecule is commonly used. The resolution limits used can also have a large influence on the quality of the rotation function. As the resolution increases the model similarity will decrease, such that resolution ranges of 6Å to 3Å are commonly used. The presence of more than one subunit in the asymmetric unit will result in a corresponding number of solutions, such that a tetramer will reveal four solutions relating in symmetry to each of the four subunits. Problems may arise if more than one molecule is present within the asymmetric unit as the solution to one of the molecules will only account for a portion of the scattering matter of the crystal and as such would give a lower signal to noise ratio than a solution accounting for almost all of the crystal scattering matter. The rotation of a molecule with respect to another in a rotation function is often expressed by the angles κ, ϕ and ω which represent the polar angles (**Figure 2.10**) with rotation κ being about an axis whose direction is defined by the spherical polar angles ϕ and ω.

2.10.2 The translation function

Once a solution to the rotation function has been found it is possible to carry out the translation function. In contrast to the rotation function the translation function as described by Crowther and Blow (Crowther & Blow, 1967) is dependent on the cross-

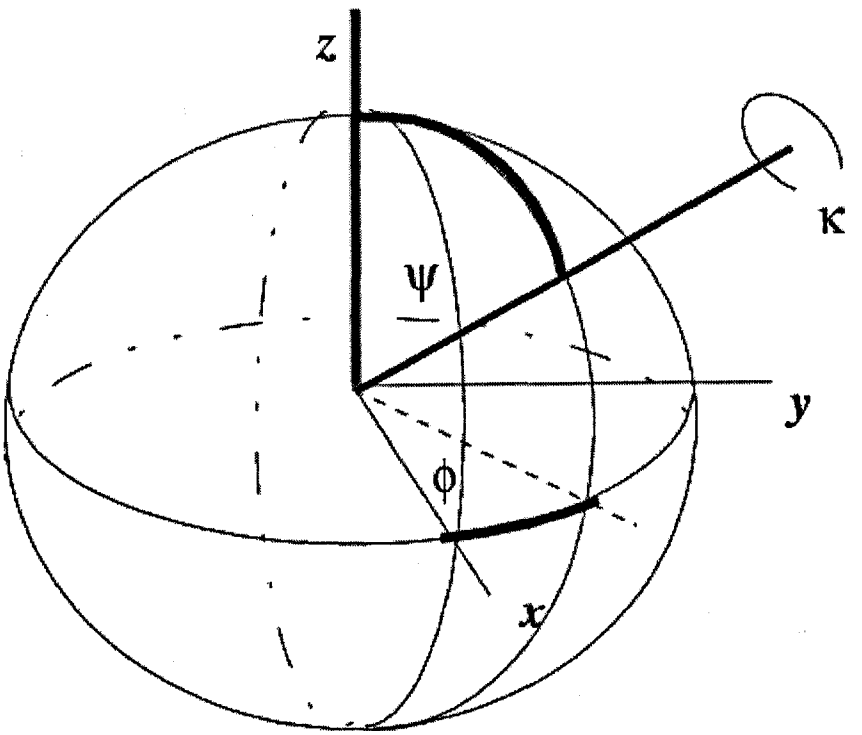


Figure 2.10. A diagrammatic view of the spherical polar angles ω, ϕ and κ used for the rotation function in molecular replacement www-strutted.corm.cam.ac.auk/course.html

vectors (intermolecular) from the Patterson, which are between atoms in adjacent asymmetric units. The intermolecular vectors are calculated by removing the intramolecular vectors calculated from the Patterson. The full symmetry translation function, T0 is two dimensional and is calculated for every pair of symmetry related molecules.

$$T0(\mathbf{t}) = \int_V P_o(\mathbf{u}) \cdot P_c(\mathbf{u},\mathbf{t})\, d\mathbf{u}$$

where $P_o(\mathbf{u})$ is the observed Patterson function, $\mathbf{u}$ is the vector in Patterson space and $\mathbf{t}$ is the translation for the identity asymmetric unit. The function $T(\mathbf{t})$ reaches a maximum when $\mathbf{t}$ is equal to the true translation vector $\mathbf{t}_o$ because the computed search model Patterson $P_c(\mathbf{u},\mathbf{t})$, fits the observed Patterson function $P_o(\mathbf{u})$ correctly.

An improved translation function, the T2 function, has been derived by Crowther and Blow (Crowther and Blow, 1967), which is a three-dimensional function which utilises all the symmetry operators simultaneously. The T2 function can be expressed by subtracting the intramolecular vectors.

$$T_2(\mathbf{t}) = \int_V (P_o(\mathbf{u}) - \sum_l P_{II}(\mathbf{u}))\,(P_c(\mathbf{u},\mathbf{t}) - \sum_l P_{II}(\mathbf{u}))\, d\mathbf{u}$$

Where P_{II} is the calculated intramolecular vector for each subunit of the search model.

A possible solution to the translation function can be checked by examining the steric clashes between molecules in the cell. This is done during the translation search by the program Molrep (Vagin & Teplyakov 2000), whereas the program Amore (Navaza, 1994) only calculates this once a solution has been obtained.

2.11 Phase improvement

The initial phases obtained for the protein through molecular replacement often contain large errors, resulting in a poor electron density map. In order to improve the map quality and remove errors a number of procedures may be used.

2.11.1 Solvent flattening

Within a protein crystal the protein is surrounded by both ordered and disordered solvent molecules. The ordered solvent molecules are hydrogen bonded to the protein and as such their electron density forms discrete features in the map. The disordered solvent however forms a flat featureless continuum beyond this. In addition the phase errors associated with initial phase sets can often contribute false peaks into the disordered solvent. To reduce the map noise caused by this, solvent flattening aims to change the phases so as to flatten out the electron density of the disordered solvent to a constant correct value. This results in improved phases for the electron density in the protein region (Wang, 1985). A molecular boundary is defined by dividing the unit cell into a grid with the density at each grid point being replaced by a new value, being proportional to the weighted sum of the density of a sphere of radius, R, from the grid point j. The "flattened" map can then be used to calculate new structure factor amplitudes and phases which are combined with the initial phases to produce a new map, which is modified and back transformed. This method has proved successful for phase refinement especially for crystals with a high solvent content however it is poor for phase extension (Wang, 1985).

2.11.2 Molecular NCS averaging

This technique utilises the presence of non-crystallographic symmetry (NCS) common to many protein crystals to improve the phases. It is assumed that differences between

molecules within the asymmetric unit are negligible and therefore that the density for the various subunits can be mapped onto each other and averaged. This process also involves flattening the solvent. The asymmetric unit is then reformed and phase angles for the new model can be calculated by back transforming the new map. The obtained phase information is then combined with the previous phases, with the 'best' phases being used to calculate the next electron density map.

2.11.3 Histogram matching

Although not as powerful a technique as solvent flattening or molecular averaging an important technique in phase improvement is histogram matching (Zhang & Main, 1990). For this the electron density distribution of the map is plotted against frequency and compared with that of a standard map at the same resolution. These two graphs are subsequently divided into bins containing an equal number of density points. Deviations from the predicted histogram result in incorrect phasing and tend to give a broader histogram than the true one. By reassigning the electron density to the predicted distribution of density, a modified map is produced which should be a truer representation of the structure.

2.12 Structure refinement

The aim of refinement is to increase the agreement between the calculated structure factor amplitudes and the observed structure factor amplitudes. The map used for the initial model building will contain errors due to inaccurate measurements and poor estimates of the phases. Refinement is improve the estimated phases and subsequently improve the quality of the electron density map. The main drawback in refinement is the low observation to parameter ratio because for each atom in the protein 4 parameters (x, y, z,

and the B-factor) are commonly refined. In order to decrease the number of parameters or increase the number of 'observations' certain restraints can be placed upon the model. These include restraints on bond lengths and angles, van der Waals contact distances, planarity of groups and maintenance of the chirality of chiral groups. Further restraints can be placed on the model based on the non-crystallographic symmetry which may exist between molecules within the asymmetric unit cell, allowing the application of symmetry constraints. Parameters may further be reduced by describing the structure as a set of rigid bodies

Crystallographic refinement can be described as the search for a global minimum of a target:

$$E = E\text{chem} + \omega_{\text{x-ray}} E_{\text{x-ray}}$$

where E is a function of the parameters of the atomic model, Echem contains the empirical information about chemical interactions, describing covalent (bond lengths, bond angles, torsion angles, chiral centres and planarity of aromatic rings) and non-bonded interactions. A measure of the differences between the observed and calculated data is given as $E_{\text{x-ray}}$ and $\omega_{\text{x-ray}}$ is a weighting term.

2.12.1 Maximum likelihood refinement

The method of maximum likelihood refinement as implemented in the program REFMAC5 (Murshdov *et al.*, 1997) is based on the likelihood of a model given estimates of the errors in the measured intensities as well as the model (Bricogne, 1997). Improvements to the model to make the observations more probable will increase the likelihood of the model being correct. Errors in the model are provided by the σ_A distribution, which is the square root of the correlation coefficient between the normalised observed and calculated structure factors from the model. As improvements

to the model are made the σ_A distribution changes and new estimations are made. Refinement that improves the model structure factors, will decrease the error estimates and this will improve the next round of refinement. In the work described in this thesis refinement was carried out by the program REFMAC (Murshudov *et al.*, 1997).

2.12.2 Rigid body refinement

The first stage of refinement is rigid body refinement in which segments of the protein are treated as a rigid unit in a manner specified by the user. Rigid body refinement typically uses low resolution data and allows for a more accurate placement of a model obtained by molecular replacement into the electron density.

2.12.3 Monitoring the progress of refinement

In order to monitor "improvement" within a crystal structure the agreement between the structure factor amplitudes calculated from the model with the observed structure factor amplitudes, the R-factor is calculated. In order to calculate the R-factor the following equation is used;

$$R_{fact} = \frac{\sum_{hkl} \| F_o | - | F_c \|}{\sum_{hkl} | F_o |}$$

when the Fobs and Fcalcs are in very good agreement then the R-factor will be low and the model is a good likeness to the protein structure. A desirable R-factor is approximately 10% of the model resolution such that for a 2Å model you would expect an R-factor of between 0.15 and 0.2. To eliminate bias and over fitting of a model the free R-factor (R_{free}) was introduced (Brunger, 1997), in which a randomly selected section of data usually 5-10%, is removed from the refinement to form a test data set, T. R-factors are produced for the test set using the following equation.

$$R_{free} = \frac{\sum_{hkl \subset T} \left| |F_o| - |F_c| \right|}{\sum_{hkl \subset T} |F_o|}$$

The value obtained for the R-free should not be higher than more than 5% of the R-factor. The R_{free} is a good indicator as to whether changes made to the model to lower the R-factor such as the addition of waters or NCS restraints are doing so through over-fitting of the data. As another check of model quality, the program PROCHECK (Laskowski *et al.*, 1993) is used to calculate a Ramachandran plot (Ramachandran, 1963) for the model and to check model stereochemistry to reveal areas for possible model improvement.

3.1 Introduction

This chapter describes the identification, cloning, overexpression, purification, crystallisation and structure determination by molecular replacement of *P.falciparum* enoyl reductase (PfENR) complexed with NAD^+ and triclosan.

3.2 Identification of the *P.falciparum* ENR gene

The identification of the PfENR sequence was carried out by conducting a BLAST search on the amino acid sequence data within the *P.falciparum* genome database, PlasmoDB, (www.PlasmoDB.org) against the *E.coli* and *B.napus* ENR sequences. Comparison of the PfENR sequence revealed approximately 44% and 18% sequence identity to the ENRs of *B.napus* and *E.coli*, respectively, consistent with the apicoplast being of an algal not bacterial origin (**Figure 3.1**). The gene for PfENR was found to be located on chromosome "BLOB", which was later assigned to chromosome 6 (Asn no. AF338731). Analysis of the pattern of sequence conservation within the region of the triclosan binding site showed that of the 11 residues shown to be within 4Å of the triclosan inhibitor within the *E.coli* and *B.napus* ENR structures 8 and 10 residues, respectively are conserved in the *P.falciparum* sequence (**Figure 3.2**). Consistent with PfENR being of an apicoplast origin, a large bipartite N-terminal signal sequence was identified, using the program PlasmoAP (Foth *et al.*, 2003) preceding the mature PfENR gene. A striking feature of the sequence alignment is the presence of a large, polar, low complexity insert of 42 amino acids within the PfENR sequence that is not mirrored in any other of the reported ENR sequences. The only possible exception to this is in MtENR, where a small 10 residue glycine-rich hydrophobic insert can be identified at the equivalent position. Within the MtENR structure this insert has been implicated in allowing the

CHAPTER 3

Structure determination of the *Plasmodium falciparum* ENR/NAD^+/triclosan complex

3.1 Introduction

This chapter describes the identification, cloning, overexpression, purification, crystallisation and structure determination by molecular replacement of *P.falciparum* enoyl reductase (PfENR) complexed with NAD^+ and triclosan.

3.2 Identification of the *P.falciparum* ENR gene

The identification of the PfENR sequence was carried out by conducting a BLAST search on the amino acid sequence data within the *P.falciparum* genome database, PlasmoDB, (www.PlasmoDB.org) against the *E.coli* and *B.napus* ENR sequences. Comparison of the PfENR sequence revealed approximately 44% and 18% sequence identity to the ENRs of *B.napus* and *E.coli*, respectively, consistent with the apicoplast being of an algal not bacterial origin (**Figure 3.1**). The gene for PfENR was found to be located on chromosome "BLOB", which was later assigned to chromosome 6 (Asn no. AF338731). Analysis of the pattern of sequence conservation within the region of the triclosan binding site showed that of the 11 residues shown to be within 4Å of the triclosan inhibitor within the *E.coli* and *B.napus* ENR structures 8 and 10 residues, respectively are conserved in the *P.falciparum* sequence (**Figure 3.2**). Consistent with PfENR being of an apicoplast origin, a large bipartite N-terminal signal sequence was identified, using the program PlasmoAP (Foth *et al.*, 2003) preceding the mature PfENR gene. A striking feature of the sequence alignment is the presence of a large, polar, low complexity insert of 42 amino acids within the PfENR sequence that is not mirrored in any other of the reported ENR sequences. The only possible exception to this is in MtENR, where a small 10 residue glycine rich hydrophobic insert can be identified at the equivalent position. Within the MtENR structure this insert has been implicated in allowing the

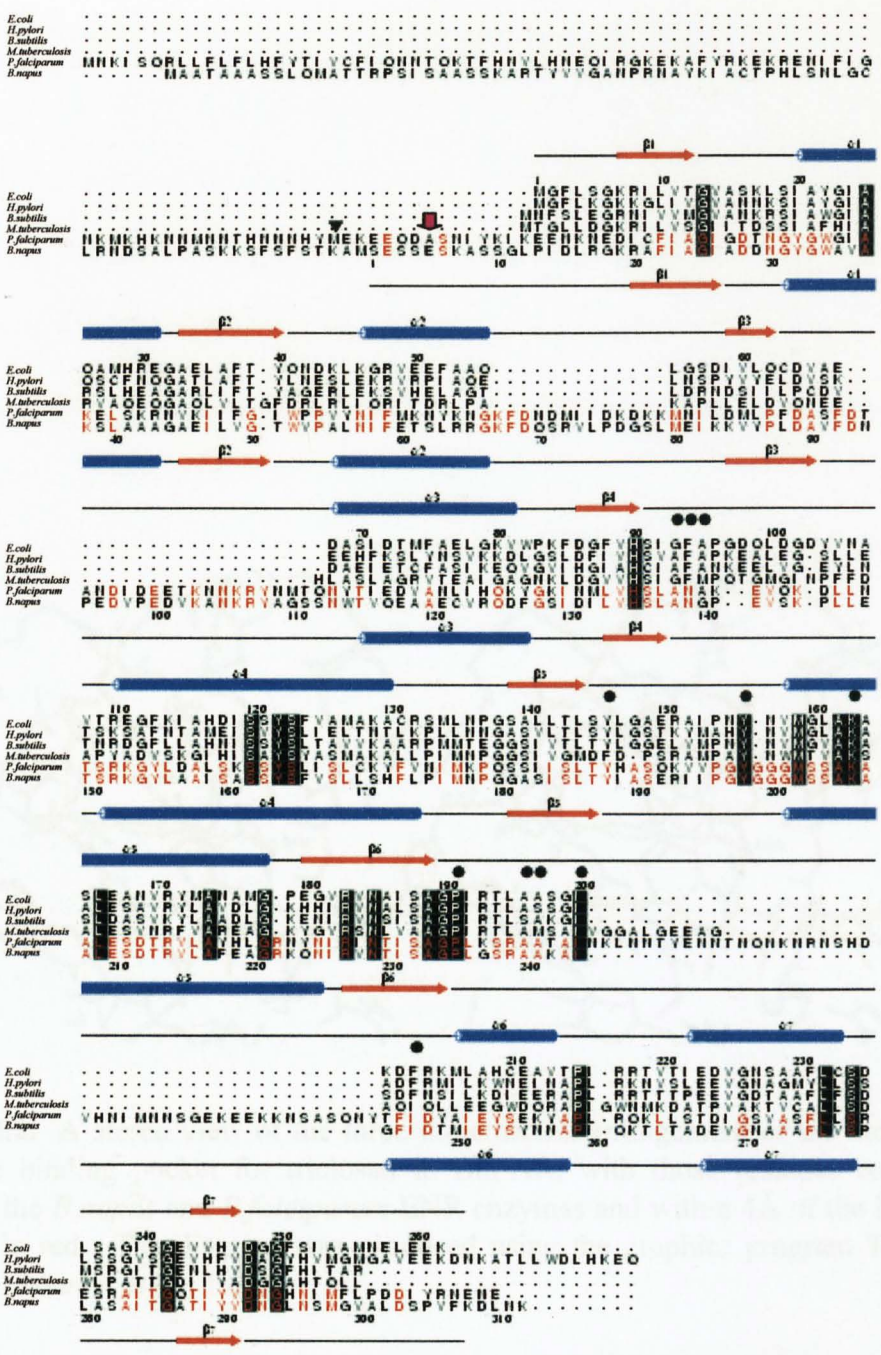


Figure 3.1. A multiple structure-based sequence alignment of the ENR enzyme from *E.coli*, *H.pylori*, *B.subtilis*, *M.tuberculosis*, *P.falciparum* and *B.napus*. The secondary structure elements of *E.coli* (top) and *B.napus* (bottom) ENR are shown with red arrows for β-strands and blue cylinders for α-helices. Sequence numbers for *E.coli* and *B.napus* ENR are given on top and below the alignment, respectively. Those residues completely conserved are in reverse type and those involved in triclosan binding within the *E.coli* structure have a black circle above. Residues conserved between *B.napus* and *P.falciparum* ENR are in red. The predicted start of the mature PfENR enzyme is shown by a black arrow and the start of the cloned PfENR sequence is indicated by a purple arrow. The sequences were aligned using ClustalW (Higgins & Thompson, 1994) and the figure was produced using the program Alscript (Barton, G.J., 1993).

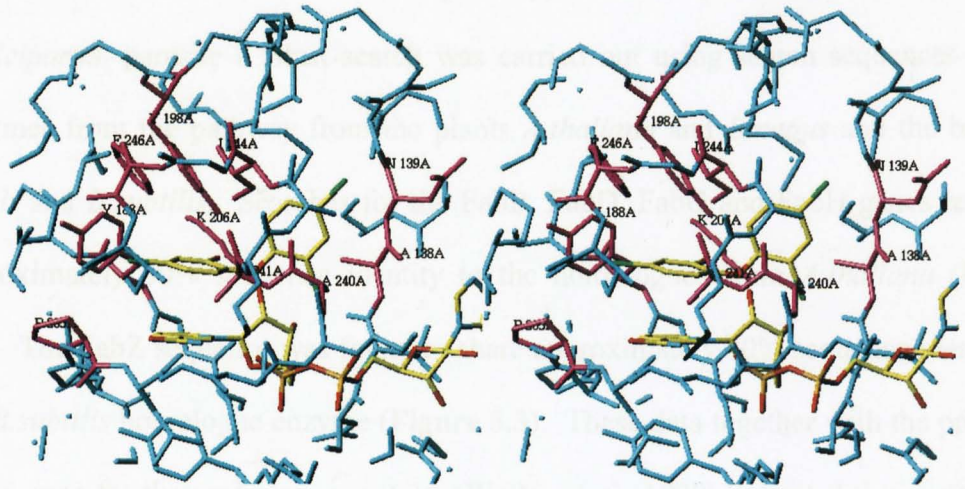


Figure 3.2. A stereo view of the three dimensional arrangement of the atoms that form the binding pocket for triclosan in BnENR, with those residues conserved between the *B.napus* and *P.falciparum* ENR enzymes and within 4Å of the inhibitor labelled in red. The diagram was produced using the graphics program TURBO-FRODO (Roussel *et al.*, 1990).

enzyme to accommodate larger substrates such as mycolic fatty acids by increasing the size of the substrate binding pocket (Rozwarski *et al.*, 1999). However the contrast between the small, predominantly hydrophobic, glycine rich sequence and the large bulky polar hydrophilic sequence may suggest an additional or novel function of this insert in the PfENR enzyme.

3.2.1 Other *P.falciparum* type II FAS genes

In order to establish whether a complete type II FAS pathway exists within the *P.falciparum* parasite a Blast search was carried out using search sequences of the enzymes from the pathway from the plants *A.thaliana* and *B.napus* and the bacteria *E.coli* and *B.subtilis*. Searches for the FabB, FabD, FabG and FabH genes revealed approximately 30% sequence identity to the homologues from *A.thaliana* (**Figure 3.3**). The FabZ sequence was found to share approximately 30% sequence identity to the *B.subtilis* homologue enzyme (**Figure 3.3**). These data together with the presence of the gene for the acyl carrier protein, (Waller *et al.*, 1998) indicated the presence of a fully functional type II FAS pathway within the *P.falciparum* parasite. Analysis of the genes showed that they all contain a bipartite N-terminal leader sequence to allow for targeting to the apicoplast but only the ENR gene contains any significant insert within the coding sequence. The location of these genes within the *P.falciparum* genome does not indicate the presence of any kind of operon system with the FabD and FabZ genes being present on chromosome 13, the FabB and FabI genes being present on chromosome 6 and the FabH and FabG being present on chromosomes 2 and 9, respectively. A Blast search was also carried out to try to detect a type I FAS pathway, using the sequences from *Homo sapiens*, *Candida albicans*, *Saccharomyces*

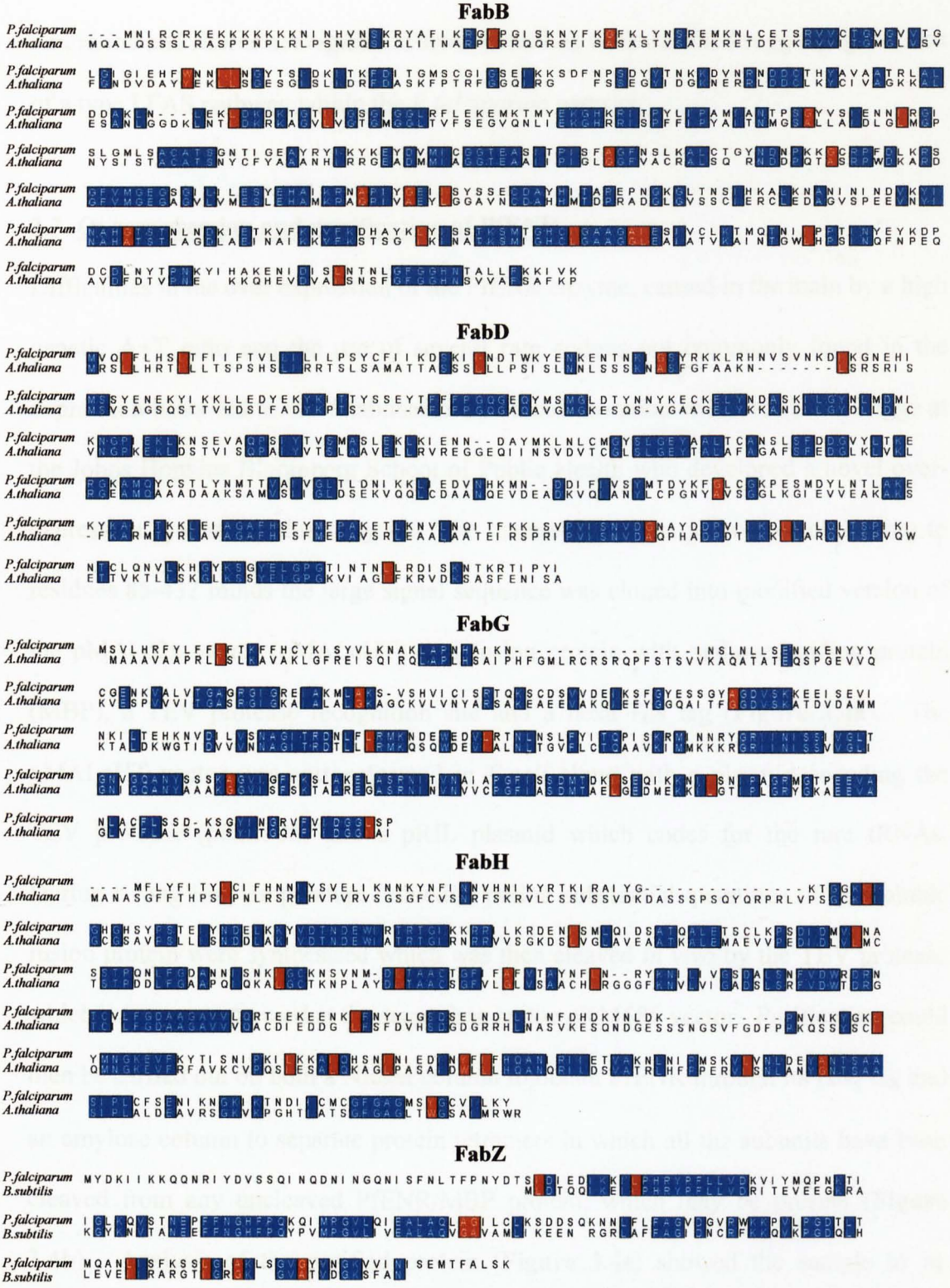


Figure 3.3. Sequence alignments of the FAS type II pathway enzymes of *P.falciparum*, against the *A.thialana* enzymes, except for the FabZ sequence which is aligned against the *B.subtilis* homologue sequence. Those residues which share sequence identity are in a blue box with reverse type. Those residues which are similar are in a red box with reverse type. The sequences for the *P.falciparum* FAS genes were obtained from the PlasmoDB (www.PlasmoDB.org). The *A.thialana* and *B.subtilis* sequences are from the swissprot database (Boeckmann *et al.*, 2003). The sequences were aligned using ClustalW (Higgins & Thompson, 1994) and the figure was produced using the program Alscript (Barton, G.J., 1993).

cerevisiae and *Rattus norvegicus*. However no hits were detected suggesting the lack of a type I FAS pathway within the *P.falciparum* parasite.

3.3 Overexpression and purification of PfENR

Difficulties in the over expression of the PfENR enzyme, caused in the main by a high genetic A+T ratio and the use of several rare codons not commonly found in the expression hosts led to the establishment of a collaboration with Prof Sean Prigge at the Johns Hopkins Bloomberg School of Public Health who developed a novel over-expression system for this enzyme. The mature PfENR sequence corresponding to residues 85-432 minus the large signal sequence was cloned into modified version of the pMALc2x vector (pMALcHT) as a fusion protein with maltose binding protein (MBP), a TEV protease recognition site and a hexa His tag (**Figure 3.4a**). The pMALcHT vector was cotransformed in *E.coli* along with a plasmid encoding the TEV protease (pKM586) and a pRIL plasmid which codes for the rare tRNAs. During a long low temperature induction period substantial quantities of the soluble fusion protein were synthesised which was then cleaved *in vivo* by the TEV protease which was over-expressed at the same time on the pKM586 vector. Purification could then be carried out on both a Nickel column to obtain PfENR through its His_6 tag and an amylose column to separate protein tetramers in which all the subunits have been cleaved from any uncleaved PfENR/MBP protein, which may be present (**Figure 3.4b**). Analysis of the purified protein (**Figure 3.4c**) showed the sample to be approximately 95% pure. The PfENR enzyme was then concentrated to 12mg ml^{-1} in a buffer of 20mM Na/K phosphate pH 8.0, 150mM NaCl for crystallization trials. The overexpressed PfENR is numbered such that the first residue, which formed the

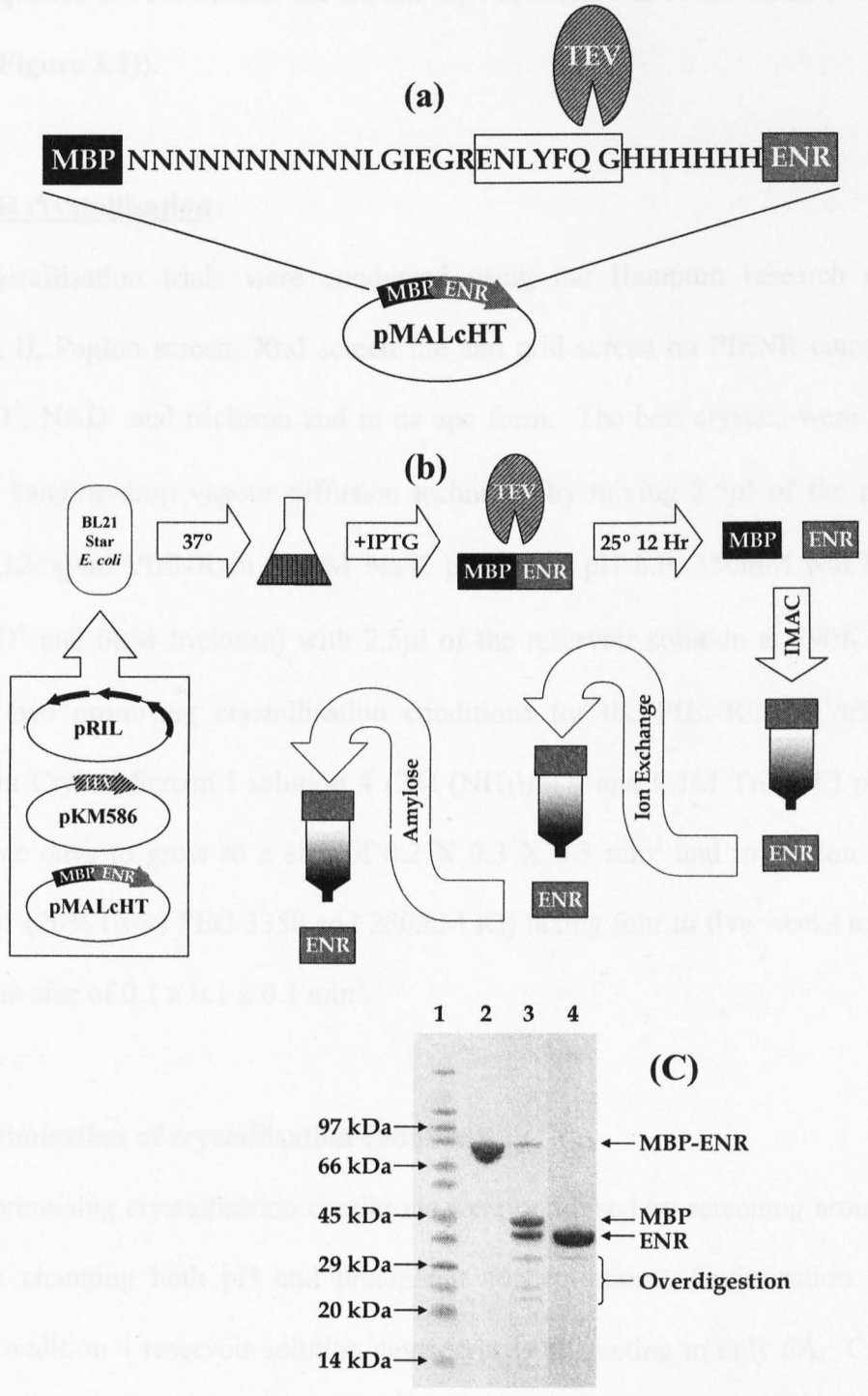


Figure 3.4. (a) A schematic representation of the pMALcHT vector. A single letter abbreviation is used to describe the sequence of the linker between the MBP fusion protein and PfENR. The seven residue TEV protease binding site is highlighted by a black box. (b) A schematic representation of the purification scheme for PfENR. (c) SDS-PAGE of PfENR purification. Lane one shows Sigma markers with pure MBP-ENR fusion protein produced by the pMALc2x vector in lane 2. Lane 3 contains the factor Xa-digested fusion and lane 4 shows the PfENR produced by *in vivo* cleavage using the pMALcHT vector.

mature sequence of PfENR after the six His tag sequence, was renumbered as residue 1 ((Ala^1)(**Figure 3.1**)).

3.4 PfENR crystallisation

Initial crystallisation trials were conducted using the Hampton research crystal Screens I, II, PegIon screen, Xtal screen lite and grid screen on PfENR complexed with NAD^+, NAD^+ and triclosan and in its apo form. The best crystals were grown using the hanging-drop vapour diffusion technique by mixing 2.5µl of the protein solution (12mg/ml PfENR in 20mM Na/K phosphate pH 8.0, 150mM NaCl, with 5µM NAD^+ and 6µM triclosan) with 2.5µl of the reservoir solution at 290K. This produced two promising crystallisation conditions for the PfENR/NAD^+/triclosan complex in Crystal Screen I solution 4 (2M $(NH_4)_2SO_4$ and 0.1M Tris-HCl pH8.5), taking three days to grow to a size of 0.2 X 0.3 X 0.5 mm^3 and in PegIon screen solution 11 (20% (w/v) PEG 3350 and 250mM KI) taking four to five weeks to reach a maximum size of 0.1 x 0.1 x 0.1 mm^3.

3.4.1 Optimisation of crystallisation conditions

The two promising crystallisation conditions were optimised by screening around the conditions changing both pH and precipitant concentration. Optimisation of the Screen I condition 4 reservoir solution gave crystals diffracting to only 6Å. Crystals of the PfENR/NAD^+/triclosan complex grown through the optimisation of PegIon screen 11 (**Figure 3.5a**) were of a better quality, with the optimised condition (19.5% (w/v) PEG 3350 and 230mM KI) producing crystals of an average size of 0.15 x 0.10 x 0.10 mm^3 diffracting to 2.8Å "in house" and to 2.0Å at the SRS Daresbury laboratory.

3.5 Data Collection

In order to perform data collection at cryogenic temperatures, crystals of the PfENR/NAD^+/triclosan complex were transferred to a solution identical to the well solution but containing 20% glycerol as a cryo protectant. These crystals were then flash frozen in a stream of nitrogen gas at 100K using an Oxford Cryosystems Cryostream device. An initial dataset to 2.8Å was collected by rotating the crystal by 1° increments over a range of 190° and exposing for 15 minutes per image on a Mar Research Mar345 image plate mounted on a Rigaku RU200 X-ray generator. Analysis of the data were carried out using the DENZO suite of programs (Otwinowski & Minor, 1997) and the CCP4 suite of programs (CCP4, 1994). A second data set was collected on a similar crystal on a Quantum Q4 CCD detector on station 14.1 at the SRS Daresbury laboratory. The crystal was exposed for 15 seconds and rotated by 1° increments per image over a range of 190° to collect data to 2.0Å resolution (**Figure 3.5b**). Analysis of the processed 2.0Å data set showed the Rmerge values at 2.18Å and 2.1Å were 0.395 and 0.5, respectively. Therefore the data were truncated to 2.18Å. The statistics for the data collection can be seen in Table 3.1

3.5.1 Space group determination

Preliminary analysis of the data indicated that the crystal belonged to the monoclinic spacegroup P2 or $P2_1$, with unit cell dimensions a= 88.2Å, b= 82.4Å, c= 94.8Å, and $\alpha = \gamma = 90°$ $\beta = 90.8°$. Analysis of the X-ray diffraction pattern showed that reflections were observed along the $0k0$ axis (b*) only where they satisfied the condition $k = 2n$, (**Figure 3.6**) indicating that the crystals belonged to the space group $P2_1$.

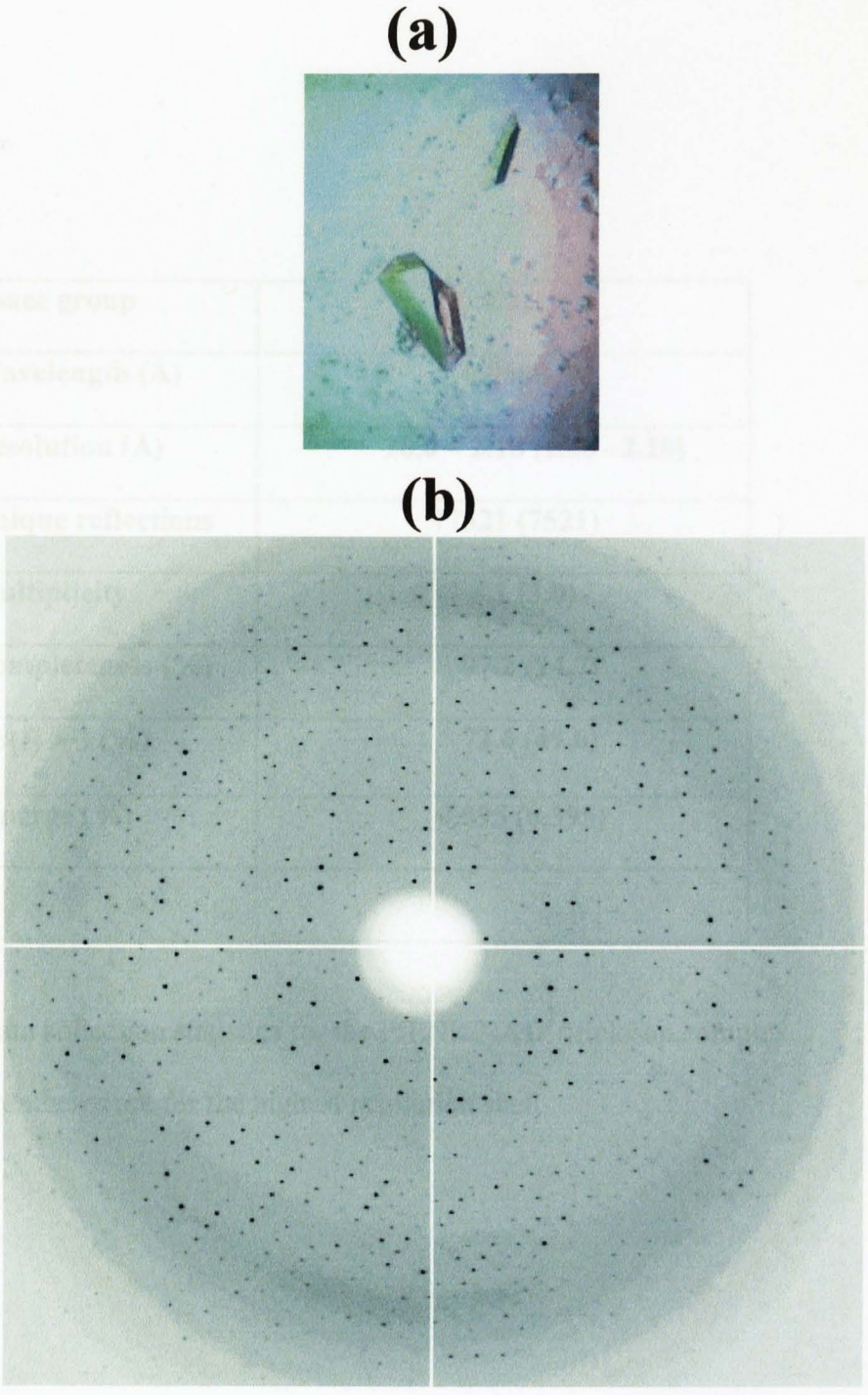


Figure 3.5. (a) Crystals of the PfENR/NAD^+/triclosan complex grown in 19.5 % (w/v) PEG 3350 and 230 mM KI, with approximate dimensions of 0.15 x 0.10 x 0.10 mm^3. (b) A 1° oscillation frame of a PfENR/NAD^+/triclosan complex crystal collected on a Quantum Q4 CCD detector at station 14.1 at the SRS Daresbury laboratory. The resolution at the edge of the diffraction plate is 2.0Å.

Space group	**$P2_1$**
Wavelength (Å)	**0.9600**
Resolution (Å)	**20.0 – 2.18 (2.26 –2.18)**
Unique reflections	**77321 (7521)**
Multiplicity	**3.1 (3.0)**
Completeness (%)	**97.2 (94.7)**
I/σ(*I*) > 3 (%)	**72.6 (49.6)**
Rmerge (%)	**0.095 (0.395)**

Table 3.1 Data collection statistics for the PfENR/NAD^+/triclosan complex. Values in parentheses are for the highest resolution shell.

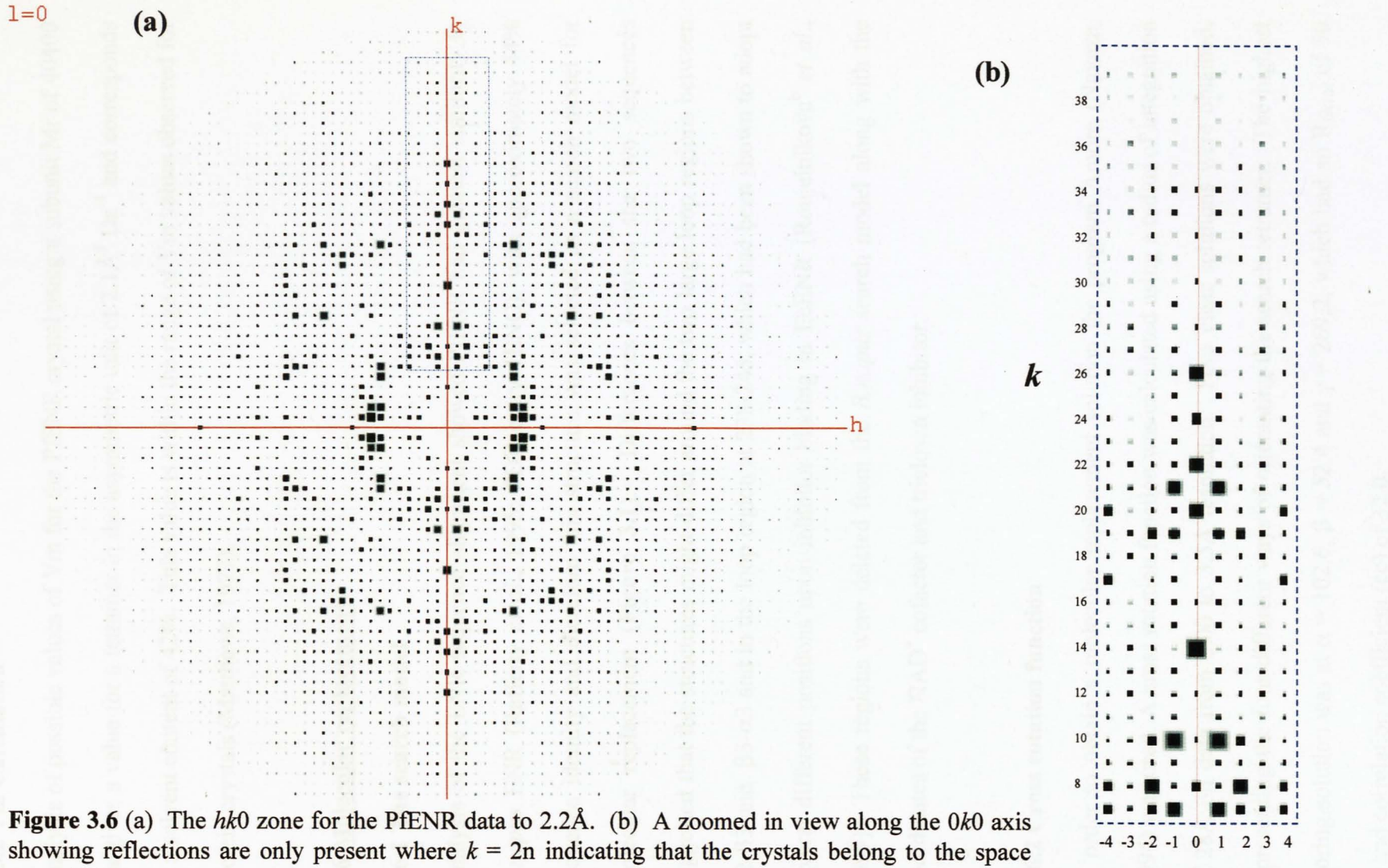


Figure 3.6 (a) The *hk*0 zone for the PfENR data to 2.2Å. (b) A zoomed in view along the 0*k*0 axis showing reflections are only present where $k = 2n$ indicating that the crystals belong to the space group $P2_1$.

3.5.2 Vm Calculation

Analysis of possible values of Vm for the PfENR crystal using a subunit Mr of 40400 Da gives a value for a tetramer in the asymmetric unit of 2.1$Å^3$ Da^{-1} and corresponds to a solvent content of 42%. This value is within the range of Vm values observed for protein crystals (Matthews, 1977).

3.6 Molecular replacement

3.6.1 The search model

Analysis of the ENR sequence alignment showed PfENR had greatest similarity to *B.napus* ENR (Rafferty *et al.*, 1995; PDB entry 1ENO), with approximately 44% sequence identity and therefore this structure was selected as a search model for molecular replacement (**Figure 3.1**). Differences between the two sequences suggested that the structures might differ around the two large loop regions between α2-β3 and β3-α3 and in the loop region at 238-245, which had been shown to adopt several different positions upon inhibitor binding in EcENR (Roujeinikova[b] *et al.*, 1999). These regions were deleted from the *B.napus* search model along with the coordinates of the NAD^+ cofactor and triclosan inhibitor.

3.6.2 Cross rotation function

In order to obtain a molecular replacement solution the program Amore (Navaza, 1994) was used. A cross rotation function was calculated using a radius of integration of 28Å on data from 8.00 to 3.5Å resolution. Four clear solutions were obtained, representing the four different ways of orientating the search tetramer. The highest scoring solution was at $\alpha = 102.6$, $\beta = 52.4$ and $\gamma = 269.2$, which had an R_{factor} of 50 % and correlation coefficient (cc) of 32.0.

3.6.3 Translation function

The translation function was calculated using the program Amore (Navaza, 1994), producing a clear solution in the space group $P2_1$ with an R_{factor} of 45.1% and cc of approximately 40.0 at x = 0.244 and z = 0.232. No translation was calculated for the y axis because in monoclinic space groups this is a "floating axis" and as such can be any value. The same procedure was tried for the space group P2, however no clear solution could be obtained, confirming the correct space group as $P2_1$.

3.7 Phase improvement, Model building and refinement

The correct molecular replacement model output by the program Amore was subject to various stages of refinement and model building to improve the PfENR model.

3.7.1 Rigid body refinement

The model from molecular replacement was subjected to 10 rounds of rigid body refinement within the program REFMAC5 (Murshudov *et al.*, 1997) using data in the resolution range 20Å to 5Å with each of the four subunits within the tetramer being treated as an individual rigid body. There was only a small drop in the R_{factor} from 0.44 to 0.42 and a drop from 0.46 to 0.43 in the R_{free}.

3.7.2 Restrained least squares refinement

The model from the rigid body refinement was then submitted to restrained positional refinement using REFMAC5 (Murshudov *et al.*, 1997). Five cycles of refinement were carried out on data in the resolution range of 20Å to 2.2Å and sigma A maps were prepared. This reduced the R_{factor} from 0.4 to 0.29 whilst the R_{free} changed from 0.42 to 0.4. Analysis of the initial electron density maps showed clear density for the

NAD^+ cofactor and triclosan. Given that both of these had been omitted from the model this provided confidence in the molecular replacement solution (**Figure 3.7**).

3.7.3 Model building

The graphics program TURBO-FRODO (Roussel *et al.*, 1990) was used to change the residues of the BnENR search model to the corresponding residues in PfENR and to include the NAD^+ and triclosan. The two extended loops which were omitted from the initial model could also be seen and were fitted accordingly. Initially only subunit A was constructed and this was rotated and superimposed onto the other three subunits. The model was refined and rebuilt in an iterative process to an R_{factor} and R_{free} value of 0.23 and 0.27, respectively. Analysis of the ENR family shows that PfENR contains a large polar insert from Asn^{241} to Tyr^{282}, positioned next to the $\alpha 7$ helix which could not be found in the electron density maps. In order to try to locate and define the polar insert the map was flooded with water atoms using the ARP/WARP function in REFMAC5. However the procedure did not result in clear density for residues in the loop. The inability to locate the loop would suggest that it does not form any ordered structure within this crystal form of PfENR complexed with NAD^+ and triclosan.

3.7.4 Addition of solvent

The 2.18Å resolution data was high enough to be able to accurately assign several water molecules into the PfENR model using the program ARP/wARP version 5 (Perrakis *et al.*, 1997) in the final stages of refinement. Waters were assigned to peaks within the $1F_{obs} - 1F_{calc}$ difference map with a sigma cut-off of >2.0 if they lay within 2.2Å to 3.3Å of a protein O or N atom. Once solvent addition was completed

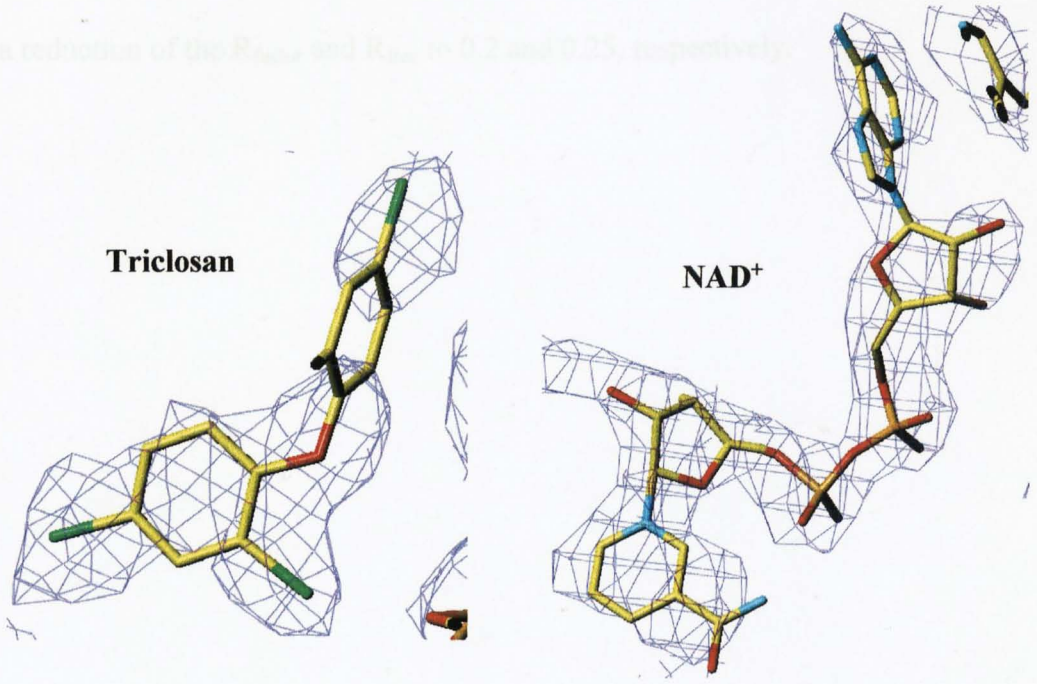


Figure 3.7 The electron density contoured at 0.8σ in the initial $2F_{obs}$-F_{calc} maps which indicated the presence of triclosan and the NAD^+ cofactor bound to PfENR. The density for subunit A is shown. The NAD^+ and triclosan molecules are coloured with yellow for carbon, blue for nitrogen, red for oxygen, green for chlorine and orange for phosphorous. The triclosan and NAD^+ molecules have been added as a guide to the electron density. The figure was produced in TURBO-FRODO (Roussel *et al.*, 1990).

each individual water molecule was visually inspected to check that it was in good electron density and that it was within hydrogen bond distance of surrounding atoms. The B-factors of all water molecules was also checked and any water which had a B-factor over 60 $Å^2$ was removed. A total of 518 water molecules were added, resulting in a reduction of the R_{factor} and R_{free} to 0.2 and 0.25, respectively.

CHAPTER 4

Analysis of the *Plasmodium Falciparum* enoyl reductase structure

4.1 Introduction

The following chapter describes the final model of *Plasmodium falciparum* enoyl ACP reductase and the mode of binding triclosan and NAD^+

4.2 Model stereochemistry

The final model of the PfENR/NAD^+/triclosan complex comprises 112 residues in a α-helical conformation (~40%) and 45 residues formed β-sheets (~16%) with a total of 483 hydrogen bonds being made as defined by the program Promotif (Hutchinson & Thornton). The final model contained the following residues defined by good density: Asp^{15} to Asn^{241}, Tyr^{283} to Arg^{345} in both the A and B subunits, Asn^{13C} to Asn^{241C} and Tyr^{283C} to Asn^{346C} in subunit C and Asp^{15D} to Ile^{240D} and Tyr^{283D} to Tyr^{344D} in subunit D together with one molecule of NAD^+ and triclosan per subunit and 518 water molecules. In all 7 residues had to be truncated back to Cβ due to weak or no density (Lys^{72A}, Asp^{73A}, Lys^{60B}, Asn^{140B}, Lys^{167B}, Lys^{42C} and Lys^{102D}). The final model has an R_{factor} of 0.2 (R_{free} of 0.25) for all data within the resolution range 20.0 to 2.18Å (**Table 4.1**). Analysis of the PfENR model geometry was carried out using the program PROCHECK (Laskowski *et al.*, 1993) and showed a tight geometry with most parameters being inside or better than the range expected for a model at this resolution (**Figure 4.1 (a), (b), (c) and (d)**). The bond length and angle r.m.s.d from ideal geometry as defined by REFMAC5 (Murshudov et al., 1997) are 0.02Å and 1.8°, respectively. Examination of the χ1-χ2 plots for all residue types showed the side-chains to be in favourable conformations in all four monomers (**Figure 4.2 (a), (b), (c)** and **(d)**). A Ramachandran plot was calculated for PfENR and showed approximately 90% of residues to be in the most favoured regions and 10% in additionally allowed regions, with no residues in the generously or disallowed

Refinement statistics and model quality	
Total non-H atoms	9911
Water sites	518
Average B-value ($Å^2$)	30
Resolution limits (Å)	20-2.18
Number of reflections	65771
R-value	0.2
Rfree Value	0.25
r.m.s. bond deviation (Å)	0.02
r.m.s. angle deviation (deg.)	1.8

Table 4.1. A table to show the refinement statistics for the refined PfENR complexed with NAD^+ and triclosan to 2.2Å resolution.

Subunit A.

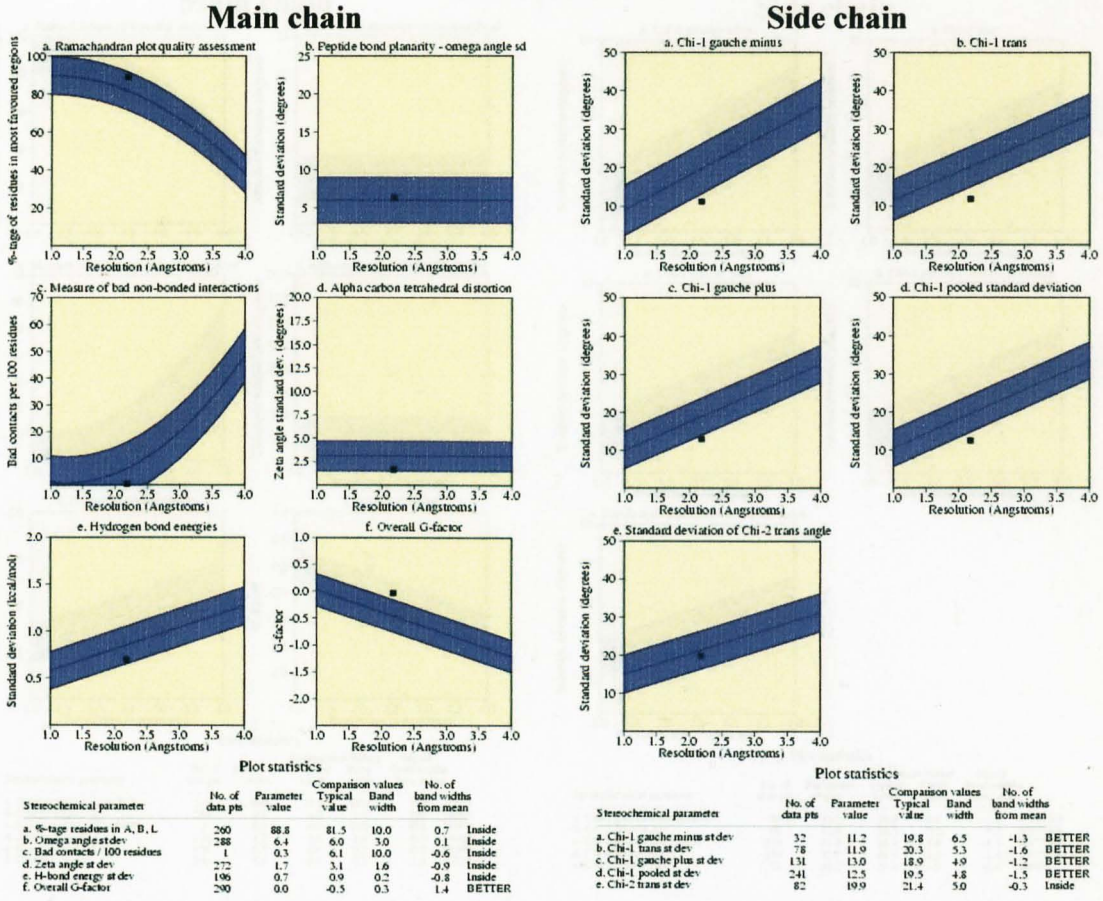


Figure 4.1. (a) Main chain and side chain parameters for subunit A, produced using the program PROCHECK (Laskowski *et al.*, 1993).

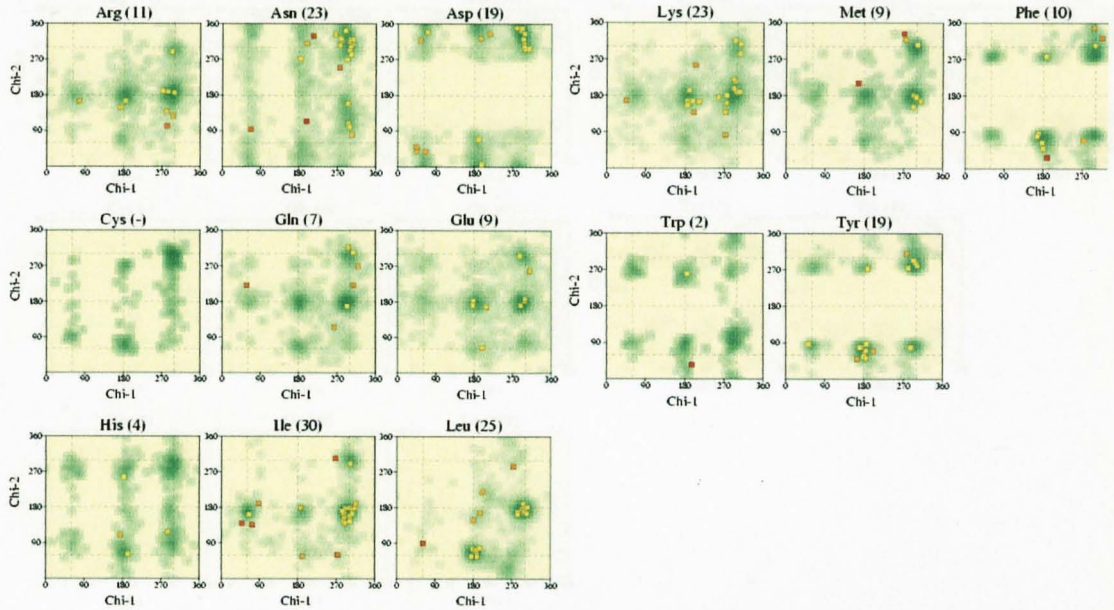


Figure 4.2. (a) The $\chi 1$ vs. $\chi 2$ plots for subunit A. Favourable conformations are represented in the shaded areas. The figure was produced using the program PROCHECK (Laskowski *et al.*, 1993).

Subunit B

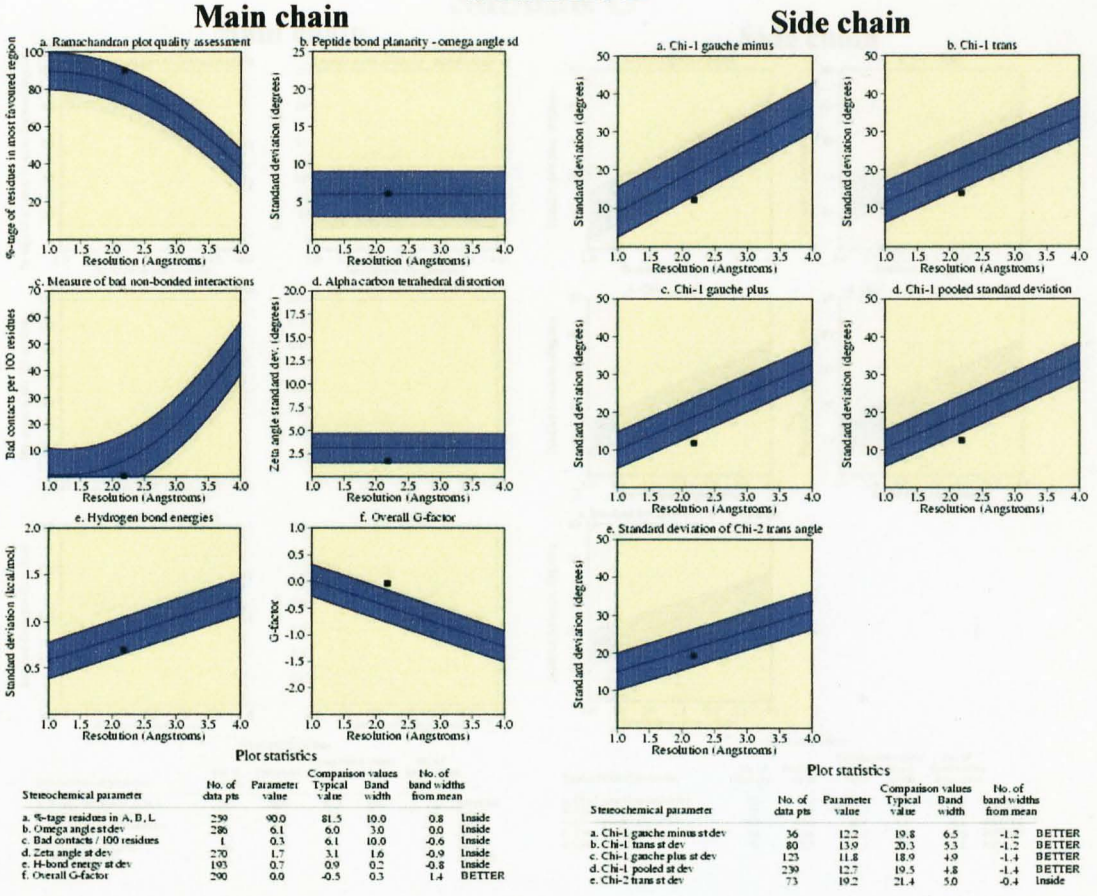


Figure 4.1. (b) Main chain and side chain parameters for subunit B, produced using the program PROCHECK (Laskowski *et al.*, 1993).

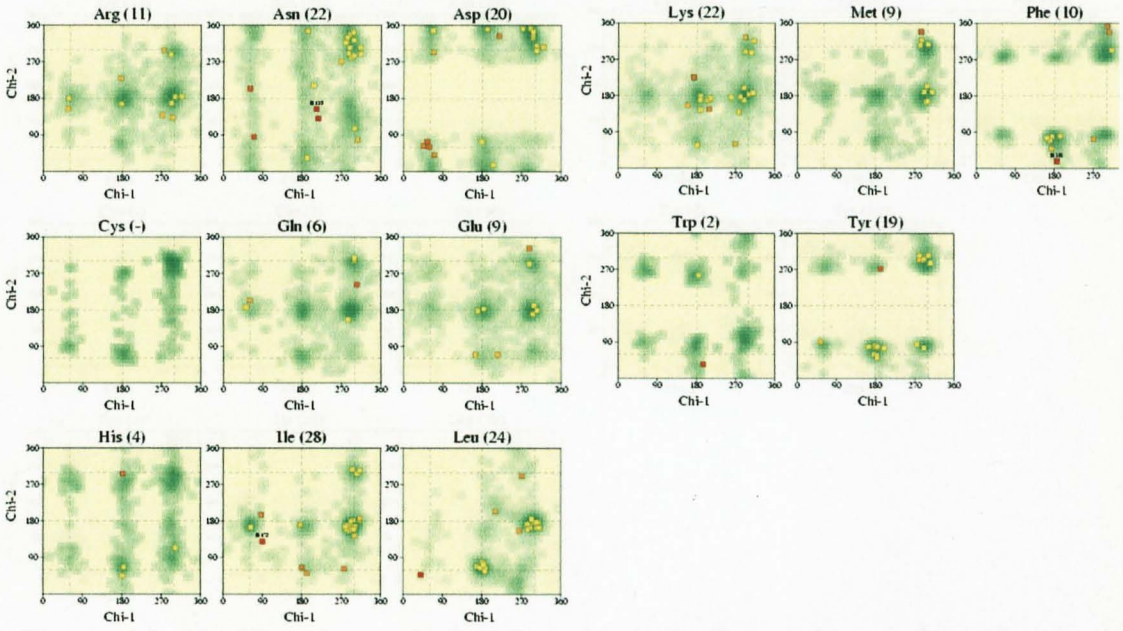
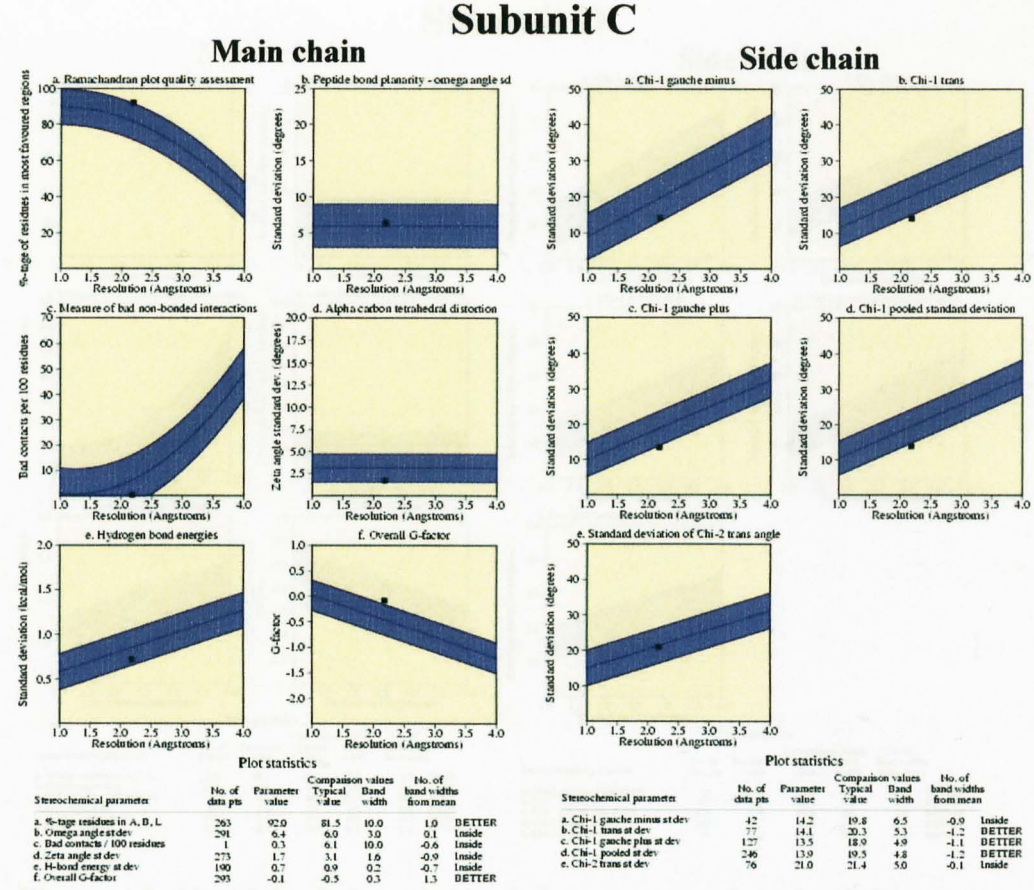


Figure 4.2. (b) The χ 1 vs. χ 2 plots for subunit B. Favourable conformations are represented in the shaded areas. The figure was produced using the program PROCHECK (Laskowski *et al.*, 1993).



Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. %-tage residues in A, B, L	263	92.0	81.5	10.0	1.0	BETTER
b. Omega angle st dev	291	6.4	6.0	3.0	0.1	Inside
c. Bad contacts / 100 residues	1	0.3	6.1	10.0	-0.6	Inside
d. Zeta angle st dev	273	1.7	3.1	1.6	-0.9	Inside
e. H-bond energy st dev	190	0.7	0.9	0.2	-0.7	Inside
f. Overall G-factor	293	-0.1	-0.5	0.3	1.3	BETTER

Stereochemical parameter	No. of data pts	Parameter value	Comparison values Typical value	Band width	No. of band widths from mean	
a. Chi-1 gauche minus st dev	42	14.2	19.8	6.5	-0.9	Inside
b. Chi-1 trans st dev	77	14.1	20.3	5.3	-1.2	BETTER
c. Chi-1 gauche plus st dev	127	13.5	18.9	4.9	-1.1	BETTER
d. Chi-1 pooled st dev	246	13.9	19.5	4.8	-1.2	BETTER
e. Chi-2 trans st dev	76	21.0	21.4	5.0	-0.1	Inside

Figure 4.1. (c) Main chain and side chain parameters for subunit C, produced using the program PROCHECK (Laskowski *et al.*, 1993).

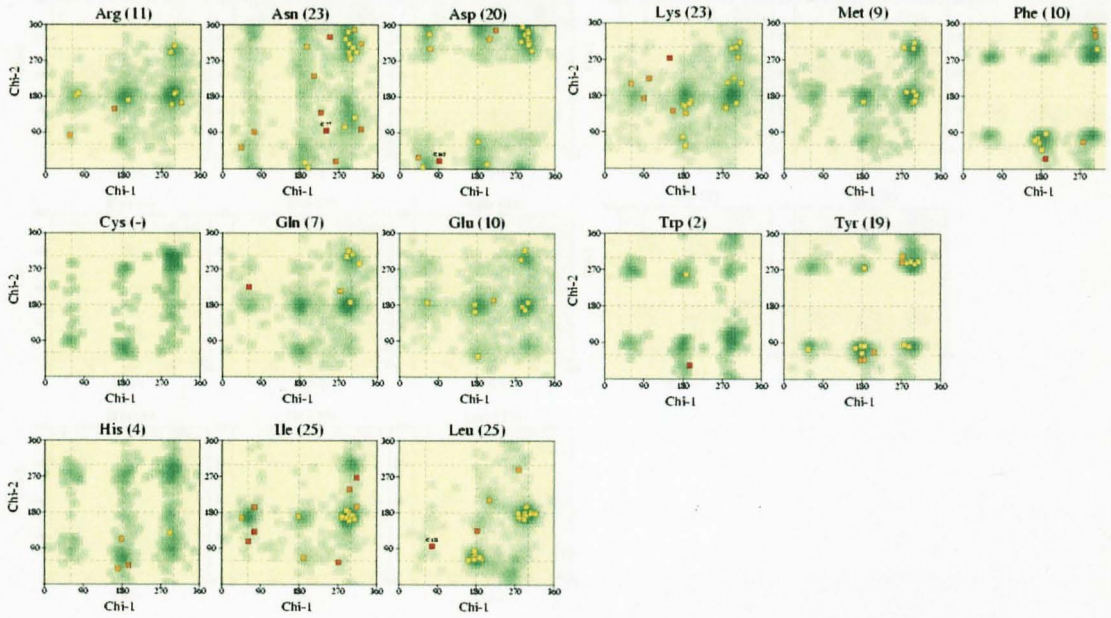


Figure 4.2. (c) The $\chi 1$ vs. $\chi 2$ plots for subunit C. Favourable conformations are represented in the shaded areas. The figure was produced using the program PROCHECK (Laskowski *et al.*, 1993).

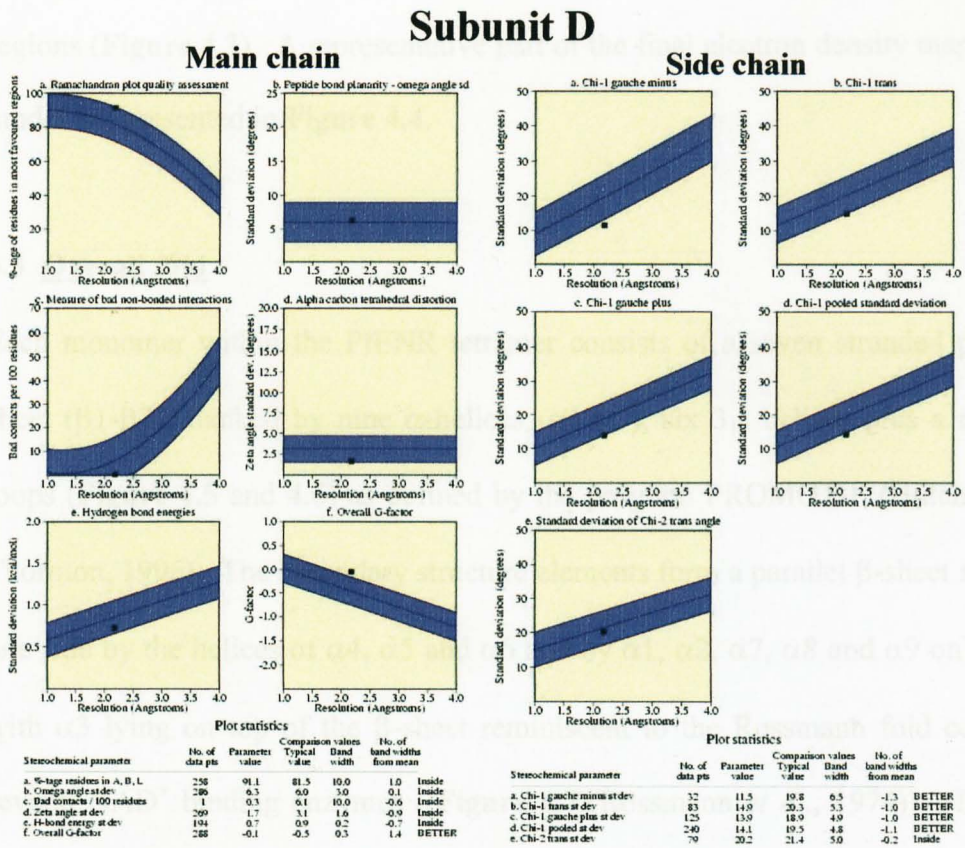


Figure 4.1. (d) Main chain and side chain parameters for subunit D, produced using the program PROCHECK (Laskowski et al., 1993).

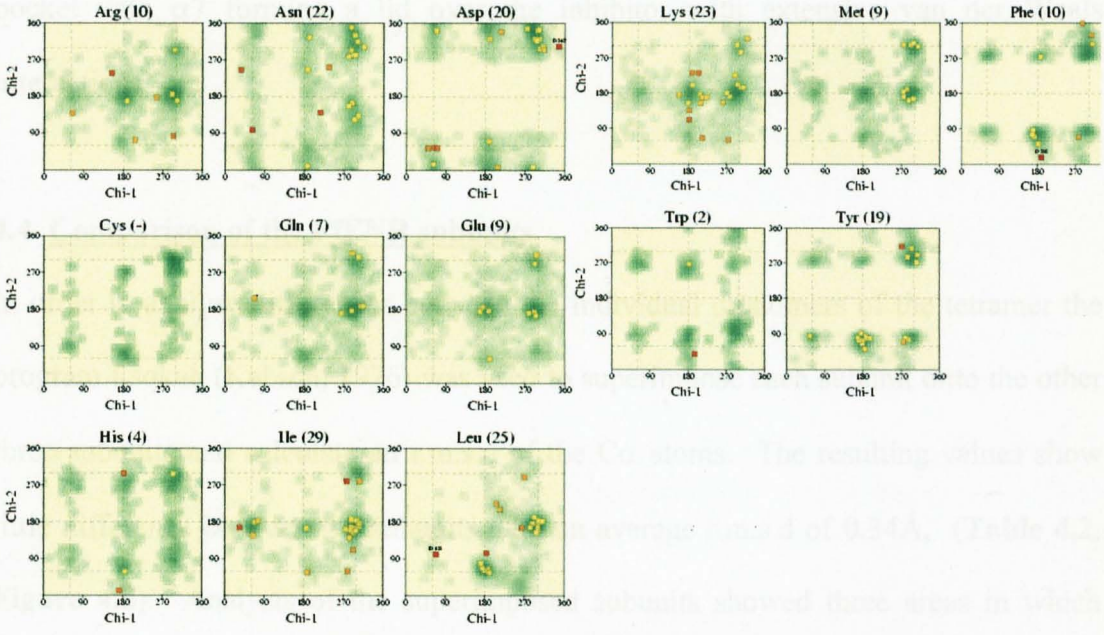


Figure 4.2. (d) The $\chi 1$ vs. $\chi 2$ plots for subunit D. Favourable conformations are represented in the shaded areas. The figure was produced using the program PROCHECK (Laskowski *et al.*, 1993).

regions (**Figure 4.3**). A representative part of the final electron density map and final model are presented in **Figure 4.4**.

4.3 Overall fold

Each monomer within the PfENR tetramer consists of a seven stranded parallel β-sheet (β1-β7), flanked by nine α-helices (α1-α9), six 3_{10} helices plus a number of loops (**Figure 4.5** and **4.6**) as defined by the program PROMOTIF (Hutchinson and Thornton, 1996). The secondary structure elements form a parallel β-sheet flanked on one side by the helices of α4, α5 and α6 and by α1, α2, α7, α8 and α9 on the other, with α3 lying on top of the β-sheet reminiscent to the Rossmann fold common to several NAD^+ binding enzymes ((**Figure 4.7**)(Rossmann *et al.*, 1974)). The NAD^+ cofactor binds at the C-terminal end of the β-sheet and makes important interactions with α4, α5 and the β1/α1 loop. Helices α5, α6 and α8 form the inhibitor binding pocket with α7 forming a lid over the inhibitor with extensive van der Waals interactions.

4.4 Comparison of the PfENR subunits

In order to analyse differences between the individual monomers of the tetramer the program Lsqkab (Kabsch, 1976) was used to superimpose each subunit onto the other three subunits and calculate an r.m.s.d of the Cα atoms. The resulting values show little difference between the subunits with an average r.m.s.d of 0.34Å, (**Table 4.2**, **Figure 4.8**). Analysis of the superimposed subunits showed three areas in which more significant differences were observed. Interestingly each of these areas corresponds to regions of the sequence, which represent insertions when compared to the bacterial ENR enzyme (Asp^{71} – Met^{76}, Thr^{90} – Asn^{109} and Thr^{281} – Glu^{288}). Re-

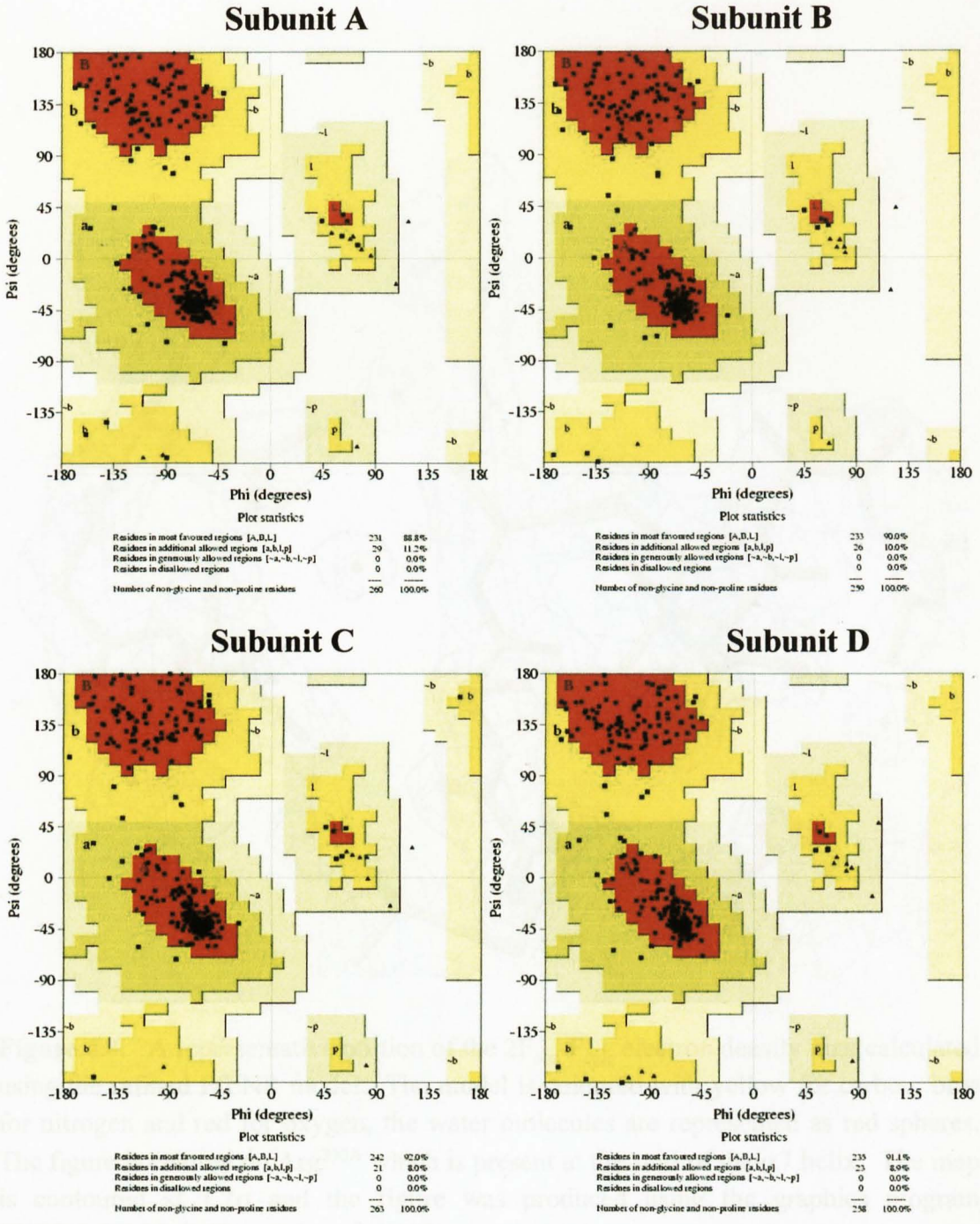


Figure 4.3. Ramachandran plots for subunits A, B, C and D of PfENR. The shading of the Ramachandran plot is with red, dark yellow, yellow and white representing the decreasingly favoured regions respectively. Produced using PROCHECK (Laskowski *et al.*, 1993).

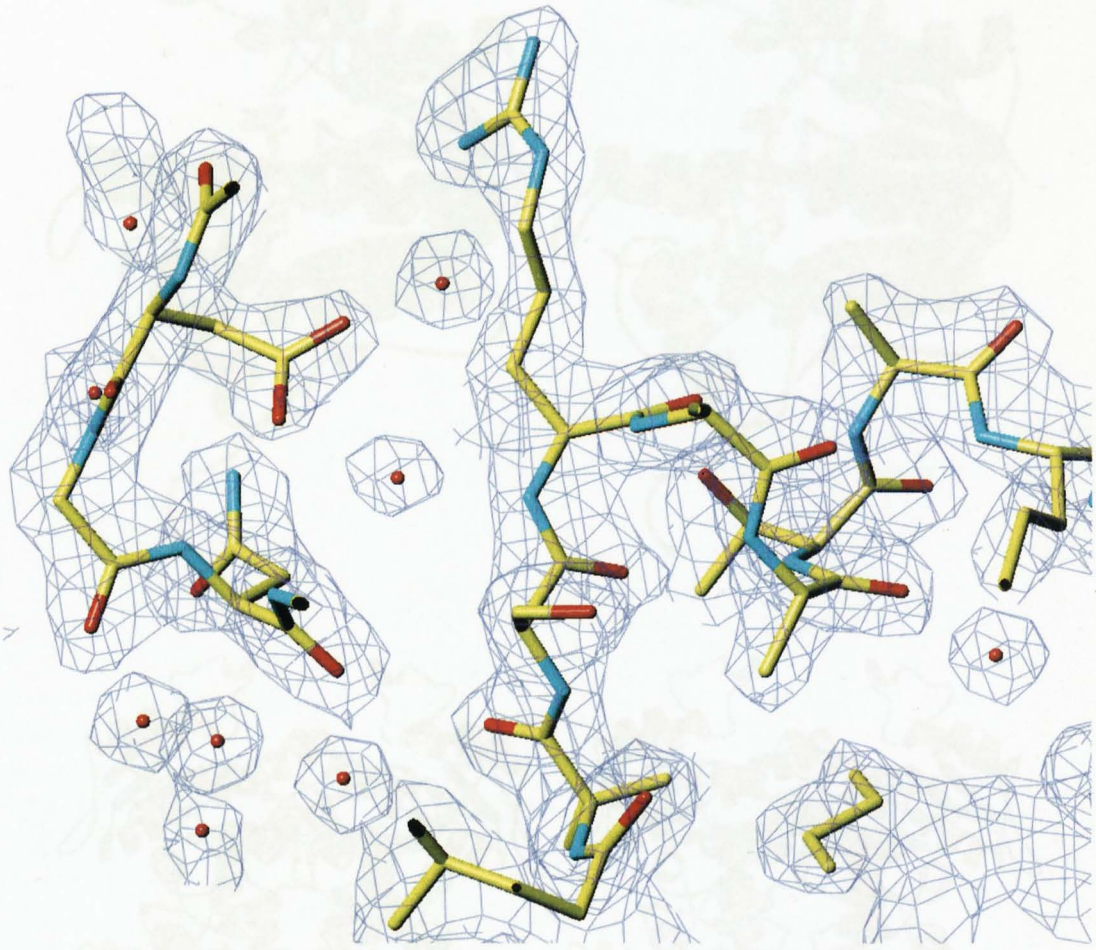


Figure 4.4. A representative portion of the $2F_{obs}$-F_{calc} electron density map calculated using the refined PfENR model. The model is coloured with yellow for carbon, blue for nitrogen and red for oxygen, the water molecules are represented as red spheres. The figure is centred on Arg^{235A} which is present at the start of the α7 helix. The map is contoured at 1.2σ and the figure was produced using the graphics program TURBO/FRODO (Roussel *et al.*, 1990).

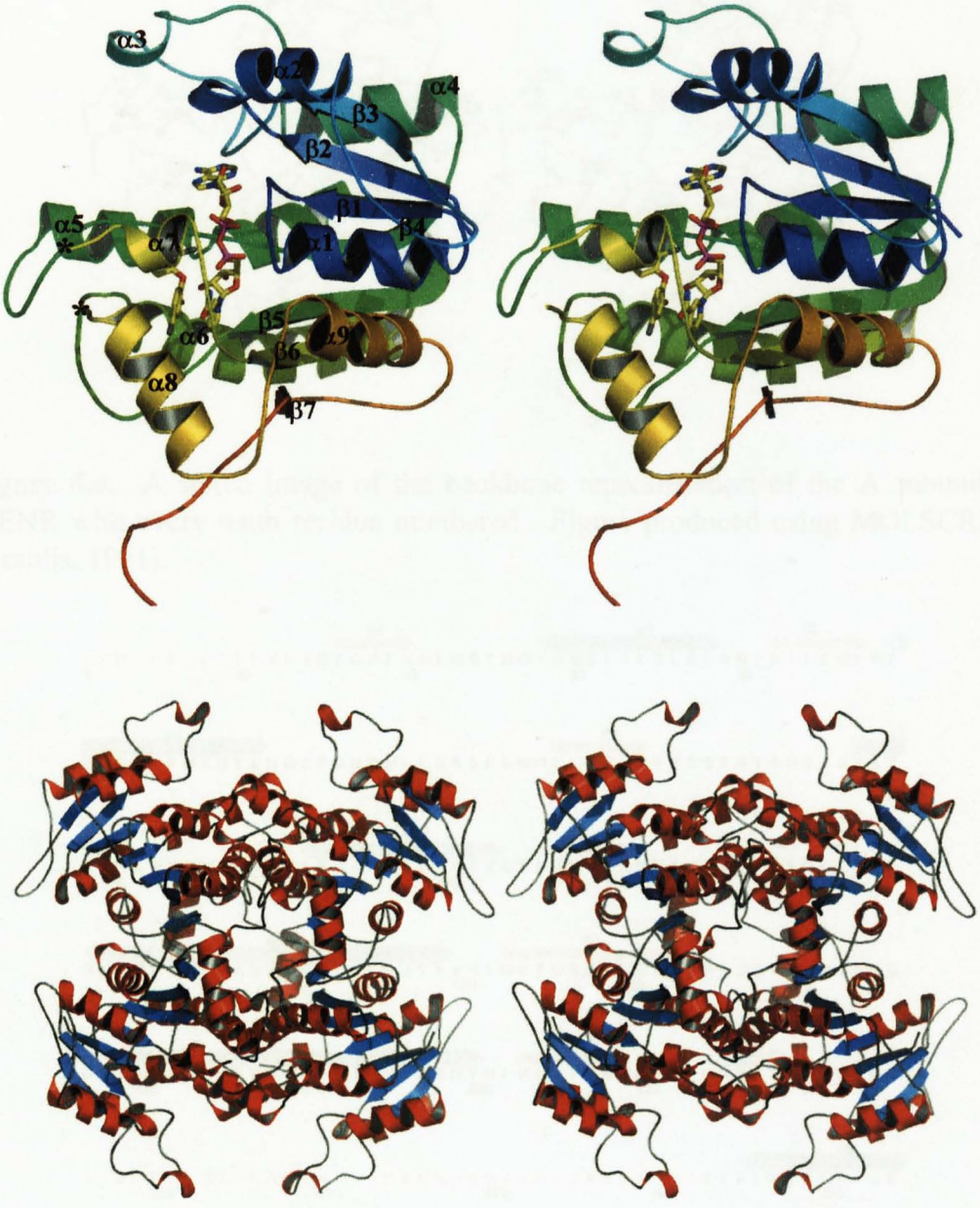


Figure 4.5. (a) A stereo view of the PfENR monomer with both NAD^+ cofactor and inhibitor bound. The PfENR monomer is coloured from blue to red with each secondary structure element being labelled, the region at which the large inserted sequence is present is indicated by a black star. **(b)** A stereo view of the PfENR tetramer. The figure was produced using the graphics program Pymol (Delano, 2002).

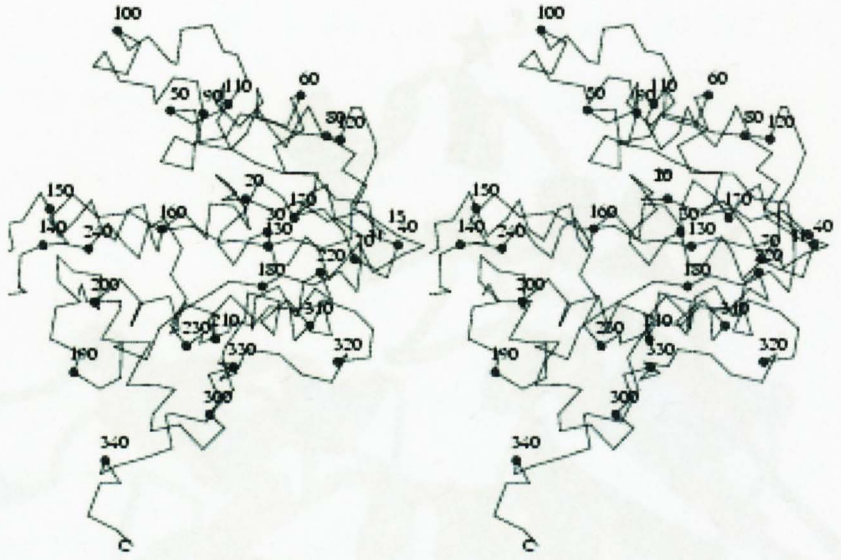


Figure 4.6. A stereo image of the backbone representation of the A subunit of PfENR with every tenth residue numbered. Figure produced using MOLSCRIPT (Kraulis, 1991).

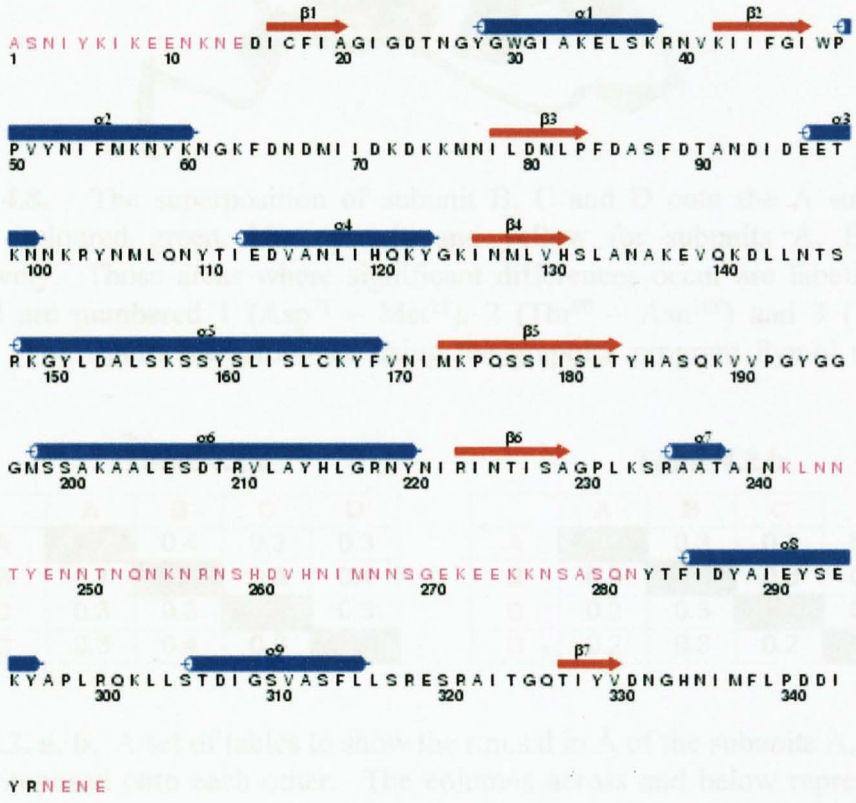


Figure 4.7. The sequence for PfENR with the secondary structural elements shown as blue cylinders and red arrows for α-helices and β-strands respectively. Those residues not defined by density within the PfENR crystal structure are coloured purple. The figure was produced using the program Alscript (Barton, G.J. 1993).

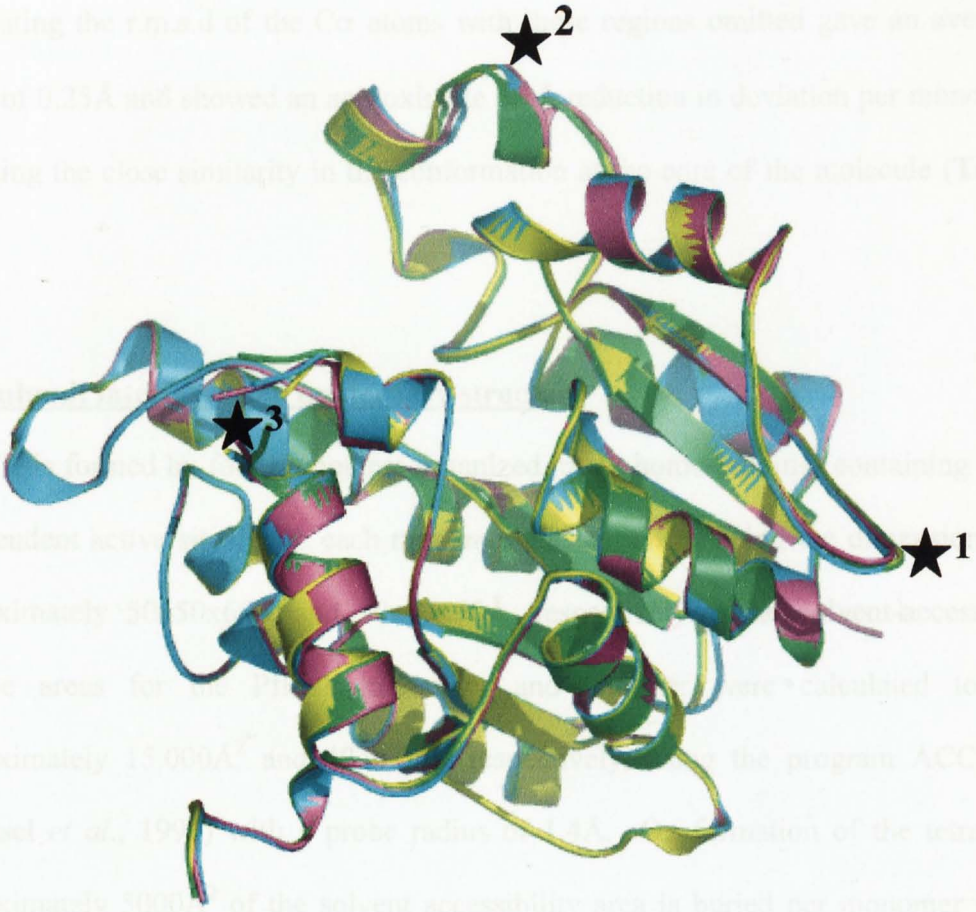


Figure 4.8. The superposition of subunit B, C and D onto the A subunit of PfENR, coloured green, blue, purple and yellow for subunits A, B, C, D respectively. Those areas where significant differences occur are labelled by a star and are numbered 1 (Asp^{71} – Met^{76}), 2 (Thr^{90} – Asn^{109}) and 3 (Thr^{281} – Glu^{288}). The figure was produced using the graphics program Pymol (Delano, 2002).

Table 4.2 a

	A	B	C	D
A		0.4	0.3	0.3
B	0.4		0.3	0.4
C	0.3	0.3		0.3
D	0.3	0.4	0.3	

Table 4.2 b

	A	B	C	D
A		0.3	0.2	0.2
B	0.3		0.3	0.3
C	0.2	0.3		0.2
D	0.2	0.3	0.2	

Table 4.2. a, b. A set of tables to show the r.m.s.d in Å of the subunits A, B, C and D superimposed onto each other. The columns across and below represent each subunit labelled A to D in red. (a) The r.m.s.d difference on each of the monomers with no residues removed. (b) The r.m.s.d deviation between the subunits after the removal of the three areas shown to differ significantly between the four subunits (Asp^{71} – Met^{76}, Thr^{90} – Asn^{109} and Thr^{281} – Glu^{288}).

calculating the r.m.s.d of the Cα atoms with these regions omitted gave an average value of 0.25Å and showed an approximate 0.1Å reduction in deviation per monomer revealing the close similarity in the conformation at the core of the molecule (**Table 4.2**).

4.5 Subunit interfaces and quaternary structure

PfENR is formed by four monomers organized into a homo-tetramer containing four independent active sites, with each monomer and tetramer having the dimensions of approximately 50x50x60Å and 90x90x50Å, respectively. The solvent-accessible surface areas for the PfENR monomer and tetramer were calculated to be approximately 15,000Å^2 and 40,000Å^2, respectively, using the program ACCESS (Roussel *et al.*, 1990) with a probe radius of 1.4Å. On formation of the tetramer approximately 5000Å^2 of the solvent accessibility area is buried per monomer with each of the monomers within the tetramer making contact with all three of the other subunits. The PfENR tetramer has three different inter-subunit interfaces about three orthogonal twofold symmetry axis designated as P, Q and R (**Figure 4.9**). Contacts between subunits A-C and B-D are found around the P axis, between subunits A-B and C-D around the Q axis and between subunits A-D and B-C around the R axis. The program CONTACT (CCP4, 1994) together with a visual inspection was used to examine the subunit interfaces.

4.5.1 The P axis

The P-axis is the most extensive of the subunit interfaces with approximately 2400Å^2 of the solvent accessible area being buried on formation of the tetramer and is formed by 44 residues (Glu^{35}, Arg^{39}, Arg^{210}, Ala^{213} to His^{215}, Gly^{217}, Arg^{218}, Asn^{221}, Ile^{222},

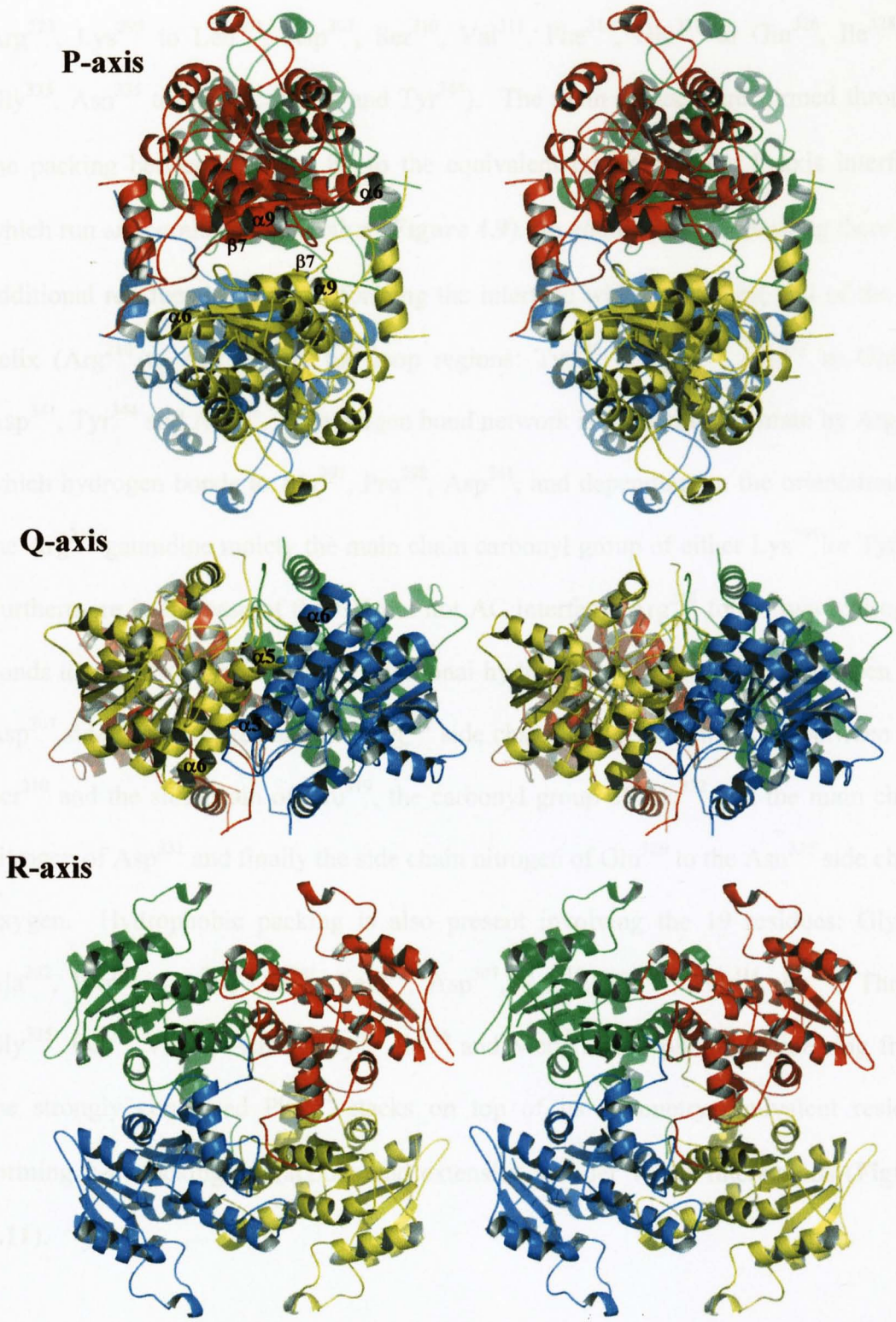


Figure 4.9. A stereo view of the PfENR tetramer viewed down each of the three molecular two fold axes. The subunits A, B, C and D are coloured yellow, blue, red and green, respectively and each of the three axes are labelled. The secondary structural elements which are important in producing major subunit contacts are labelled. The figure was created using the graphics program Pymol (Delano, 2002).

Arg^{223}, Lys^{295} to Leu^{304}, Asp^{307}, Ser^{310}, Val^{311}, Phe^{314}, Glu^{319} to Gln^{326}, Ile^{328} to Gly^{333}, Asn^{335} to Met^{337}, Asp^{341} and Tyr^{344}). The main contacts are formed through the packing between α9 and β7 to the equivalent residues in the P axis interface which run antiparallel to each other (**Figure 4.9**). In addition to this packing there are additional residues involved in forming the interface which are at the end of the α6 helix (Arg^{210} to Arg^{223}) and the loop regions: Tyr^{295} to Leu^{304}, Glu^{319} to Gln^{326}, Asp^{341}, Tyr^{344} and Arg^{345}. A hydrogen bond network is formed in the main by Arg^{218}, which hydrogen bonds to Ala^{297}, Pro^{298}, Asp^{341}, and depending on the orientation of the Arg^{218} gaunidine moiety the main chain carbonyl group of either Lys^{295} or Tyr^{296}. Furthermore in the case of the BD but not AC interface, Arg^{218} forms two hydrogen bonds to Arg^{300} (**Figure 4.10a**). Additional hydrogen bonds are formed between the Asp^{307} side chain oxygen and the Arg^{321} side chain nitrogen atom. Also between the Ser^{310} and the side chain of Glu^{319}, the carbonyl group of Ala^{322} and the main chain nitrogen of Asp^{331} and finally the side chain nitrogen of Gln^{326} to the Asn^{335} side chain oxygen. Hydrophobic packing is also present involving the 19 residues: Gly^{217}, Ala^{232}, Pro^{298}, Leu^{299}, Gln^{301}, Leu^{303}, Asp^{307}, Ser^{310}, Val^{311}, Phe^{314}, Ile^{323}, Thr^{324}, Gly^{325}, Ile^{328}, Val^{330}, Asn^{332}, Gly^{333} Ile^{336} and Met^{337}. The phenylalanine ring from the strongly conserved Phe^{314} stacks on top of its symmetry equivalent residue forming π-π stacking interactions and extensive van der Waals interactions (**Figure 4.11**).

4.5.2 The Q axis

The interface down the Q axis is dominated by a four helix bundle composed of the α5 and α6 helices and their respective symmetry equivalents. Looking down the axis of the AB interface the first interaction is between the $\alpha5^{A}$ helix and preceding loop

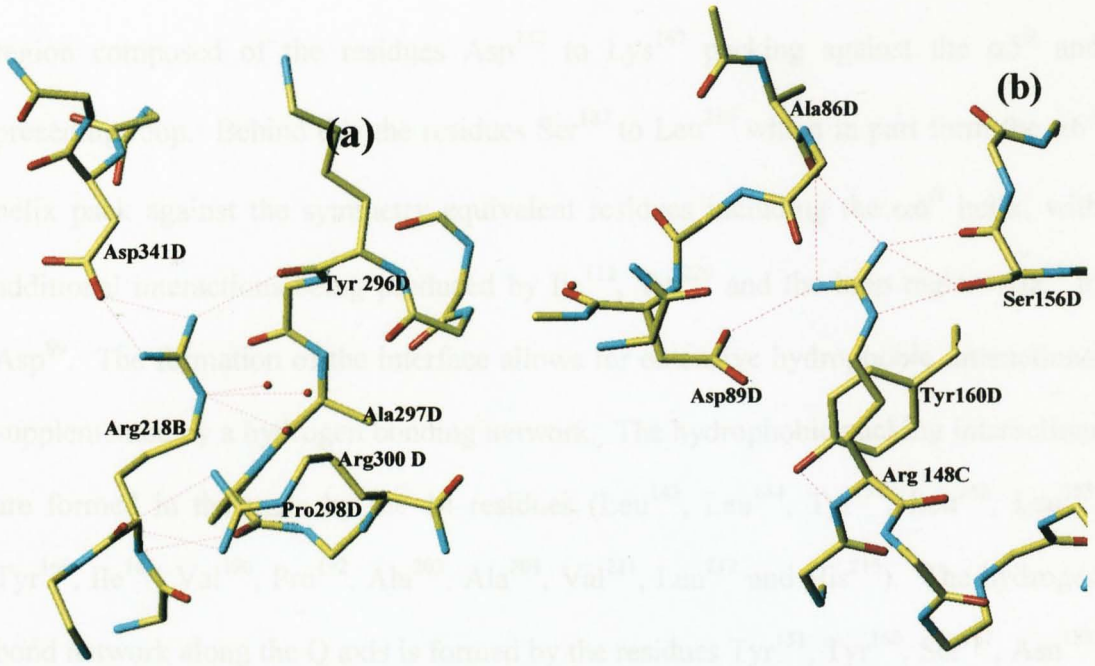


Figure 4.10. (a) The hydrogen bonding network created by the Arg^{218B} residue to the Asp^{341}, Tyr^{296}, Ala^{297}, Pro^{298} and Arg^{300} of subunit D forming part of the P-axis interface. (b) The hydrogen bonding network which forms part of the Q axis interface in which Arg^{148} from subunit C is hydrogen bonded to Asp^{89}, Ala^{86}, Ser^{156} and Tyr^{160} of subunit D. For both figure hydrogen bonds are represented by red dashed lines, all residues are coloured with yellow for carbon, red for oxygen and blue for nitrogen. Water atoms are represented by red spheres. Those residues involved in forming the hydrogen bonding network are labelled. The figure was produced using the graphics program TURBO/FRODO (Roussel *et al.*, 1990)

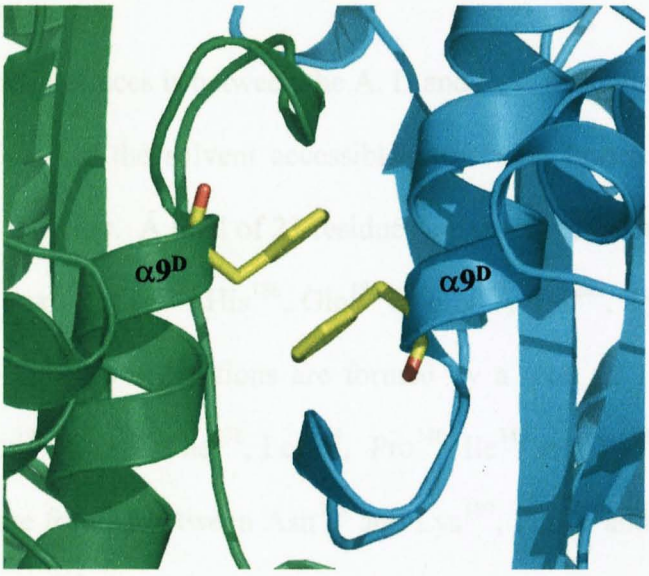


Figure 4.11. A graphical representation of the packing interface between the PfENR subunits B and D, coloured green and blue respectively. The Phe^{314} residue is shown with yellow for carbon and red for oxygen. The figure was produced using the graphics program Pymol (Delano, 2002).

region composed of the residues Asp^{142} to Lys^{167} packing against the $\alpha5^{B}$ and preceding loop. Behind this the residues Ser^{187} to Leu^{216} which in part form the $\alpha6^{A}$ helix pack against the symmetry equivalent residues including the $\alpha6^{B}$ helix, with additional interactions being produced by Ile^{112}, Tyr^{220} and the loop region Ala^{86} to Asp^{89}. The formation of the interface allows for extensive hydrophobic interactions, supplemented by a hydrogen bonding network. The hydrophobic packing interactions are formed in the main by the 14 residues (Leu^{143}, Leu^{144}, Tyr^{151}, Leu^{152}, Leu^{155}, Tyr^{160}, Ile^{163}, Val^{190}, Pro^{192}, Ala^{203}, Ala^{204}, Val^{211}, Leu^{212} and His^{215}). The hydrogen bond network along the Q axis is formed by the residues Tyr^{151}, Tyr^{160}, Ser^{187}, Asn^{188}, Asp^{208}, Arg^{210} and Asn^{218}. Furthermore the Arg^{148} residue forms up to 7 hydrogen bonds to the neighbouring Ser^{156}, Ala^{86}, Asp^{89} and Tyr^{160} with some residues forming more than one hydrogen bond to Arg^{148} (**Figure 4.10b**). On formation of the interface approximately 1700Å^2 of the solvent accessible area is buried per monomer.

4.5.3 The R axis

The smallest of the interfaces is between the A, D and B, C subunits (**Figure 4.9**) with approximately 900Å^2 of the solvent accessible area being buried per monomer on production of the tetramer. A total of 22 residues are involved in the formation of the R axis interface: Asp^{142}, Tyr^{184}, His^{186}, Gln^{188} to Gly^{193}, Tyr^{288}, Tyr^{292} and His^{334} to Tyr^{344}. The hydrophobic interactions are formed by a total of 10 residues Val^{190}, Val^{191}, Pro^{192}, Gly^{193}, Leu^{334}, Phe^{338}, Leu^{339}, Pro^{340}, Ile^{343} and Tyr^{344}. A total of five hydrogen bonds are formed between Asn^{335} and Lys^{189}, Phe^{340} and Tyr^{288} and Val^{190} to both Ile^{336} and Phe^{338}.

4.6 Cofactor and inhibitor binding sites

4.6.1 Triclosan binding site

The electron density for the triclosan inhibitor within the PfENR/NAD^+/triclosan complex was of very good quality and allowed for its unambiguous placement into the active site in all four subunits (**Figure 4.12a**). The triclosan binds in a non-covalent manner to the active site, forming one hydrogen bond to Tyr^{184} and extensive van der Waals interactions. The 4-chloro phenoxy ring of triclosan (ring A) interacts in a face to face manner with the NAD^+ nicotinamide unit allowing π-π stacking interactions and the formation of van der Waals interactions with the side chains of Tyr^{184}, Tyr^{194}, Pro^{231}, Phe^{285}, Ala^{237}, Ile^{240} and Ile^{286}. The 2,4-dichlorophenoxy ring of triclosan (ring B) is located within a pocket bounded by the peptide backbone of residues Leu^{133} to Ala^{136}, the pyrophosphate and nicotinamide moieties of NAD^+ and by the side chains of Ala^{236}, Met^{198} and Val^{139} (**Figure 4.12b**). It has been shown using the *E.coli* ENR that triclosan binding is reversible and that neither the NAD^+ or triclosan are modified during inhibition, consistent with the non-covalent binding seen within the active site (Ward *et al.*, 1999).

4.6.2 NAD^+ binding site

In all four subunits of the PfENR model the nicotinamide adenine dinucleotide (NAD^+) cofactor can be clearly seen within the electron density permitting its unambiguous identification (**Figure 4.13a**). The adenine ring binds within a pocket formed by the main chain atoms of residues: Ala^{20} to Gly^{22}, Ser^{132} to Asn^{135}, Phe^{84} to Ser^{87} and Ala^{229} to Ala^{234} and the side chains of residues: Trp^{48}, Leu^{133} and Asn^{135}. The adenine unit interacts with the Trp^{48} side chain in a face to face manner allowing π–π stacking interactions, with Leu^{133} making van der Waals interactions with the

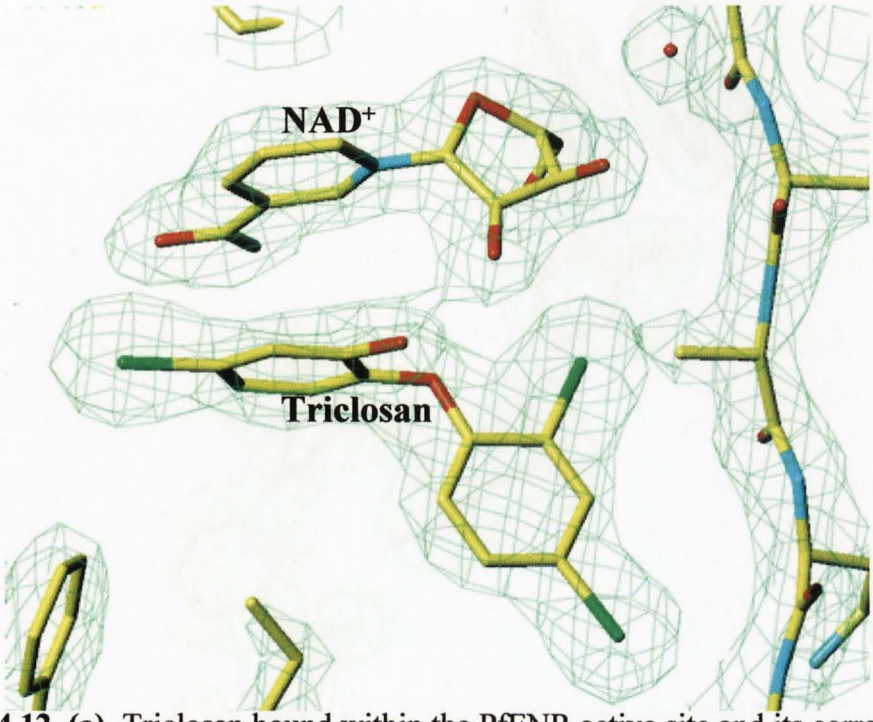


Figure 4.12. (a) Triclosan bound within the PfENR active site and its corresponding $2F_{obs}$-F_{calc} electron density map contoured at 1.0σ. The NAD^+ cofactor can also be seen packing in a face to face manner with the triclosan inhibitor. The model is coloured with yellow for carbon, blue for nitrogen, red for oxygen and green for chlorine. This figure is based on subunit C of PfENR. The figure was produced using the program TURBO-FRODO (Roussel *et al.*, 1990).

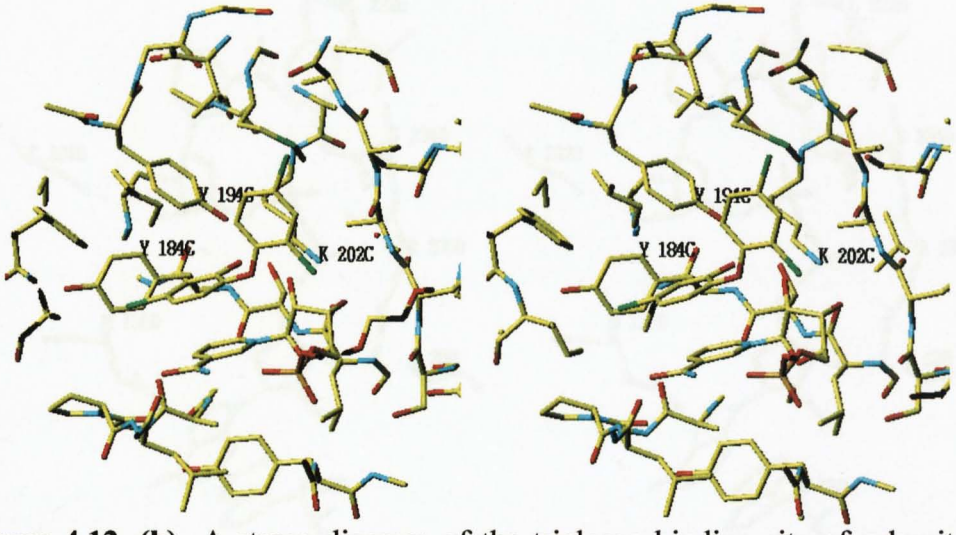


Figure 4.12. (b) A stereo diagram of the triclosan binding site of subunit C of PfENR with the active site residues Tyr^{191}, Tyr^{201}, Lys^{209} and Met^{205} labelled. The figure is coloured with yellow for carbon, red for oxygen, orange for phosphorous, blue for nitrogen and green for the chlorine atoms of triclosan and the sulphur atom of the active site Met^{205}. The figure was produced using the program TURBO-FRODO (Roussel *et al.*, 1990).

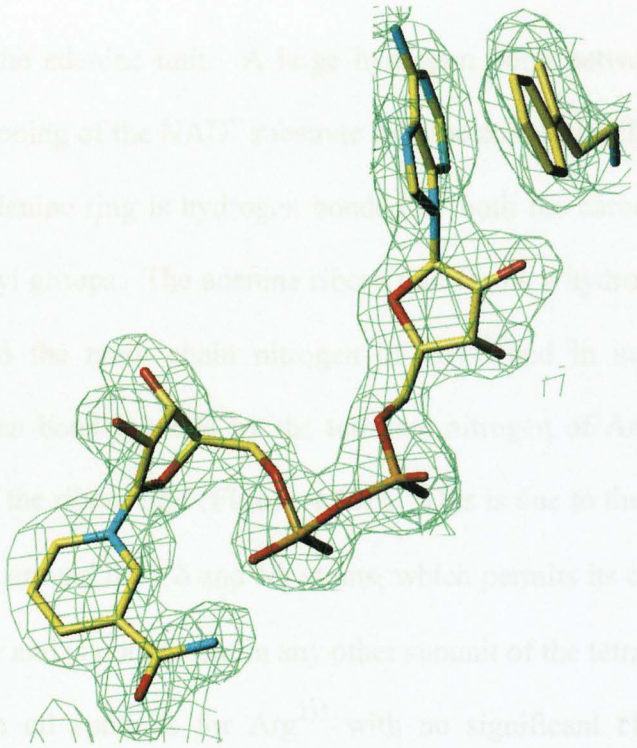


Figure 4.13. (a) The NAD^+ cofactor bound within the PfENR active site and its corresponding $2F_{obs}$-F_{calc} electron density map contoured at 1.0σ. The model is coloured with yellow for carbon, blue for nitrogen, red for oxygen and orange for phosphorous. The figure was produced using the program TURBO-FRODO (Roussel *et al.*, 1990).

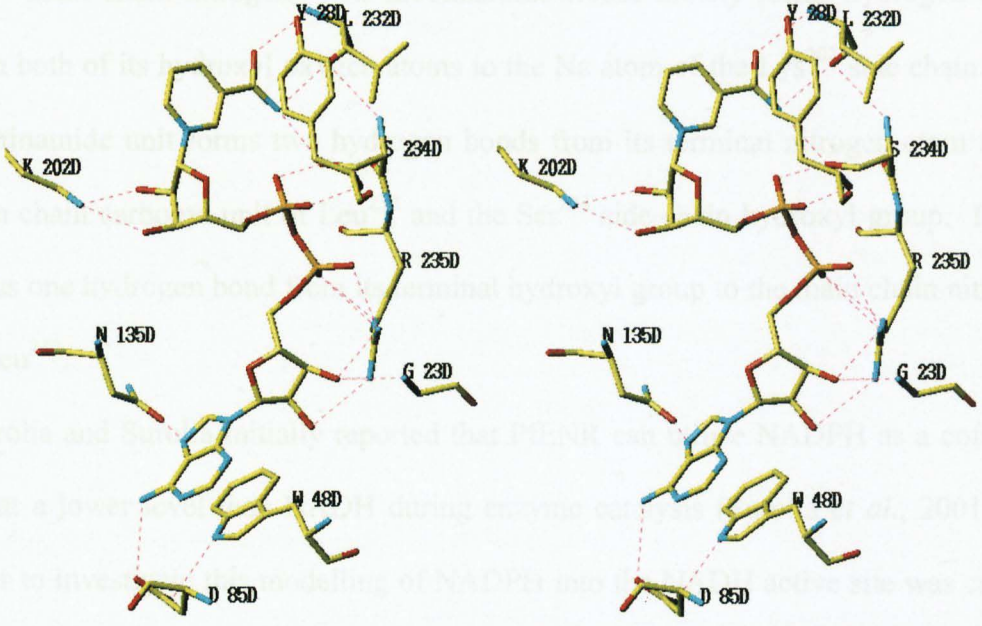
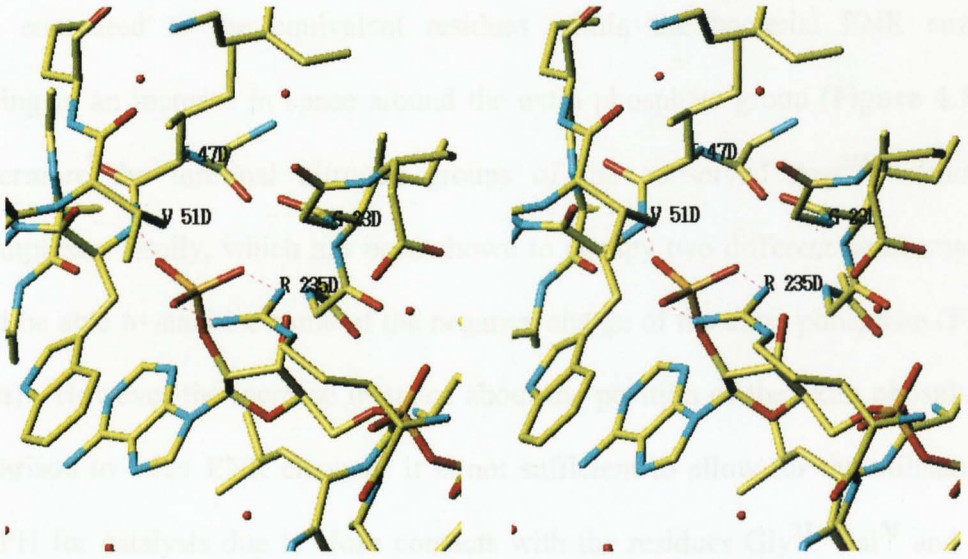


Figure 4.13. (b) A stereo diagram of the residues responsible for forming a hydrogen bonding network to the NAD^+ cofactor. All the residues are labelled and coloured with yellow for carbon, blue for nitrogen, red for oxygen and orange for phosphorous. The Arg^{242} residue is only present in a hydrogen bonding position in the D subunit and does not form hydrogen bonds in any of the other subunits. The figure was produced using the program TURBO-FRODO (Roussel *et al.*, 1990).

opposite face of the adenine unit. A large hydrogen bond network allows for the binding and positioning of the NAD^+ substrate into the active site (**Figure 4.13 b**). In all subunits the adenine ring is hydrogen bonded by both the carbonyl of Asp^{85} and the Asn^{135} hydroxyl groups. The adenine ribose unit forms a hydrogen bond from its hydroxyl group to the main chain nitrogen of Gly^{23} and in subunit D only an additional hydrogen bond is made by the terminal nitrogen of Arg^{235D} to the same hydroxyl group of the ribose unit (**Figure 4.13b**). This is due to the movement of the Arg^{235D} side chain around the Cδ and Nε atoms, which permits its closer contact with the NAD^+ cofactor and is not present in any other subunit of the tetramer. The density is well defined in all subunits for Arg^{235} with no significant change in B-factor between the two conformations. The pyrophosphate group is involved in hydrogen bonding to the main chain nitrogen of Ala^{236}, the Ser^{234} hydroxyl side chain and the Tyr^{28} main chain nitrogen. The nicotinamide ribose moiety forms hydrogen bonds from both of its hydroxyl oxygen atoms to the Nε atom of the Lys^{202} side chain. The nicotinamide unit forms two hydrogen bonds from its terminal nitrogen atom to the main chain carbonyl unit of Leu^{232} and the Ser^{234} side chain hydroxyl group. It also forms one hydrogen bond from its terminal hydroxyl group to the main chain nitrogen of Leu^{232}.

Surolia and Surolia initially reported that PfENR can utilise NADPH as a cofactor, but at a lower level than NADH during enzyme catalysis (Surolia *et al.*, 2001). In order to investigate this modelling of NADPH into the NADH active site was carried out (**Figure 4.14 a**). Interestingly a sequence change of alanine to glycine found within the PfENR sequence but not in the equivalent position of the plant or bacterial ENR enzymes produces an increase in pocket size around the position of the extra phosphate (**Figure 4.14 b**). In addition those residues which form the lid of the

(a)



(b)

E.coli	A	X_{23}	Y	Q	N	D	K
H.pylori	A	X_{23}	Y	L	N	E	S
B.subtilus	A	X_{23}	Y	A	G	E	R
S.aureus	A	X_{23}	Y	L	P	D	E
B.napus	A	X_{23}	T	W	V	P	A
P.falciparum	G	X_{23}	I	W	P	P	V
	23		47				51

Figure 4.14. (a) A stereo diagram of the modelling of NADPH into the PfENR active site. The extra phosphate group can be seen to protrude into a cavity formed in the main by the side chains of Val^{51} and Ile^{47}. The side chain of Arg^{235} has been moved within allowable $\chi 1$-$\chi 2$ values to maximise the hydrogen bonding potential of this residue but this does not represent its position within the PfENR structure. The diagram is coloured with yellow for carbon, blue for nitrogen, red for oxygen and orange for phosphorous. The hydrogen bonds are represented by dashed red lines. The diagram was produced using the graphics program TURBO/FRODO (Roussel *et al.*, 1990).

(b) A table to show the equivalent residues found within the *E.coli*, *H.pylori*, *B.subtilis*, *S.aureus* and *B.napus* ENR enzymes at the site of the extra phosphate compared to PfENR. The residues are numbered with respect to the *P.falciparum* sequence and are also shown in **Figure 4.11a**. The amino acids are given in one letter code and X_{23} represents a 23 amino acid gap.

binding pocket of the extra phosphate within the PfENR structure are relatively small when compared to the equivalent residues within the bacterial ENR enzymes resulting in an increase in space around the extra phosphate group (**Figure 4.14 b**). Furthermore the terminal nitrogen groups of the conserved Arg^{235}, within the apicomplexan family, which has been shown to occupy two different conformations, would be able to stabilise some of the negative charge of the extra phosphate (**Figure 4.14 a**). However the increase in space about the position of the extra phosphate in comparison to other ENR enzymes it is not sufficient to allow for the utilisation of NADPH for catalysis due to close contacts with the residues Gly^{23}, Val^{51} and Trp^{48} which are approximately 2.4Å, 2.5Å and 3.0Å away from the extra phosphate group. This is consistent with more recent studies which found no activity was present for PfENR in the presence of NADPH (Perrozo *et al.*, 2002).

4.7 Substrate modelling and enzyme mechanism

Substrate binding to PfENR was investigated by modelling, based on both the MtENR/NADH/C16 fatty acid substrate complex and the previously published modelling studies based on butyryl-CoA (Roujeinikova[a] *et al.*, 1999). In order to model the butyryl-CoA into the PfENR active site the coordinates were taken from the X-ray crystallographic structure of acyl-CoA dehydrogenase (Djordjevic *et al.*, 1995; PDB entry 1BUC). The double bond between the C2 and C3 position of butyryl-CoA was superimposed upon the C5 and C6 positions in the phenolic ring of triclosan, with the triclosan mimicking the enolate anion intermediate (Roujeinikova[a] *et al.*, 1999). The growing fatty acid chain is positioned in the base of the active site formed by β5 and β6 and the helices α6 and α8, the substrate exits the binding pocket and protrudes into the solvent (**Figure 4.15 a and b**). The binding pocket is formed by the residues:

Trp^{48}, Glu^{97}, Ala^{134}, Asn^{135}, Ala^{136}, Lys^{137}, Val^{139}, Lys^{149}, Tyr^{184}, Tyr^{194}, Met^{198}, Lys^{202}, Pro^{231}, Ala^{236}, Ala^{237}, Ala^{239}, Ile^{240}, Phe^{285}, Ile^{286} with hydrogen bonds being made to the fatty acid substrate by Asn^{135}, Glu^{97} and Lys^{137}. The residues producing this binding pocket are predominantly hydrophobic and are strongly conserved between the *P.falciparum* and *B.napus* ENR sequences with full sequence conservation of all but Ala^{136} and Lys^{137}. However a lower level of sequence conservation is observed for those residues which make up the binding pocket within the bacterial sequences of *E.coli, H.pylori, B.subtilis* and *M.tuberculosis* with approximately 30% sequence identity although they are predominantly hydrophobic.

Substrate modelling into the active site of PfENR showed the proposed distance between the C4 position of the nicotinamide ring and the C3 position of the modelled substrate, between which hydride transfer occurs, to be approximately 3.4Å. The phenolic oxygen atom of the catalytic Tyr^{194} is approximately 2.4Å away from the enolate anion, consistent with the tyrosine residue being the proton donor to the enolate anion. The catalytic Lys^{202} residue is positioned within 4.5Å of the substrate and may allow its positive charge to play a role in stabilising the intermediate of the reaction (**Figure 4.16**). This is consistent with the proposed catalytic mechanism of the ENR enzyme (Section **1.2.3**).

Modelling of the C16 fatty acid substrate in a "U" shaped conformation as described for the MtENR enzyme ((**Figure 4.15**)(Rozwarski *et al.*, 1999)) demonstrated that without increasing the size of the binding pocket there is not sufficient room for the substrate to bind in the PfENR/NAD^+/triclosan complex (**Figure 4.15c**). However of those residues which form close contacts with the substrate, in particular Tyr^{184}, Ala^{186} and Phe^{285}, both the Ala and Phe residues are situated in an area predicted to be flexible within the ENR structure (Roujeinikova[b] *et*

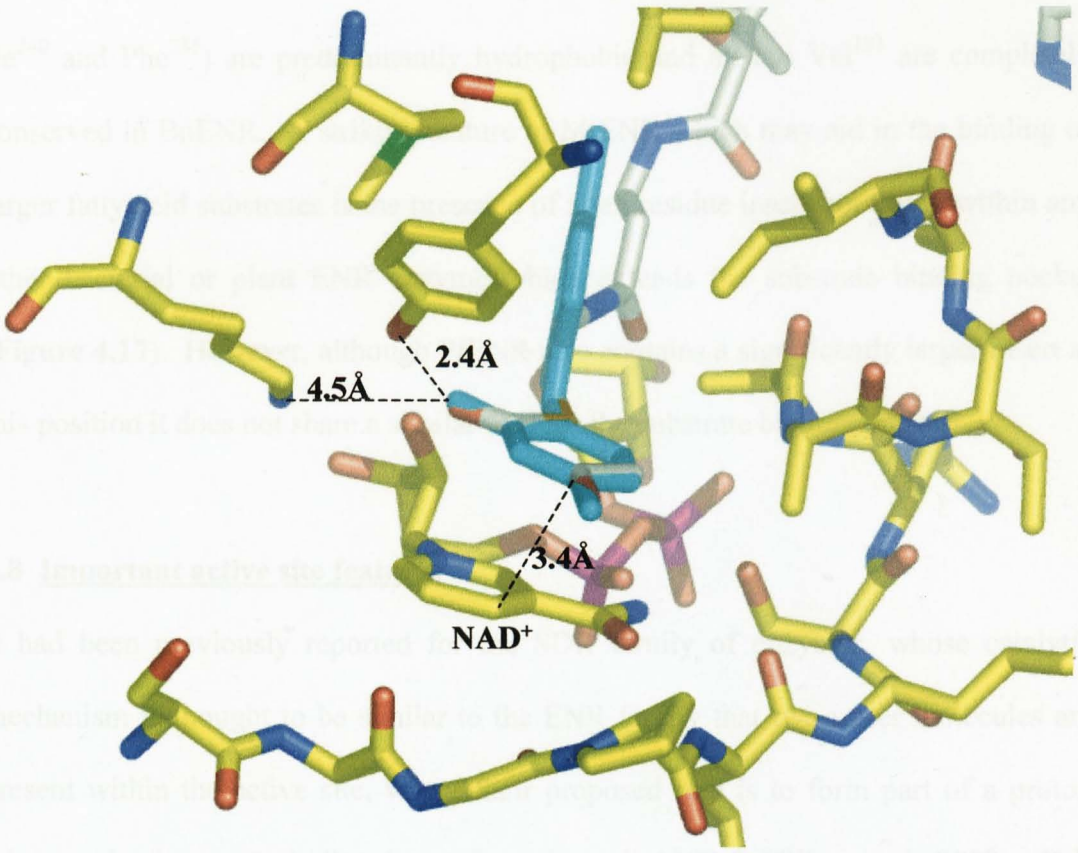


Figure 4.16. Modelling of the butyryl-CoA substrate into the active site of PfENR showing the proximity of the critical catalytic residues. The distances from the substrate to the active site residues are in Å and shown by a black dashed line. The C2 and C3 positions of the substrate are superimposed upon the C5 and C6 atoms of the triclosan phenolic ring. The triclosan inhibitor is shown in blue, the fatty acid substrate is coloured white for carbon, blue for nitrogen and red for oxygen. The PfENR enzyme is coloured with yellow for carbon, blue for nitrogen, red for oxygen, green for sulphur and purple for phosphorous. The figure was produced using the graphics program Pymol (Delano, 2002).

al., 1999). Those residues involved in forming the binding pocket (Ala^{134}, Asn^{135}, Val^{139}, Tyr^{184}, Ala^{186}, Val^{191}, Pro^{192}, Gly^{193}, Tyr^{194}, Met^{198}, Lys^{202}, Ala^{236}, Ala^{237}, Ile^{240} and Phe^{285}) are predominantly hydrophobic and all but Val^{191} are completely conserved in BnENR. A striking feature of MtENR which may aid in the binding of larger fatty acid substrates is the presence of a ten residue insert not found within any other bacterial or plant ENR enzyme which extends the substrate binding pocket (**Figure 4.17**). However, although PfENR also contains a significantly larger insert at this position it does not share a similar fold for the substrate binding pocket.

4.8 Important active site features

It had been previously reported for the SDR family of enzymes, whose catalytic mechanism is thought to be similar to the ENR family that two water molecules are present within the active site, where their proposed role is to form part of a proton relay mechanism to the bulk solvent (Benach *et al.*, 1998 & Filling *et al.*, 2002). This proton relay system may act as a proton bridge linking the catalytic Lys residue to the bulk solvent, creating a temporary proton reservoir during catalysis. In order to investigate a possible role for water molecules within the ENR family, the positions of any active site waters were investigated. Within the PfENR active site two water molecules are present which are clearly defined by density at an equivalent positions to the SDR enzymes. The position of these water molecules is such that the terminal hydroxyl atoms at the Nε group of Lys^{202} make a hydrogen bond to the first of the water molecules, W^{29} as well as to both of the oxygen atoms of the nicotinamide ribose group and the main chain carbonyl of Leu^{182}, forming a hydrogen bond network (**Figure 4.18**). The second of the water molecules is positioned 3.5Å away from W^{29} and forms a hydrogen bond to the main chain nitrogen of Leu^{182}. These two

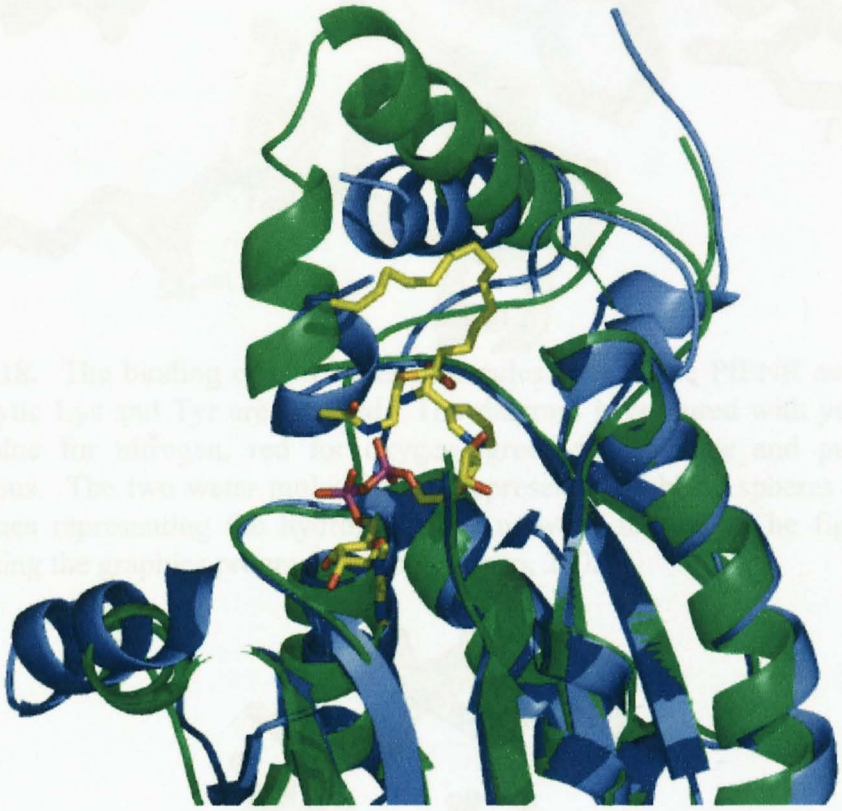


Figure 4.17. The structures of the PfENR/NAD^+/triclosan complex in blue and MtENR/NADH/C-16 fatty acid substrate complex in green. The fatty acid can be seen in a "U" shaped conformation within the binding pocket with the NAD^+ cofactor underneath, both are coloured with yellow for carbon, red for oxygen, blue for nitrogen and purple for phosphorous. The substrate binding helix which lies on top of the helix can be seen to adopt different positions for the two enzymes. The figure was produced in Pymol (Delano, 2002).

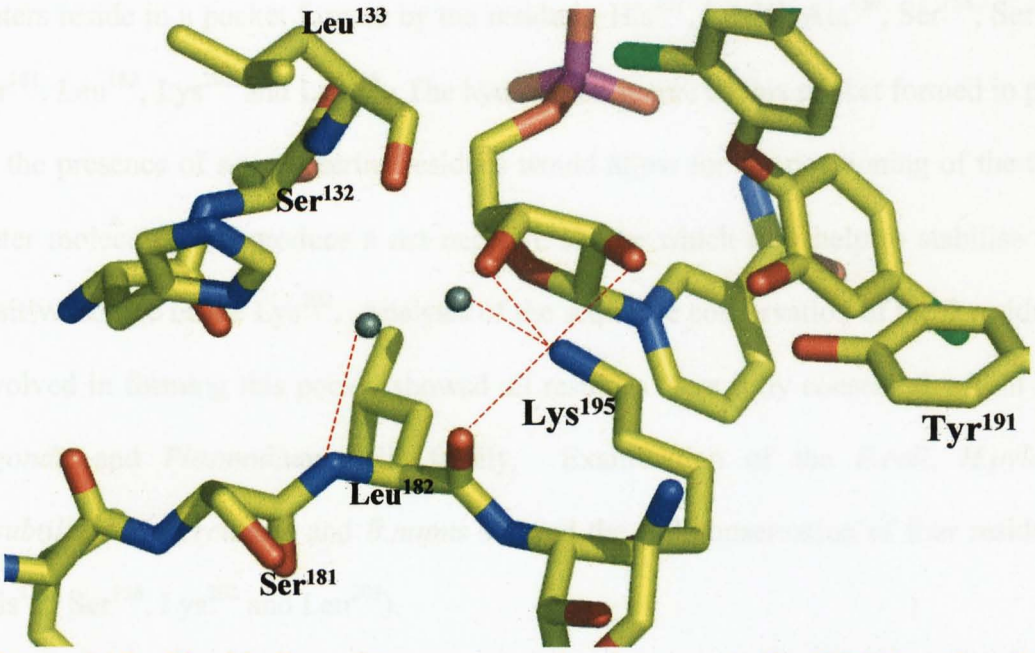


Figure 4.18. The binding of two water molecules near to the PfENR active site. The catalytic Lys and Tyr are labelled. The diagram is coloured with yellow for carbon, blue for nitrogen, red for oxygen, green for chlorine and purple for phosphorous. The two water molecules are represented as black spheres with red dashed lines representing the hydrogen bond network formed. The figure was created using the graphics program Pymol (Delano, 2002).

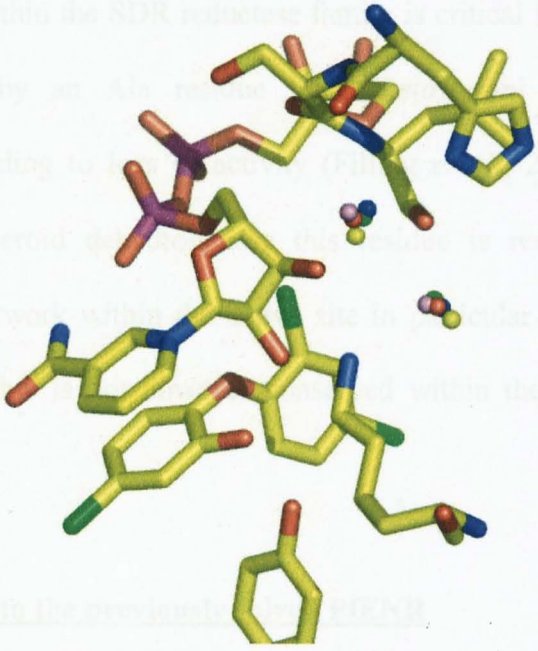


Figure 4.19. The PfENR active site with bound triclosan and NAD^+. The superimposed water molecules are coloured green, magenta and red representing the *E.coli*, *B.napus* and *P.falciparum* complexed with NAD^+ and triclosan respectively. The yellow dots represent the active site waters of *M.tuberculosis* ENR complexed with NADH and the blue represents the *M.tuberculosis* ENR complexed with fatty acid substrate and NAD^+. The figure was created using the graphics program Pymol (Delano, 2002).

waters reside in a pocket formed by the residues: His^{131}, Leu^{133}, Ala^{134}, Ser^{158}, Ser^{159}, Ser^{181}, Leu^{182}, Lys^{202} and Leu^{205}. The hydrophilic nature of this pocket formed in part by the presence of several serine residues would allow for the positioning of the two water molecules and produce a net negative charge which may help to stabilise the positive charge of the Lys^{202}. Analysis of the sequence conservation of the 9 residues involved in forming this pocket showed all residues were fully conserved within the *T.gondii* and *Plasmodium* ENR family. Examination of the *E.coli*, *H.pylori*, *B.subtilis*, *M.tuberculosis* and *B.napus* showed the full conservation of four residues (His^{131}, Ser^{158}, Lys^{202} and Leu^{205}).

The superposition of the high resolution structures of ENR from *B.napus*, *M.tuberculosis* and *E.coli* shows that the position of these two water molecules is conserved (**Figure 4.19**). Interestingly it has been reported that a conserved Asn residue, present within the SDR reductase family is critical for enzyme catalysis with the replacement by an Ala residue in *C.Testosteroni* 3β/17β Hydroxysteroid dehydrogenase leading to loss of activity (Filling *et al.*, 2001). In *C.Testosteroni* 3β/17β Hydroxysteroid dehydrogenase this residue is responsible for forming a hydrogen bond network within the active site in particular to one of the active site waters. This residue is not however conserved within the ENR family or present within the PfENR.

4.9 Comparisons to the previously solved PfENR

During the structure solution of PfENR as reported within this chapter another group also solved and published the structure for PfENR in complex with NAD^+ and both triclosan and two triclosan analogues (Perozzo *et al.*, 2002). The superimposition of their structure with that described here showed that they are very similar especially in

the mode of triclosan and NAD^+ binding. Significant differences between the two structures occur at the final three residues of the N terminal region and between residues Asp^{93} to Leu^{107} (**Figure 4.20**). The large polar 42 amino acid insert could not be assigned in either crystal structure and would suggest that this loop does not adopt an ordered fold within either crystal structure.

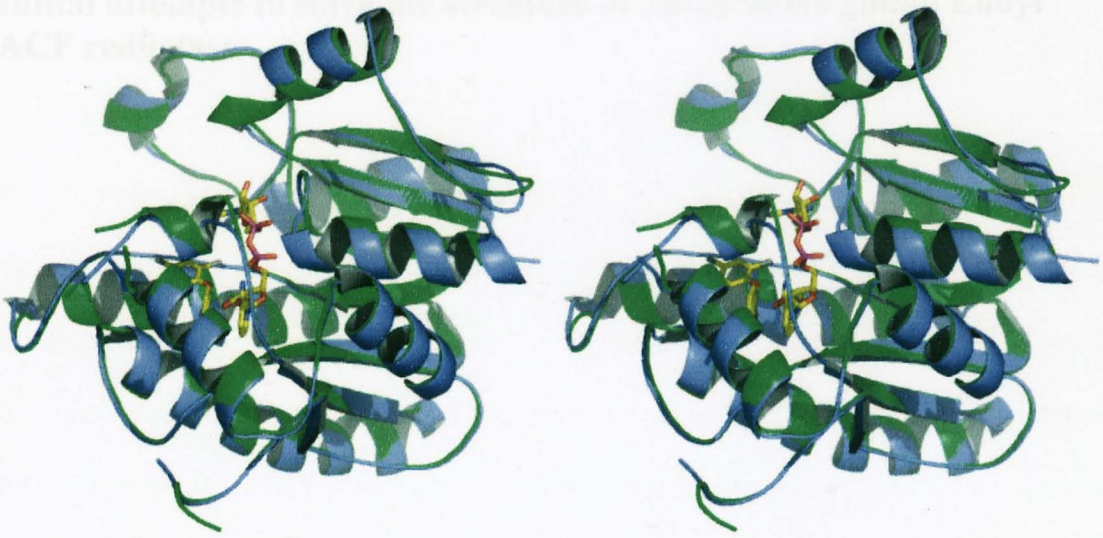


Figure 4.20. A stereo diagram of the superposition of PfENR as reported in this chapter and as solved by Perozzo *et al.*, 2002, coloured green and blue, respectively. The figure was produced using Pymol (Delano, 2002).

CHAPTER 5

Initial attempts to solve the structure of *Toxoplasma gondii* Enoyl ACP reductase

5.1 Introduction

This chapter outlines the initial attempts to clone, overexpress, purify, crystallise and solve the structure of *Toxoplasma gondii* enoyl acyl carrier protein reductase (TgENR).

5.2 Location and sequence analysis of the TgENR gene

In order to identify and characterise TgENR, a cDNA library was screened using ENR DNA sequences from *P.falciparum* by the collaborating group of Prof Rima McLeod (Samuel *et al.*, 2003). Analysis of the sequence showed it to be similar to *B.napus* and *P.falciparum* ENR to which it showed approximately 57 and 42% sequence identity, respectively (**Figure 5.1**). In addition a 27 amino acid signal sequence was identified using SIGNALP V1.1 as well as a 66 residue chloroplast transit peptide sequence using CHLOROP 1.1. Analysis of a structure based sequence alignment showed not only strong sequence identity around the active site region, but also the presence of a smaller six residue insert at an equivalent position to the 42 amino acid seen between the *B.napus* and *P.falciparum* enzymes.

5.3 Overexpression and purification of TgENR

A construct of TgENR containing residues 103–417 (lacking the putative signal and transit peptides) was over expressed and purified in the same manner as that reported for PfENR by the collaborating group of Prof Sean Prigge (Section 3.4, Muench *et al.*, 2003). The TgENR sequence was renumbered such that the first residue after the 6 His tag which was part of the mature TgENR sequence became residue 1 (Ser^1) (**Figure 5.1**).

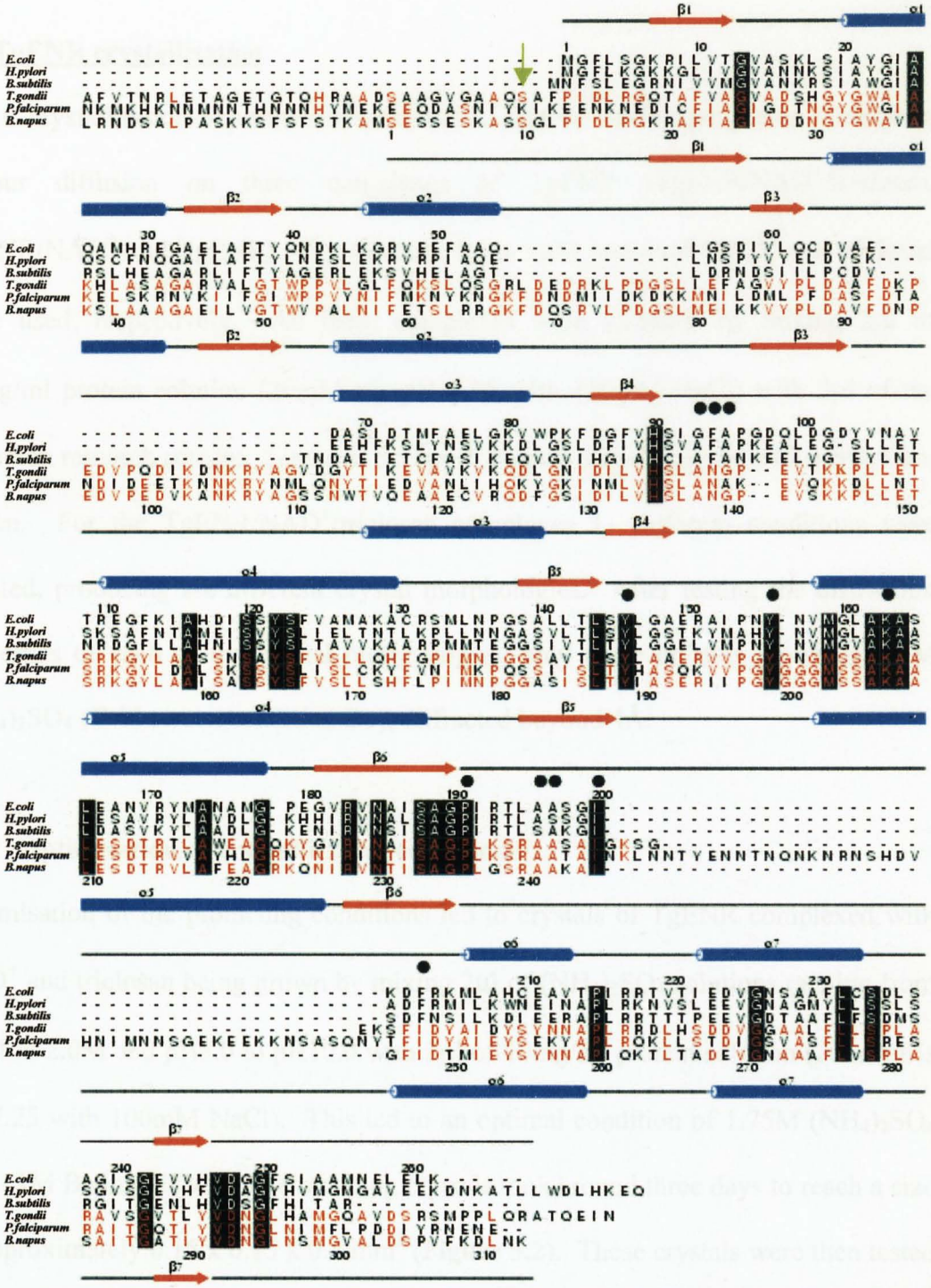


Figure 5.1. A multiple structure-based alignment of the ENR enzyme from *E.coli*, *H.pylori*, *B.subtilis*, *T.gondii*, *P.falciparum* and *B.napus*. The secondary structural elements shown as red arrows and blue cylinders for the β-strands and α-helices, respectively and sequence numbers for the *E.coli* and *B.napus* ENR are given on top and below the alignment, respectively. Those residues completely conserved are in reverse type and those involved in triclosan binding have a black circle above. Residues conserved between the *B.napus, T.gondii* and *P.falciparum* ENR enzyme are in red. The start of the mature TgENR sequence used for crystallographic studies is indicated by a green arrow. The sequences were aligned using ClustalW (Higgins & Thompson 1994) and the figure was produced using the program Alscript (Barton, G.J. 1993).

5.4 TgENR crystallisation

Initial crystallisation experiments were carried out by the hanging drop method of vapour diffusion on three complexes of TgENR (TgENR/NAD^+/triclosan, TgENR/NAD^+ and TgENR). For the complexes 5μM and 6μM NAD^+ and triclosan were used, respectively. All three complexes were screened by mixing 2μl of 20mg/ml protein solution (20mM tris pH 7.25 with 100mM NaCl) with 2μl of the Hampton research screens, Screen 1, Screen 2, PegIon Screen and the $(NH_4)_2SO_4$ Grid screen. For the TgENR/NAD^+/triclosan complexes 11 different conditions were isolated, producing six different crystal morphologies. After testing the diffraction properties of each of the crystal morphologies, only those crystals grown in 1.6M $(NH_4)_2SO_4$ (Grid screen condition B6), diffracted beyond 4Å.

5.4.1 Optimisation of crystallisation conditions

Optimisation of the promising conditions led to crystals of TgENR complexed with NAD^+ and triclosan being grown by mixing 2μl of $(NH_4)_2SO_4$ solutions ranging from 1.0M to 2.0M and pH8.0 to pH10.0 with 2μl of 20mg/ml protein solution (20mM tris pH 7.25 with 100mM NaCl). This led to an optimal condition of 1.75M $(NH_4)_2SO_4$ and 0.1M Bicine pH 9.2. The resulting crystals took around three days to reach a size of approximately 0.15 x 0.15 x 0.4 mm^3 (**Figure 5.2**). These crystals were then tested in several cryo conditions, with 20% glycerol producing crystals whose diffraction on the in-house rotating anode X-ray source was beyond 2.5Å at 100K.

5.5 Initial data collection

An initial data set was collected in house on a Mar Research Mar345 image plate mounted on a Rigaku RU200 X-ray generator using crystals of the

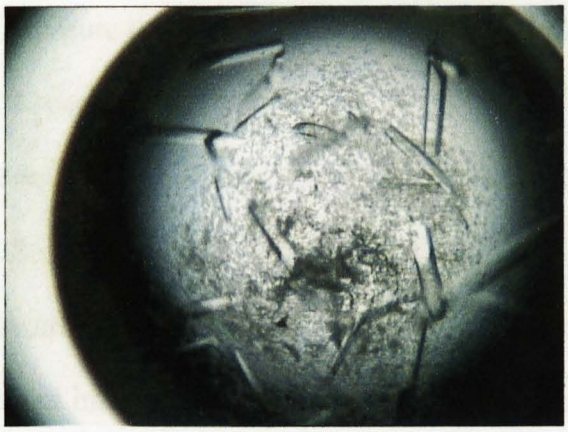


Figure 5.2. Crystals of the TgENR/NAD^+/triclosan complex grown in 1.6M $(NH_4)_2SO_4$ and 0.1M Bicine pH9.2.

(a)

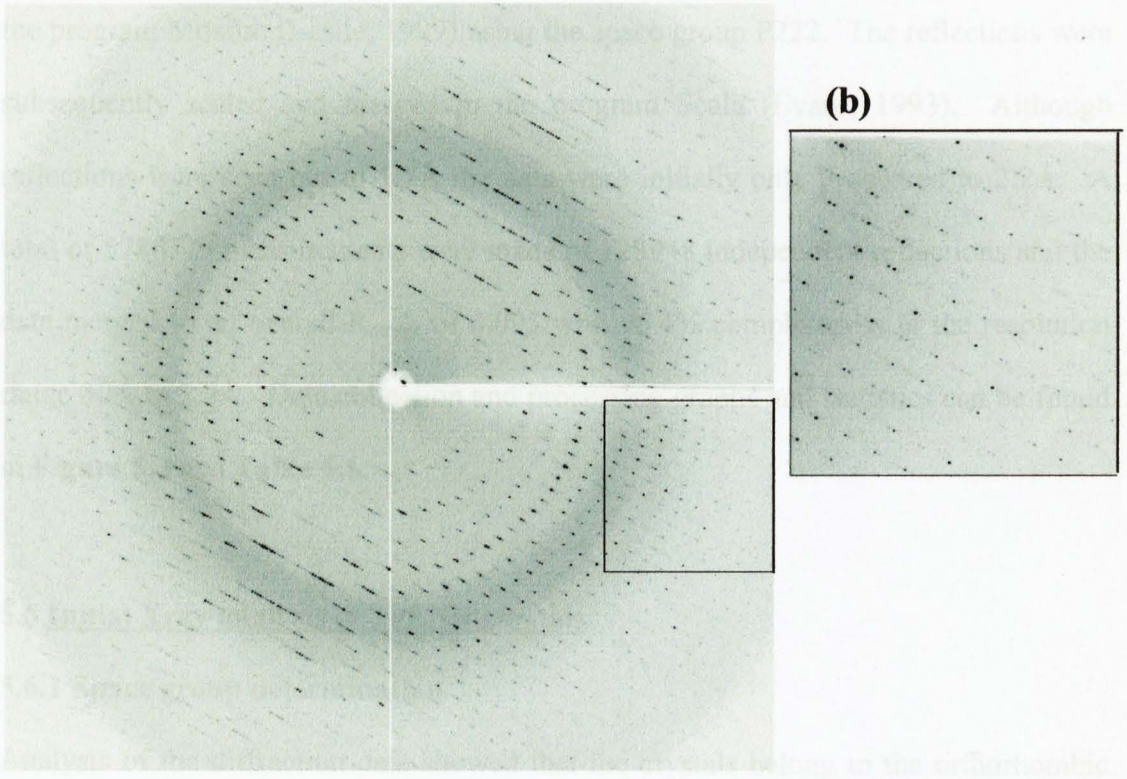


Figure 5.3 (a) A representative 0.1° oscillation frame of data collected from the TgENR complexed with NAD^+ and triclosan crystal collected on a CCDQ4 image detector on station 14.4 at the European synchrotron radiation source (ESRF). The resolution at the corners of the image plate is 1.6Å resolution. **(b)** An enlarged view of part of the image, highlighted by a black square showing the close spot separation at high resolution.

TgENR/NAD^+/triclosan complex. An oscillation of 0.5° per image was used over a total of 100° with an exposure time of 10 minutes per image. Initial analysis using the DENZO/SCALEPACK package (Otwinowski & Minor, 1997) showed the crystals to belong to the orthorhombic P222 system with approximate unit cell parameters a = 61Å, b = 153Å, c = 283Å, and $\alpha = \beta = \gamma = 90.0°$. These crystals were then taken to the European Synchrotron Radiation Facility (ESRF) station ID 14.4 for data collection. Nine hundred images were collected with a 0.1° oscillation and 1 second exposure times on an ADSC Quantum 4 detector (**Figure 5.3**). A small oscillation angle was used to overcome the poor spot separation on the images, caused by the large unit cell dimension along c of 283Å. The data were indexed and integrated in the program Mosflm (Leslie, 1999) using the space group P222. The reflections were subsequently scaled and merged in the program Scala (Evans, 1993). Although reflections were seen out to 1.7Å the data were initially only processed to 2.2Å. A total of 5786779 measurements were made of 128948 independent reflections and the data merged to an overall R_{merge} of 0.095 with 96.4% completeness in the resolution range 30Å to 2.2Å. Data collection and processing graphs and statistics can be found in **Figure 5.4** and **Table 5.1.**

5.6 Initial Xray analysis of TgENR crystals

5.6.1 Space group determination

Analysis of the diffraction data showed that the crystals belong to the orthorhombic P222 system, with unit cell parameters a = 60.6 Å, b = 152.9 Å, c = 282.9 Å, and $\alpha = \beta = \gamma = 90.0°$. Initial analysis of the X-ray diffraction pattern showed that no reflections were recorded along the 0*k*0 axis (b*) but along the *l*00 (a*) they satisfied the condition $l = 2n$ (**Figure 5.5**). Similarly reflections were recorded along the *h*00

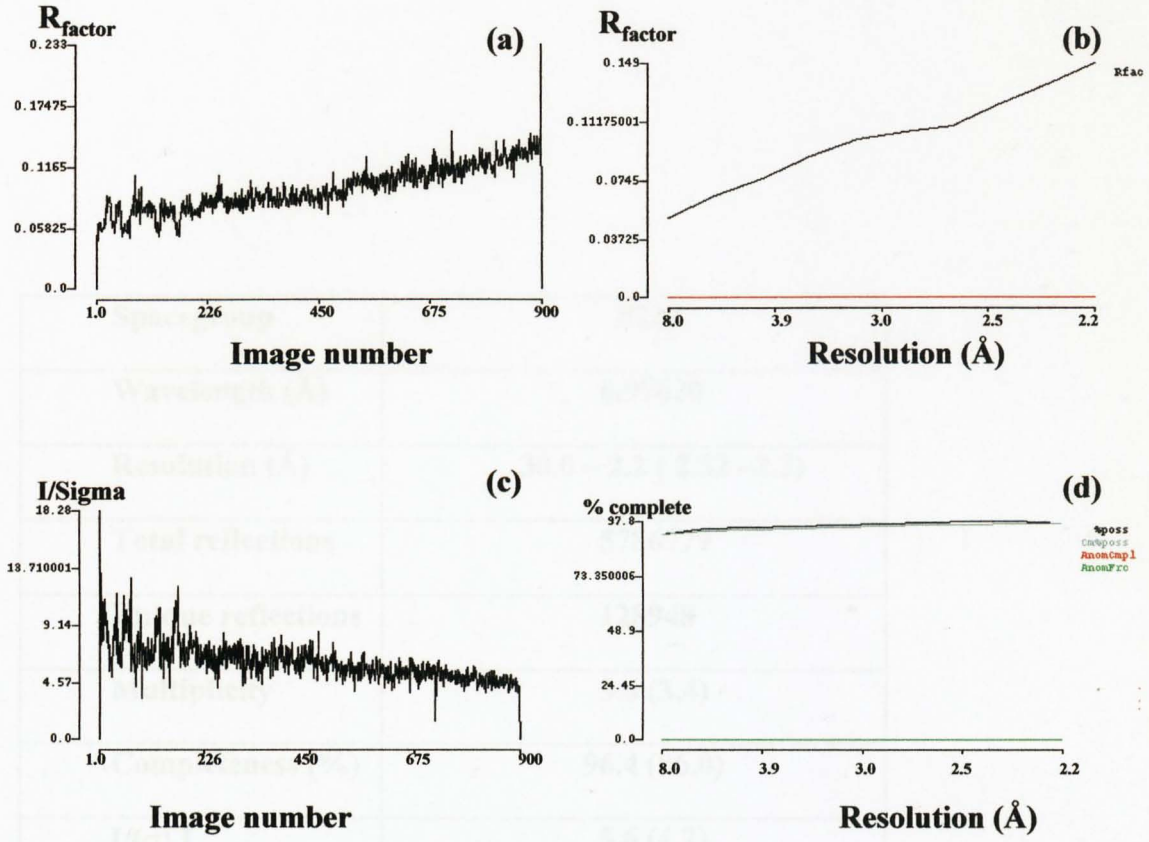


Figure 5.4 a, b, c and d. Mosflm statistics for the TgENR/NAD$^+$/triclosan complex data to 2.2Å. The graphs show **(a)** R_{factor} vs. image number (ϕ 0 to 90°) **(b)** R_{factor} vs. resolution **(c)** I/σI vs. image number (ϕ 0 to 90°) **(d)** completeness vs. resolution.

Spacegroup	**P222**
Wavelength (Å)	**0.97620**
Resolution (Å)	**30.0 – 2.2 (2.32 –2.2)**
Total reflections	**5786779**
Unique reflections	**128948**
Multiplicity	**3.6 (3.4)**
Completeness (%)	**96.4 (96.0)**
I/(σ) I	**5.6 (4.2)**
Rmerge (%)	**0.095 (0.149)**

Table 5.1 Data collection statistics for the TgENR/NAD^+/triclosan complex collected at the ESRF source to 2.2Å. Values in parentheses are for the highest resolution shell.

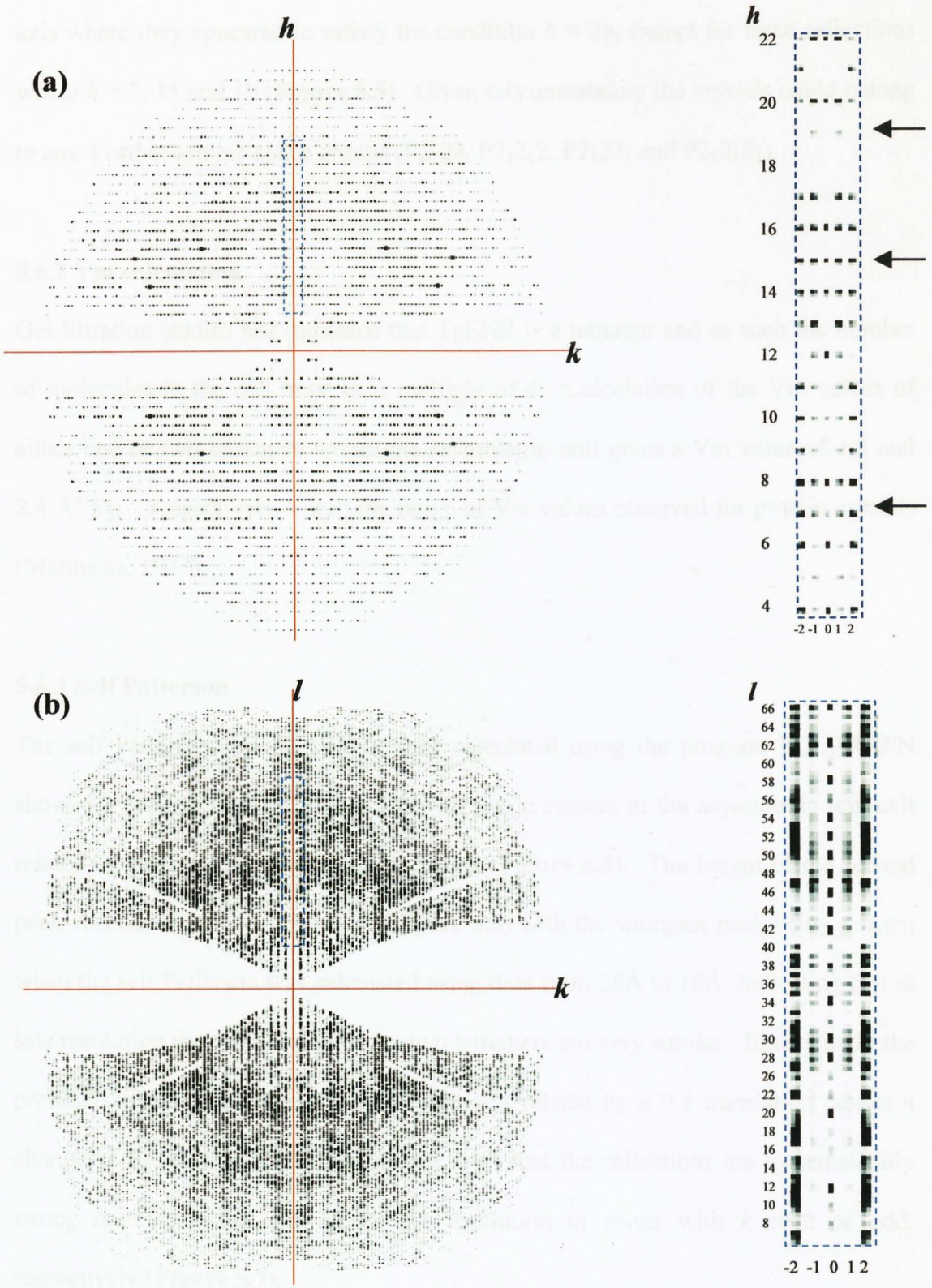


Figure 5.5 (a) The *hk*0 zone for the TgENR data to 2.2 Å, with a zoomed in view on the right hand side showing reflections along the *h* axis where *h* = 2n except for three exceptions (h=7, 15 and 19) labelled with an arrow. **(b)** The 0*kl* zone for the TgENR data to 2.2 Å, with a zoomed in view on the right hand side showing reflections along the *l* axis.

axis where they appeared to satisfy the condition h = 2n, except for three reflections where h = 7, 15 and 19 (**Figure 5.5**). Given this uncertainty the crystals could belong to any 4 orthorhombic space groups ($P2_122$, $P2_12_12$, $P2_122_1$ and $P2_12_12_1$).

5.6.2 Vm calculation

Gel filtration studies had indicated that TgENR is a tetramer and as such the number of molecules in the cell must be a multiple of 4. Calculation of the Vm values of either one or two tetramers within the asymmetric unit gives a Vm value of 4.8 and 2.4 $Å^3$ Da^{-1}, respectively within the range of Vm values observed for protein crystals (Matthews, 1977).

5.6.3 Self Patterson

The self Patterson for the TgENR data calculated using the program POLARRFN showed a peak at (0 ½ 0), consistent with two tetramers in the asymmetric unit cell related by a 0.5 translation along the b axis (**Figure 5.6**). The height of the second peak was dependent on resolution (**Figure 5.6**) with the strongest peak being present when the self Patterson was calculated using data from 20Å to 10Å, indicating that at low resolution the orientations of the two tetramers are very similar. Interestingly the presence of two tetramers within the unit cell related by a 0.5 translation causes a change in the distribution of reflections such that the reflections are systematically strong and weak for data below 5Å resolution in zones with k even or odd, respectively (**Figure 5.7**).

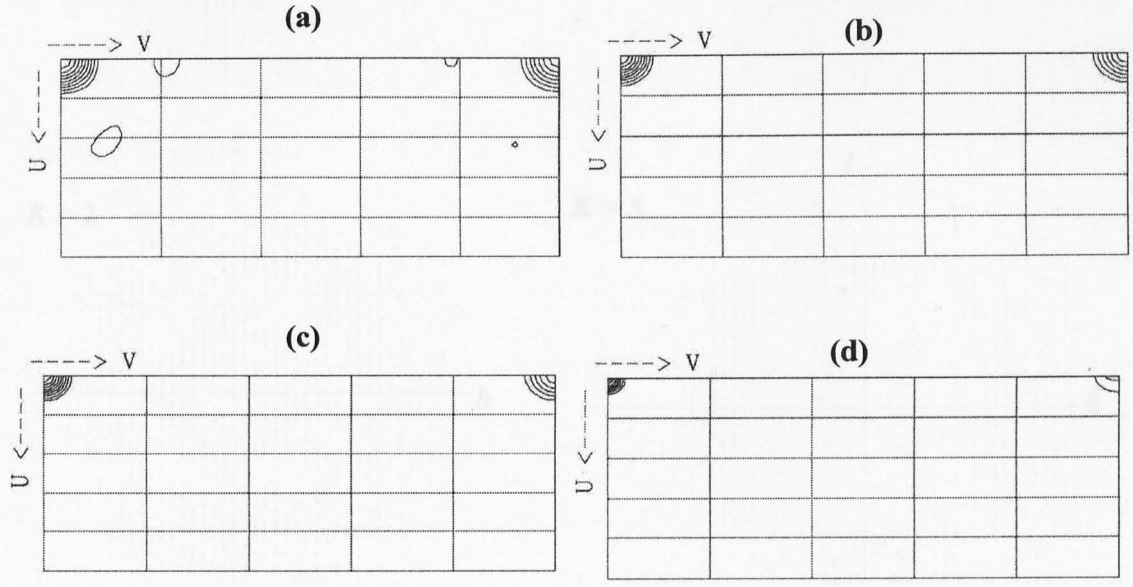


Figure 5.6. The y=0 section for self Pattersons of the TgENR crystal data calculated using data in the resolution ranges (a) 20-10Å, (b) 20-8.0Å, (c) 20-6.0Å and (d) 20-4.0Å in spacegroup P222 using POLARRFN (CCP4, 1994). All the sections cover the range $x = 0 - {}^{1}/_{2}$ and $z = 0 - {}^{1}/_{2}$. The map is contoured with the maximum at the origin set to 100 and the contours are drawn at intervals of 10%. In all maps a clear peak can be seen at 0, 0.5, 0.0. This peak decreases from 77% of the origin for the data to 10Å resolution to 70%, 58%, and 20% using data to 8Å, 6Å and 4Å, respectively.

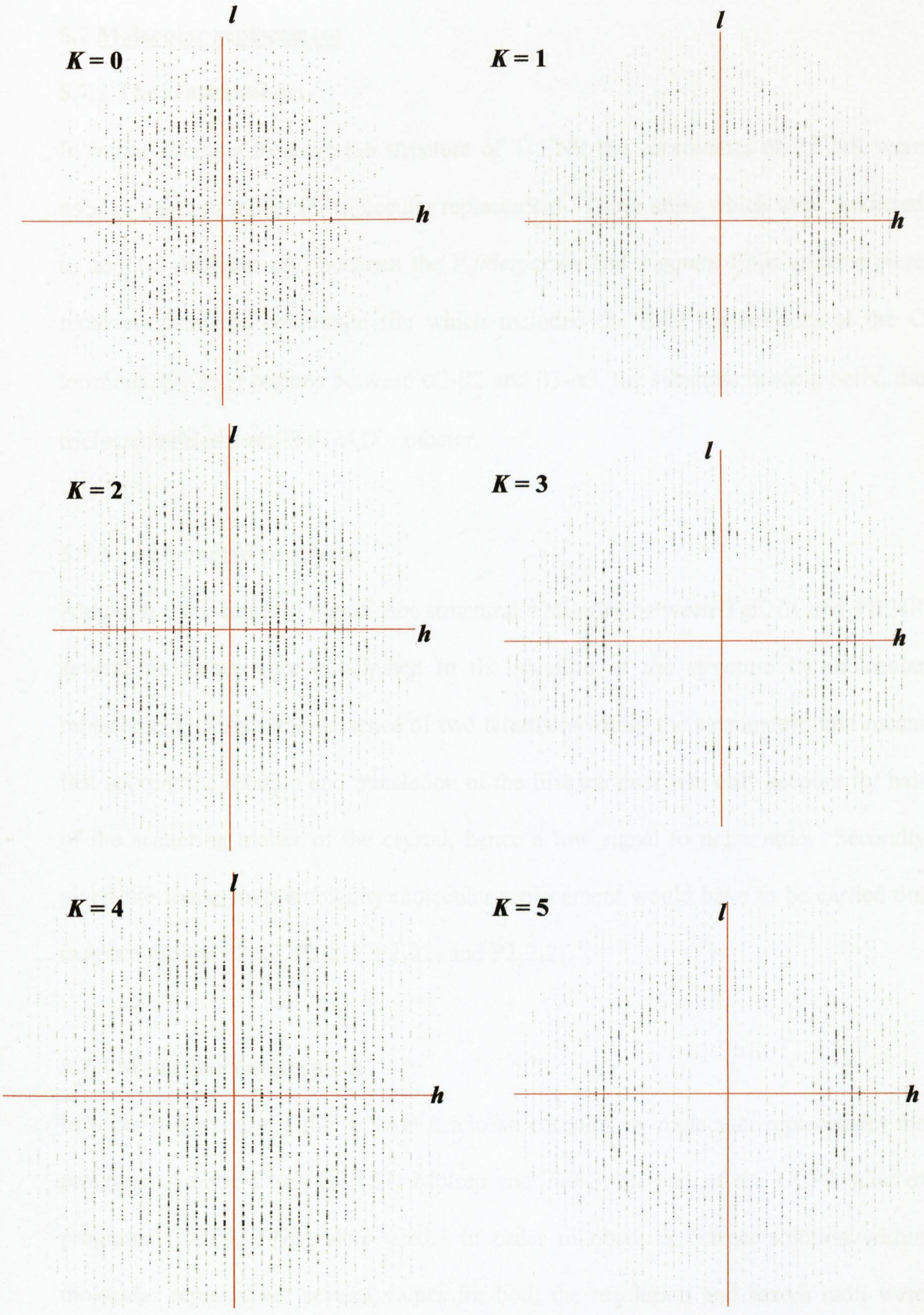


Figure 5.7. The *h*0*l*, *h*1*l*, *h*2*l*, *h*3*l*, *h*4*l* and *h*5*l* zones for the TgENR data to 2.2Å. The weak area in the centre of the *h*1*l*, *h*3*l* and *h*5*l* zones extends to approximately 5Å resolution.

5.7 Molecular replacement

5.7.1 The search model

In initial attempts to solve the structure of TgENR the coordinates of PfENR were used as a search model for molecular replacement. Those areas which were predicted to adopt a different fold between the *P.falciparum* and *T.gondii* ENR enzyme were removed from the coordinate file which included the final ten residues at the C terminus, the loop regions between α2-β3 and β3-α3, the substrate binding helix, the triclosan inhibitor and the NAD^+ cofactor.

5.7.2. Anticipated problems

Although there must be significant structural similarity between TgENR and PfENR several problems were anticipated in the solution of the structure by molecular replacement. Firstly the presence of two tetramers within the asymmetric unit meant that solving the rotation and translation of the first tetramer will only account for half of the scattering matter of the crystal, hence a low signal to noise ratio. Secondly given the spacegroup ambiguity molecular replacement would have to be carried out in space groups $P2_122$, $P2_12_12$, $P2_122_1$ and $P2_12_12_1$.

5.7.3 Molecular replacement

In order to solve the TgENR/NAD^+/triclosan complex by molecular replacement the programs Amore, CNS, BEAST, Molrep and TFFC, all part of the CCP4 suite of programs (CCP4, 1996) were used. In order to obtain a correct solution within molecular replacement several values for both the resolution and search radii were used in conjunction with changing the overall model B-factors to 20 and using a

polyalanine search model. For each program a clear rotation function was obtained, however a clear solution could not be found for the translation function.

5.8 Intrinsic problems with the data

5.8.1 Twinning

The formation of twinned crystals is a feature which occurs naturally in crystals when different domains are joined together according to a specific symmetry operation (Parsons, 2003). This results in a crystal made up of two different domains whose diffraction patterns are reflected, inverted or rotated with respect to each other, depending on the symmetry operation: the twin law (**Figure 5.8**). The ratio of the two different domains within the crystal (twinning fraction), determines the contribution of each of the domains diffraction to the overall crystal diffraction. Twinning occurs broadly into two categories, in which either all or part of certain zones are affected by overlap.

Within the orthorhombic space group P222 and its derivatives only one twinning operation can be present in which a monoclinic space group $P2_1$ whose unique angle, β, is equal to or very close to 90°, results in an apparent higher P222 spacegroup. This results in a crystal whose lattice has point group mmm but a crystal structure point group of 2/m. Furthermore, the point group mmm contains symmetry elements not in point group m and as such mistakes may occur in which unit cells may be related by symmetry operations about the a or c axes during crystal growth. This ambiguity not only produces problems associated with processing in the wrong space group, but also problems are encountered with measuring the correct intensities of reflections, when some are the sum of two separate reflections.

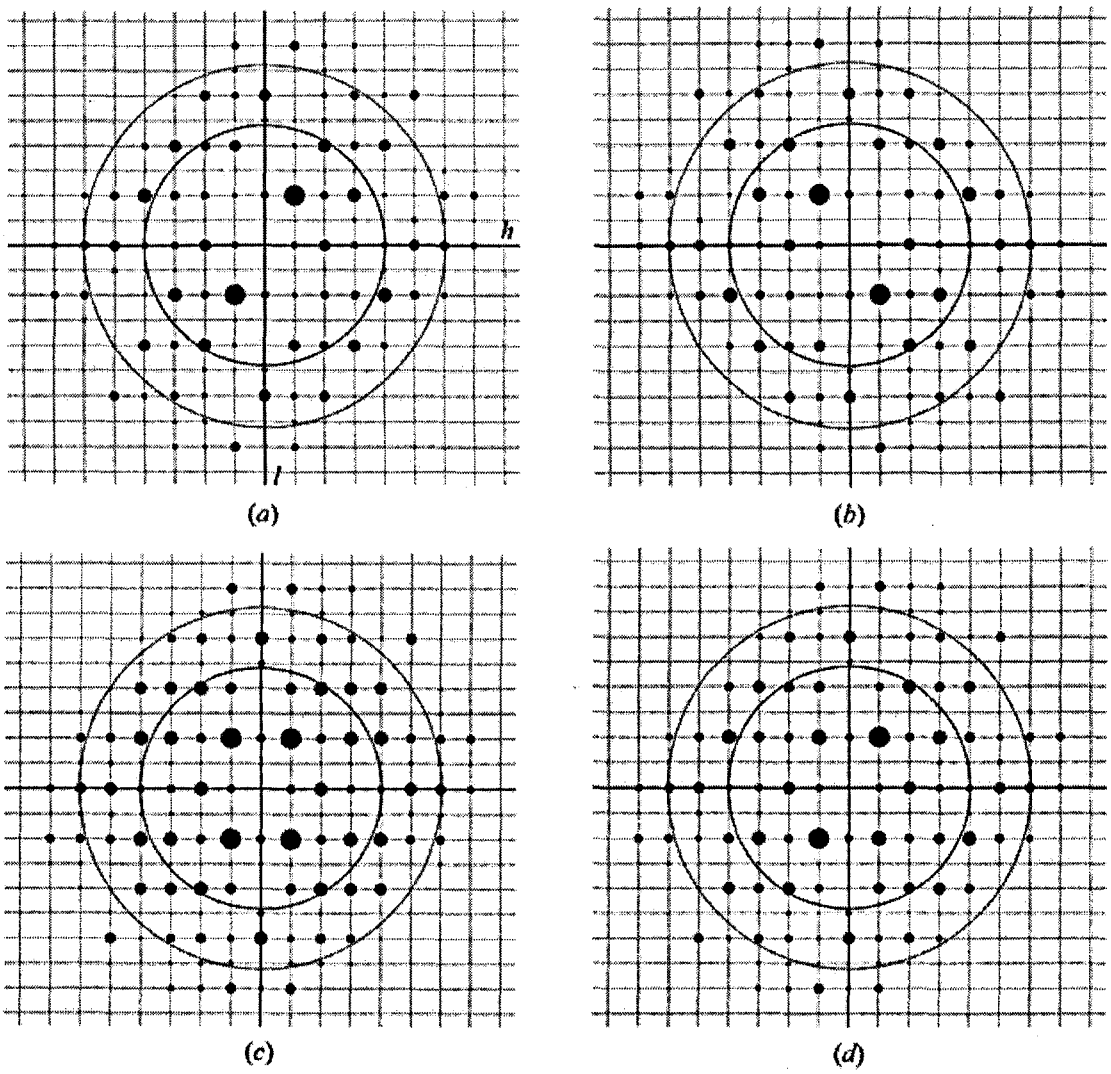


Figure 5.8 a,b, c and d. The effect of twinning on the diffraction pattern of a monoclinic crystal. **(a)** represents the diffraction pattern from one domain of a crystal. **(b)** represents the diffraction pattern from the second domain of the twinned crystal where it has been rotated by 180° about the *h* axis. Both **(c)** and **(d)** represent a twin caused by a crystal containing both domains **(a)** and **(b)**. **(c)** Representation of a 0.5 twin fraction in which the crystal is made up of an equal amount of domain **(a)** and **(b)**. **(d)** A crystals whose twin fraction is 0.2 such that the crystal contains 80% of domain **(a)** and 20% of domain **(b)**. The figure is taken from Parsons, 2003.

5.8.2 Calculating the twinning fraction of the TgENR crystal

There are several methods which may be employed to calculate the correct twinning fraction, if any, within a crystal. The presence of twinning will effect the distribution of the weak and strong reflections within the dataset and as such the calculation of the Wilson ratio $\langle F \rangle^2 / \langle I \rangle$, the "moments of intensities distribution". In a "normal" untwined dataset a value of 0.785 would be expected for this ratio, whereas with a twinned dataset a higher value of around 0.885 would be expected. Calculation of the $\langle F \rangle^2 / \langle I \rangle$ ratio for the TgENR data showed the value to be approximately 0.87 (**Figure 5.9**). However he presence of two tetramers related by a 0.5 fractional shift along the b axis within the unit cell would also affect the moments of intensities distribution and as such this ratio could not be used as a reliable indicator for twinning in the case of TgENR. As an alternative a Britton plot calculated by applying a range of twinning fractions from 0.0 to 0.5 and calculating the number of negative intensities produced (Britton, 1972) can be used. The basis of detwinning data requires the assessment of the twinning fraction, which determines the contribution of the reflections from one domain onto another and determination of the twinning operator which states which reflections will be affected by the twinning within the data set. By applying too great a twinning fraction the contribution of a reflection to its twinned counterpart is overestimated and the reflection becomes negative. Applying a twinning fraction to data which is not twinned results in the production of negative intensities which increases linearly with the twinning fraction applied. Calculation of too low a twinning fraction on twinned data will result in the production of no negative intensities. This can be seen by calculating a Britton plot for the PfENR data which is in $P2_1$ and is not twinned, within the program Detwin (CCP4, 1994). A straight line can be seen which passes through 0 indicating no twinning is present (**Figure 5.10a**).

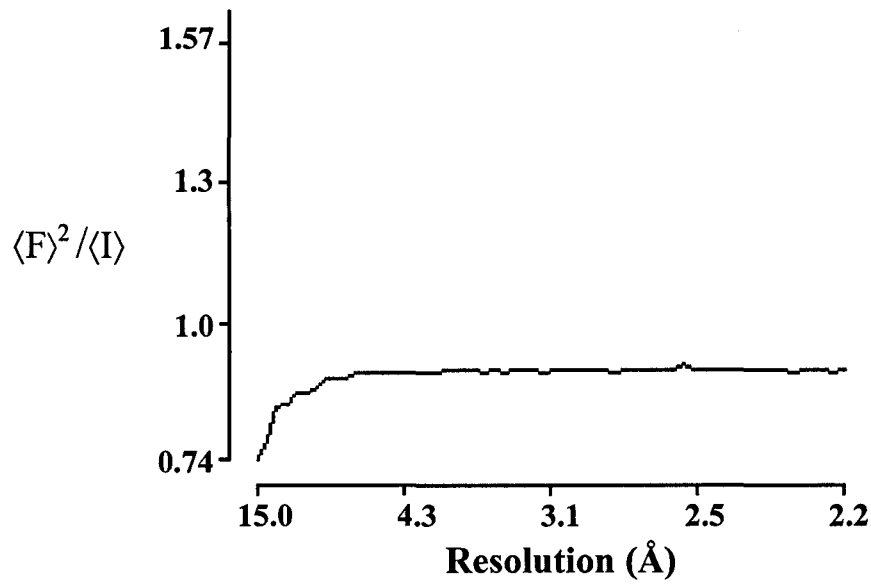


Figure 5.9. The acentric moments for the TgENR/NAD^+/triclosan data to 2.2 Å. The approximate value for the ratio of $\langle F \rangle^2 / \langle I \rangle$ is 0.87. The figure was produced in the program truncate.

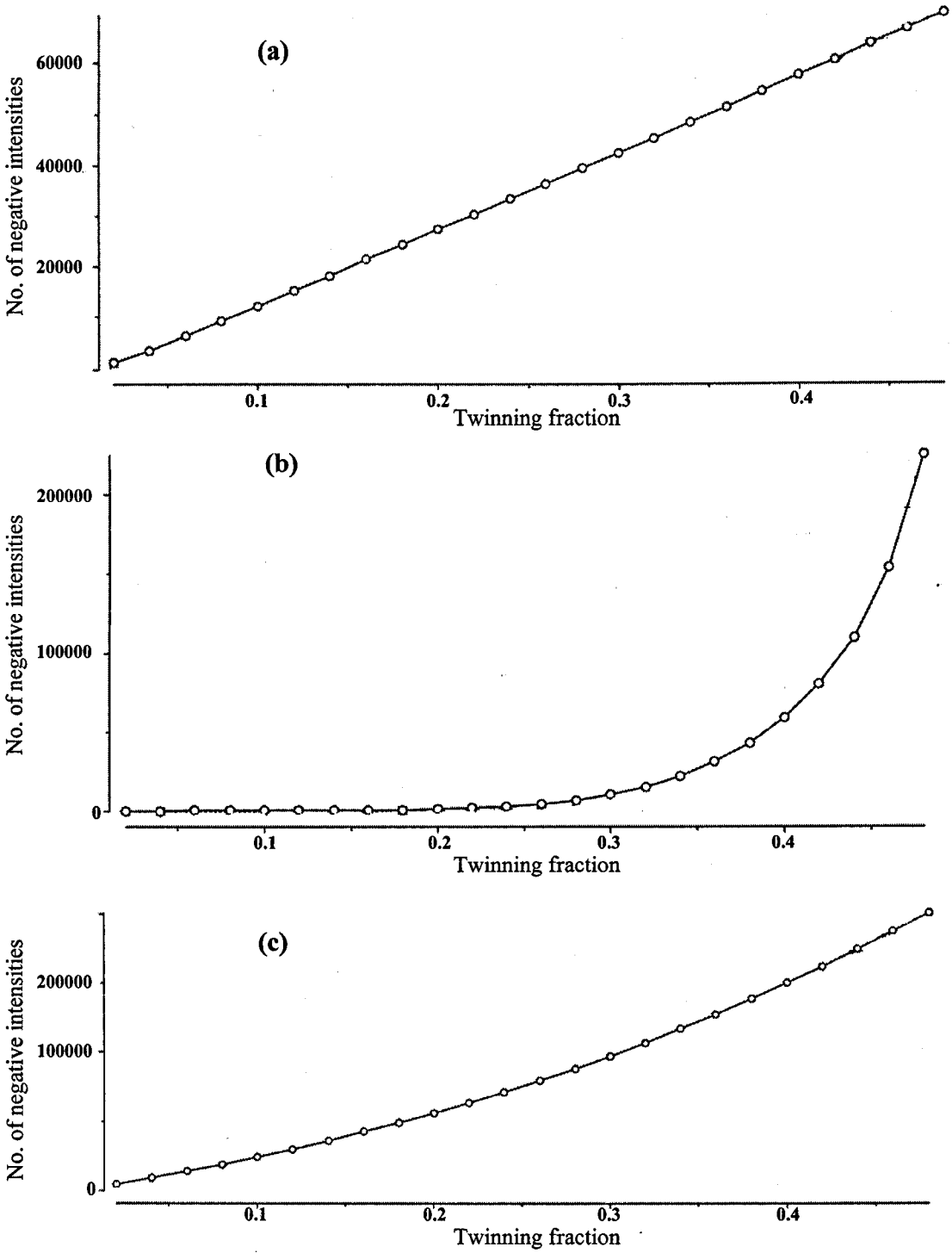


Figure 5.10. Britton plots created using the program Detwin (CCP4, 1994), showing the number of negative intensities as a function of the assumed twin fraction on data from **(a)** the PfENR/NAD^+/triclosan complex, **(b)** the TgENR/NAD^+/triclosan complex and **(c)** the TgENR/NAD^+/triclosan data after detwinning with a twinning fraction of 0.45. The Britton plot shows the number of negative intensity reflections and the applied twin fraction across the y and x axes, respectively.

Carrying out the same procedure on the TgENR data processed in P2 produces a very different graph in which a linear relationship between the number of negative intensities and the twinning fraction is not observed (**Figure 5.10b**).

5.8.3 Detwining the TgENR data

The program Detwin (CCP4, 1994) was used to detwin the TgENR data which had been processed in spacegroup P2 using a twinning fraction of 0.45 and the twinning operator h, -k, -l. After detwining the data a Britton plot was recalculated which showed a plot with a near linear line passing through 0 (**Figure 5.10c**). However the data does not look fully detwinned indicating that a twinning fraction of 0.5 is present which is indicative of a perfect merohedral twin which in the absence of strong non crystallographic symmetry is very difficult to solve. Therefore a search for a new crystal form was initiated.

CHAPTER 6

Structure determination and refinement of the *Toxoplasma gondii* ENR/NAD^+/triclosan complex

6.1 Introduction

This chapter describes the structure determination of *Toxoplasma gondii* enoyl ACP reductase (TgENR) by molecular replacement and its subsequent model building and refinement.

6.2 TgENR crystallisation

6.2.1 Preliminary crystallisation

The inability to solve the structure of TgENR from the twinned crystals described in chapter 5 led to a new set of crystallisation trials being conducted, using the Hampton research crystal Screens I, II, PegIon screen, Xtal screen lite and grid screen on TgENR complexed with NAD^+, NAD^+ and triclosan and for the free enzyme. In comparison to previous crystallisation attempts where a protein concentration of 20mg/ml was used, the TgENR concentration was lowered to between 5 and 10mg/ml. The crystals were grown using the hanging-drop vapour diffusion technique by mixing 1.5µl of the protein solution (20mM Tris 7.5 and 100mM NaCl) with 1.5µl of the reservoir solution at 290K. In order to form the complexes the protein solution was brought to a concentration of 5 and 10mg/ml by diluting the protein solution in additional buffer (20mM Tris 7.5 and 100mM NaCl) after which 5µM NAD^+ and 6µM triclosan were added to form the respective complexes. Of those conditions which produced crystals, the best quality crystals were of the TgENR/NAD^+/triclosan complex and these grew in crystal screen 1 condition 36 (0.1M Tris-HCl pH8.5 and 8% PEG 8,000), reaching a size of approximately 0.2 x 0.2 x 0.3 mm^3. Several images were collected on a tube mounted crystal to 4.5Å on a Mar Research Mar345 image plate mounted on a Rigaku RU200 X-ray generator and processed in the DENZO suite of programs (Otwinowski & Minor 1997). Analysis of

these data showed the crystals to belong to the trigonal space group P3 with approximate unit cell dimensions a = 78Å, b = 78Å, c = 188Å, $\alpha = \beta = 90°$ and $\gamma = 120°$. This crystal form had been noted in the initial crystal screening experiment in which 20mg/ml of TgENR was used as detailed in chapter 5, however the diffraction properties of the original crystals were poor with a diffraction limit of approximately 6.0Å.

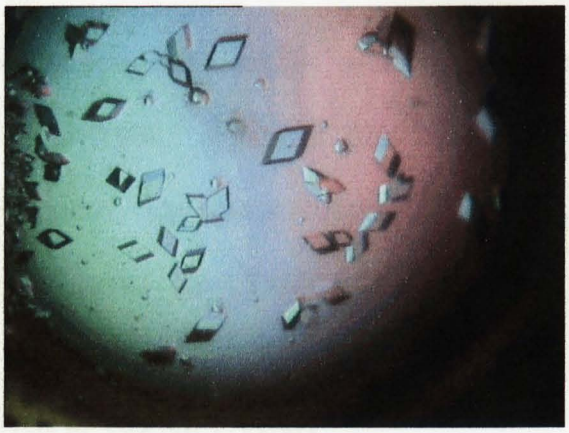
6.2.2 Optimisation of crystallisation conditions

Optimisation of the crystallisation conditions for the TgENR/NAD^+/triclosan complex was carried out by varying the concentration of TgENR from 5 to 10mg/ml, the percentage of PEG 8,000 from 0% to 20% and by changing the pH of the Tris-HCl buffer in the range of 7.5 to 9.5. The concentration of the NAD^+ and triclosan were not changed from 5μM and 6μM, respectively. This resulted in an optimal condition of 0.1M Tris-HCl pH9.0 with 6% PEG 8,000 and 10mg/ml of protein (**Figure 6.1a**).

6.3 Data Collection

Crystals of TgENR complexed with NAD^+ and triclosan were tested in a variety of cryo protectants, the most effective of which was to flash freeze them in 30% glycerol at 100K. A dataset was collected on a Mar Research Mar345 image plate mounted on a Rigaku RU200 X-ray generator to a resolution of 3.2Å. These crystals were then sent to the BESSY synchrotron radiation source for data collection. A dataset to 2.8Å was collected on a Mar165CCD image plate using a rotation oscillation of 1° over 165° and an exposure time of 15 seconds (**Figure 6.1b**). Analysis of the data was carried out using the HKL suite of programs (Otwinowski and Minor, 1997). Analysis of each image showed that those images representing 130° and 131° had a

(a)



(b)

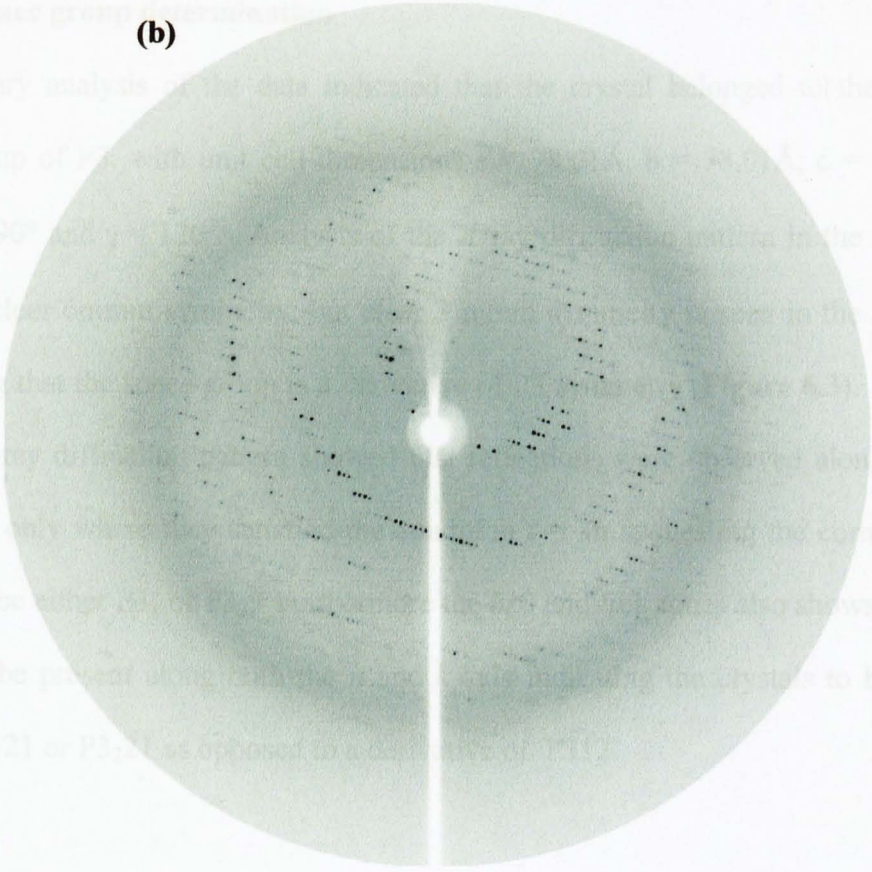


Figure 6.1 (a) Crystals of the TgENR/NAD^+/triclosan complex grown in 0.1M Tris-HCL pH 9.0 and 6% PEG 8,000 with approximate dimensions of 0.2 x 0.2 x 0.3 mm^3. **(b)** A 1° oscillation frame of a TgENR/NAD^+/triclosan crystal taken on a Mar165 CCD image plate. The resolution at the edge of the plate is 2.5Å.

significantly higher R_{sym} and Chi^2 value than the other images in the dataset. These images were subsequently removed from the data processing, reducing the overall R_{merge} value from 0.071 to 0.069. A total of 102465 measurements were made of 14012 independent reflections. The data processing graphs and statistics for the data collection can be seen in **Figure 6.2** and **Table 6.1**.

6.4 Initial X-ray analysis of TgENR crystals

6.4.1 Space group determination

Preliminary analysis of the data indicated that the crystal belonged to the trigonal spacegroup of P3, with unit cell dimensions a = 78.01Å, b = 78.01Å, c = 188.45Å, $\alpha = \beta = 90°$ and $\gamma = 120°$. Analysis of the X-ray diffraction pattern in the *hk*0 zone showed clear 6mmm symmetry, but clear 3 mmm symmetry is seen in the *hk*1 zone indicating that the space group is a derivative of P3 symmetry (**Figure 6.3**). Analysis of the X-ray diffraction pattern showed that reflections were observed along the 00*l* axis (c*) only where they satisfied the condition $l = 3n$ suggesting the correct space group to be either $P3_1$ or $P3_2$. Furthermore the *hk*0 and *hk*1 zones also shows a mirror plane to be present along both the *h* and *k* axis indicating the crystals to belong to either $P3_121$ or $P3_221$ as opposed to a derivative of P312.

6.4.2 Vm Calculation

Analysis of the Vm for the TgENR/NAD^+/triclosan crystals suggests that the asymmetric unit contains a dimer, with a Vm of 2.4$Å^3$ Da^{-1}, which is within the range of Vm values observed for protein crystals (Matthews, 1977).

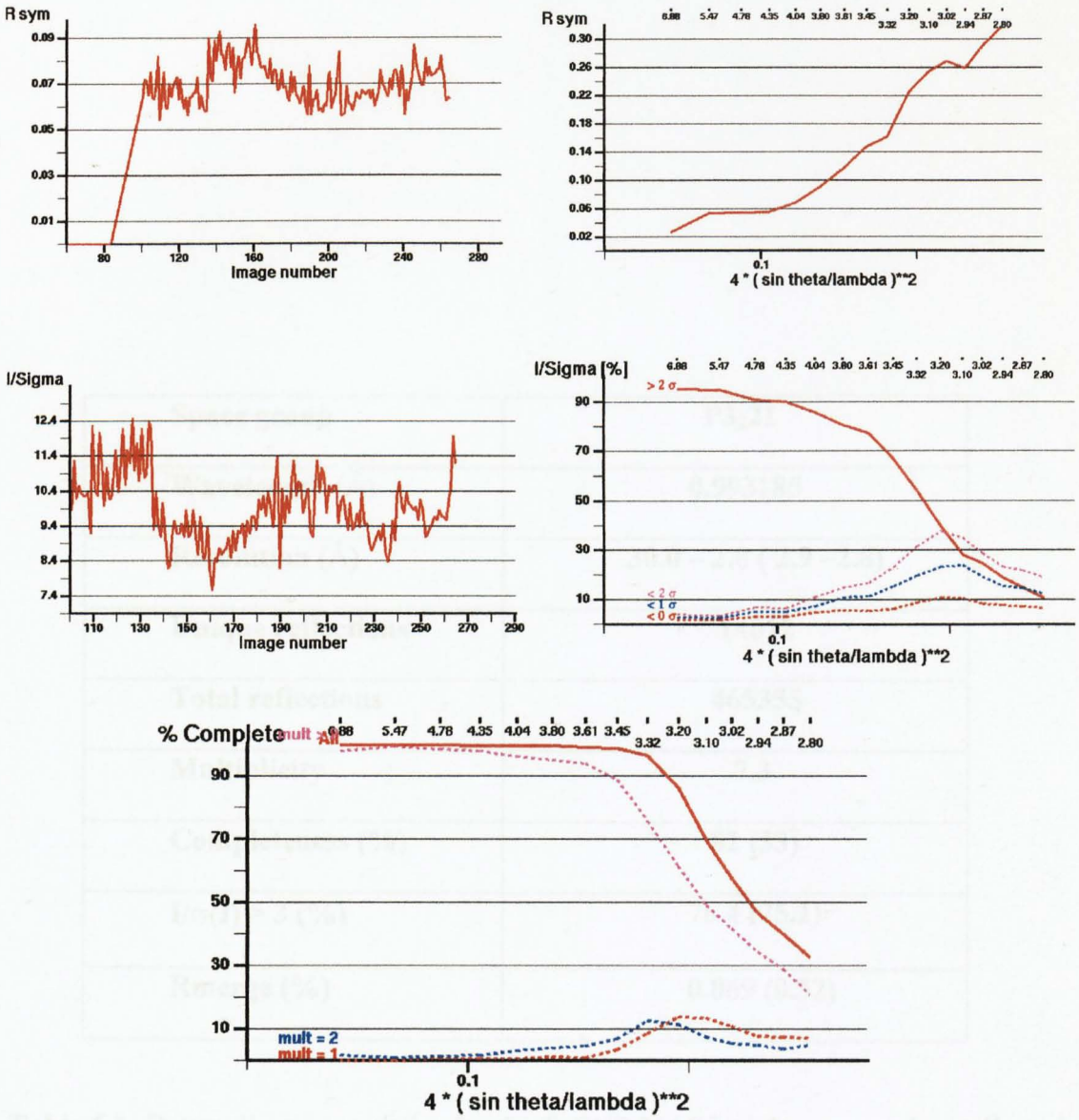


Figure 6.2. Scalepack statistics for the TgENR/NAD$^+$/triclosan complex data to 2.8Å. The graphs show (a) R_{sym} vs. image number (ϕ 0 to 165°) (b) R_{sym} vs. resolution (c) I/σI vs. image number (ϕ 0 to 165°) (d) I/σI vs. resolution (e) completeness vs. resolution. The images were numbered from 101 to 265 and as such no values are given within the R_{sym} vs. image number for the first 100 images.

Space group	**$P3_221$**
Wavelength (Å)	**0.993185**
Resolution (Å)	**30.0 – 2.8 (2.9 –2.8)**
Unique reflections	**14012**
Total reflections	**465355**
Multiplicity	**7.3**
Completeness (%)	**82 (33)**
I/σ(*I*) > 3 (%)	**76.4 (75.1)**
Rmerge (%)	**0.069 (0.32)**

Table 6.1 Data collection statistics for the TgENR/NAD^+/triclosan complex collected at the Bessy synchrotron radiation source to 2.8Å. Values in parentheses are for the highest resolution shell.

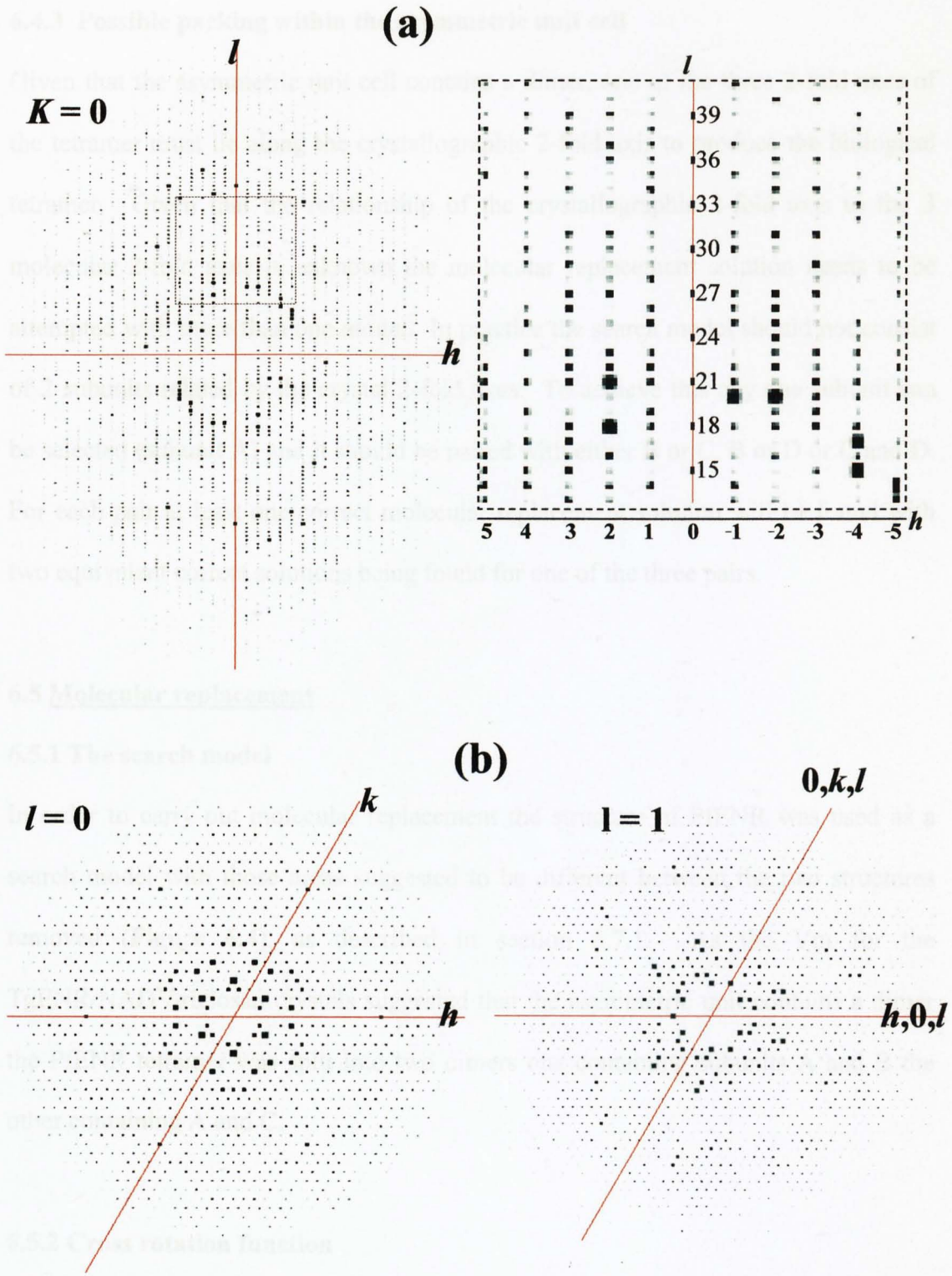


Figure 6.3 a, b and c. (a) The *h*0*l* zone for the TgENR data to 2.8Å, with a zoomed in view on the right hand side showing reflections along the *l* axis where *l* = 3n when *h* and *k* = 0. **(b)** The *hk*0 and *hk*1 zone for the TgENR data to 2.8Å which shows mirror symmetry along both the *h* and *k* axis indicating the spacegroup to be a derivative of P321 symmetry.

6.4.3 Possible packing within the asymmetric unit cell

Given that the asymmetric unit cell contains a dimer, one of the three 2-fold axes of the tetramer must lie along the crystallographic 2-fold axis to produce the biological tetramer. Given that the relationship of the crystallographic 2-fold axis to the 3 molecular 2-fold axes is unknown the molecular replacement solution needs to be attempted with more than one model. In practice the search model should not consist of 2 subunits related by the crystal 2-fold axes. To achieve this any one subunit can be selected (subunit A) and it should be paired with either B or C, B or D or C and D. For each pair at least one correct molecular replacement solution will be found with two equivalent correct solutions being found for one of the three pairs.

6.5 Molecular replacement

6.5.1 The search model

In order to carry out molecular replacement the structure of PfENR was used as a search model with those areas suggested to be different between the two structures removed (**Figure 6.4**), as described in section 5.7.1. As the Vm for the TgENR/NAD^+/triclosan crystals suggested that the asymmetric unit contains a dimer the PfENR tetramer was split into two dimers one containing subunits A and B the other containing A and C.

6.5.2 Cross rotation function

The program Amore (Navaza, 1994) was used to carry out both the rotation and translation functions. Both the PfENR AB and AC dimers were used as search models to calculate the cross rotation function, using a radius of integration of 30.0Å on data from 8.0 to 3.0Å resolution. Of the two dimer search models it was the AB

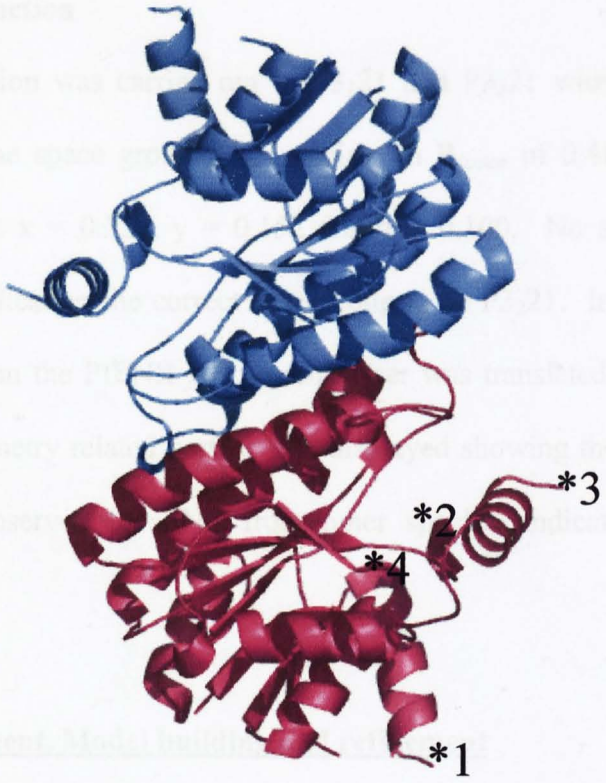


Figure 6.4 The PfENR search dimer formed by subunits A and B coloured blue and red, respectively. The areas where residues were removed are shown as stars with number 1 being 63 to 75, 2 being 93 to 108, 3 being 238 to 241 and 4 being 336 to 346. The figure was produced using the graphics program Pymol (Delano, 2002).

dimer which produced the clearer solutions of which the highest scoring was at α = 11.7, β = 85.9 and γ = 127.4, achieving an R_{factor} of 0.55 and correlation coefficient of 0.46.

6.5.3 Translation function

The translation function was carried out in $P3_121$ and $P3_221$ with a clear solution being obtained for the space group $P3_221$ giving an R_{factor} of 0.48 and correlation coefficient of 0.56 at x = 0.328, y = 0.152 and z = 0.100. No solution could be obtained in $P3_121$ indicating the correct spacegroup to be $P3_221$. In order to test the validity of the solution the PfENR AB search dimer was translated to the suspected solution and its symmetry related partner was displayed showing the production of a tetramer like that observed for ENR from other species, indicative of a correct solution (**Figure 6.5**).

6.6 Phase improvement, Model building and refinement

6.6.1 Rigid body refinement

The solution output by Amore was subjected to 20 rounds of rigid body refinement within the program REFMAC5 (Murshudov *et al.*, 1997). Only data in the resolution range 20.0 Å to 4.0 Å were used and each of the subunits within the dimer was treated as an individual rigid body or domain. This resulted in a small drop in R_{factor} to 0.46 (R_{free} 0.47).

6.6.2 Restrained least squares refinement

The model was then subject to x, y, z refinement over 10 cycles on data in the resolution range of 20.0Å to 2.8Å, with sigma A weighted maps being calculated

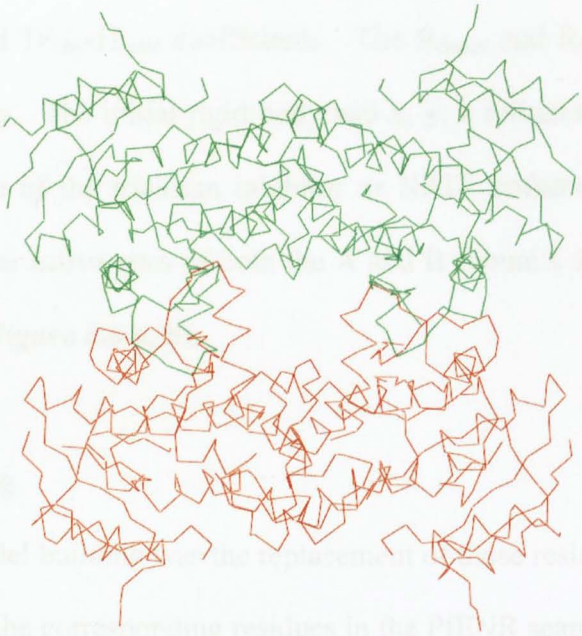


Figure 6.5. The TgENR dimer shown in green with its symmetry generated partner shown in red producing a TgENR tetramer. The figure was produced using the graphics program TURBO/FRODO (Roussel *et al.*, 1990).

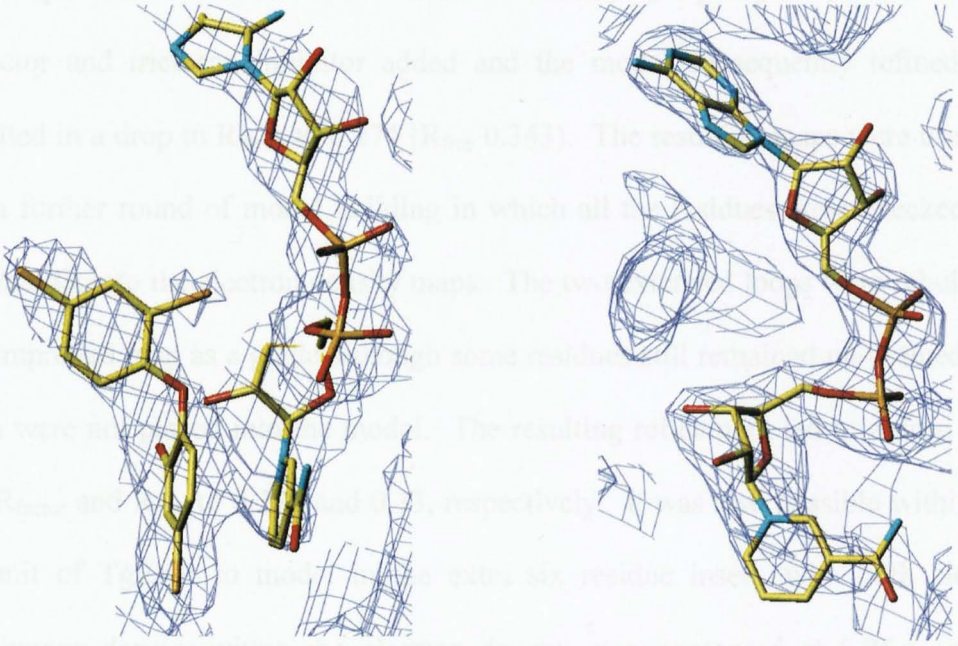


Figure 6.6. The $2F_{obs}$-F_{calc} electron density maps of the triclosan and NAD^+ cofactor within the TgENR model contoured at 0.8σ. The maps were produced through refinement of the initial TgENR model without the addition of either NAD^+ or triclosan. The NAD^+ and triclosan molecules are coloured with yellow for carbon, blue for nitrogen, red for oxygen and orange for phosphorous in NAD^+ and chlorine in triclosan. The triclosan and NAD^+ molecules have been added as a guide to the electron density. The maps were produced using the graphics program TURBO/FRODO (Roussel *et al.*, 1990).

using $2F_{obs}$-$1F_{calc}$ and $1F_{obs}$-$1F_{calc}$ coefficients. The R_{factor} and R_{free} dropped to 0.30 and 0.42, respectively. The initial rigid body and x, y, z refinement was carried out without the inclusion of the triclosan inhibitor or NAD^+ cofactor, yet clear density was present within the active sites of both the A and B subunits showing a low level of bias in the maps (**Figure 6.6 a, b**).

6.6.3 Model building

The first stage in model building was the replacement of those residues within TgENR which differed from the corresponding residues in the PfENR search model. Analysis of the maps showed a low level of bias for those residues that were incorrect. The two loops which were removed in the search model, were partly rebuilt and the NAD^+ cofactor and triclosan inhibitor added and the model subsequently refined. This resulted in a drop in R_{factor} to 0.274 (R_{free} 0.343). The resulting maps were used again for a further round of model building in which all the residues were checked for an optimal fit into the electron density maps. The two extended loops were rebuilt using the improved map as a guide although some residues still remained un-defined and as such were not placed into the model. The resulting refinement led to a drop in both the R_{factor} and R_{free} to 0.270 and 0.33, respectively. It was also possible within the A subunit of TgENR to model in the extra six residue insert, with both clear and continuous density within the electron density map contoured at 0.75σ and large positive peaks within the difference map contoured at 2.5σ. After the addition of the loop within the A subunit the model was refined resulting in clear and continuous density within the refined map with no negative peaks within the difference map. There was however a slight increase in the R_{free} of 0.01, but the quality of the refined electron density map indicated a correct solution in the placing of the chain.

6.6.4 NCS averaging and TLS refinement

Throughout the initial refinement medium to loose NCS restraints were placed upon the dimer. The poor resolution of the data resulted in an inability to refine the individual B-factors of the atoms due to the poor observation to parameter ratio. As a result an overall B factor was refined and the TLS program was used to model the motion of the molecule. By defining each of the subunits and its respective NAD^+ and triclosan cofactors as independent TLS units both the R_{factor} and R_{free} dropped to 0.25 and 0.3, respectively.

6.6.5 The final 2.8Å TgENR model

Although the model was not at this point fully refined or built, with some residues within the extended loop regions not yet fitted model building was halted until a higher resolution data set could be obtained.

6.7 Collection of a higher resolution data set

A major problem with the TgENR/NAD^+/triclosan crystal form was its fragile nature which results in the cracking of the crystals upon freezing, resulting in a split diffraction pattern. Several crystals diffracted to beyond 2.8Å but all had two or more lattices present on the image. In order to obtain crystals with an increased stability within the cryo protectant solution, the crystals were grown in 0.1M Tris-HCl pH9.0 and 6% PEG 8,000 with 5μM NAD^+ and 6μM triclosan with an additional 5%, 10%, 15%, 25% or 30% glycerol. Crystals grew under all conditions but those containing a higher percentage of glycerol diffracted poorly (approximately 4Å). However those crystals which grew in 5% glycerol were more stable when placed into a cryo solution containing the same conditions as in the hanging drop but with an additional 25%

glycerol. Flash freezing of the crystals within a stream of nitrogen gas at 100K with the new cryo conditions resulted in a higher resolution data set being collected to 2.6Å.

6.7.1 Data collection

A dataset on TgENR/NAD^+/triclosan crystals was collected to beyond 2.6Å on a Mar Research Mar345 image plate mounted on a Rigaku RU200 X-ray generator using a rotation oscillation of 1° over 90° and an exposure time of 10 minutes (**Figure 6.7**). Analysis of the data was carried out using the HKL suite of programs (Otwinowski & Minor, 1997) with a total of 110634 measurements being made of 21105 independent reflections. The data processing graphs and statistics can be seen in **Figure 6.8** and **Table 6.2**. The data were cut to 2.6Å due to the poor quality of data beyond this resolution.

6.7.2 Higher resolution refinement

Analysis of the processed data showed approximately the same unit cell dimensions of a = 78.075Å, b = 78.075Å, c = 188.45Å, $\alpha = \beta = 90°$ and $\gamma = 120°$ as those refined for the crystal used for the data collection to 2.8Å. The refined model of the TgENR was subsequently refined using REFMAC5 (Murshudov *et al.*, 1997), with the new data to 2.6Å giving an R_{factor} and R_{free} value of 0.24 and 0.31, respectively. The maps from refinement were then inspected to look for improvements in quality. The map showed areas of clear improvements allowing for the positioning of several previously un-defined residues (**Figure 6.9**).

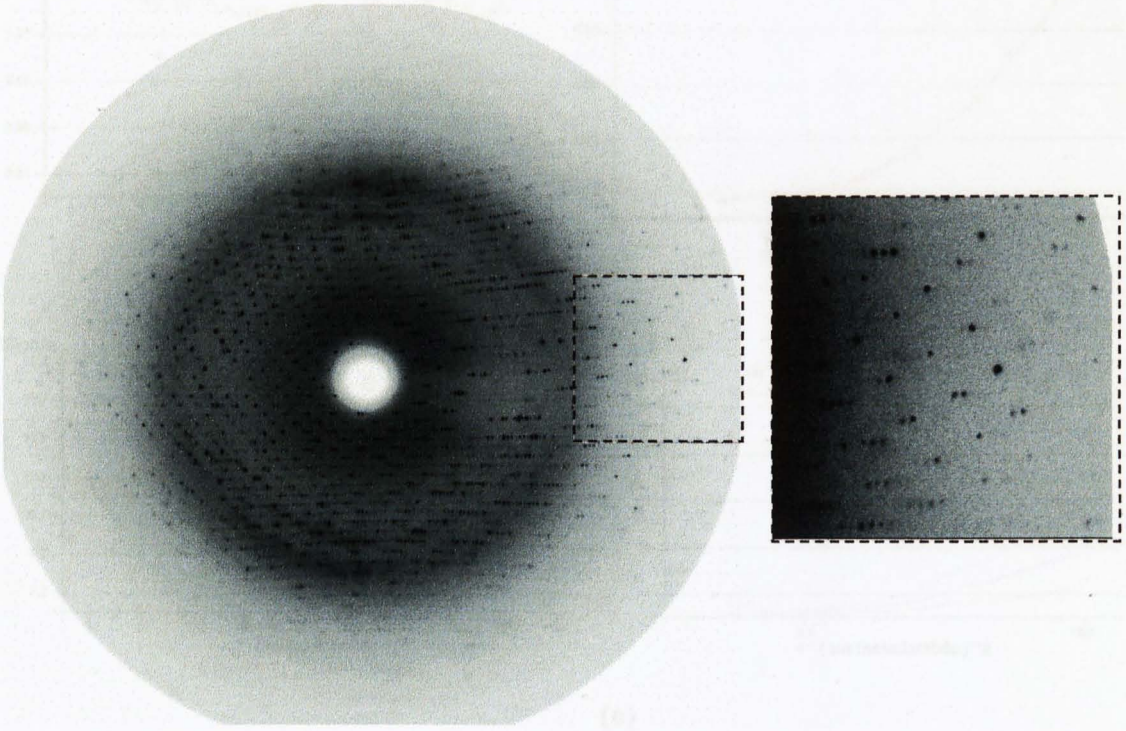


Figure 6.7. A 1° oscillation frame of a TgENR/NAD^+/triclosan complex crystal taken on a Mar Research Mar345 image plate mounted on a Rikagu rotating copper anode. The resolution at the edge of the plate is 2.3Å.

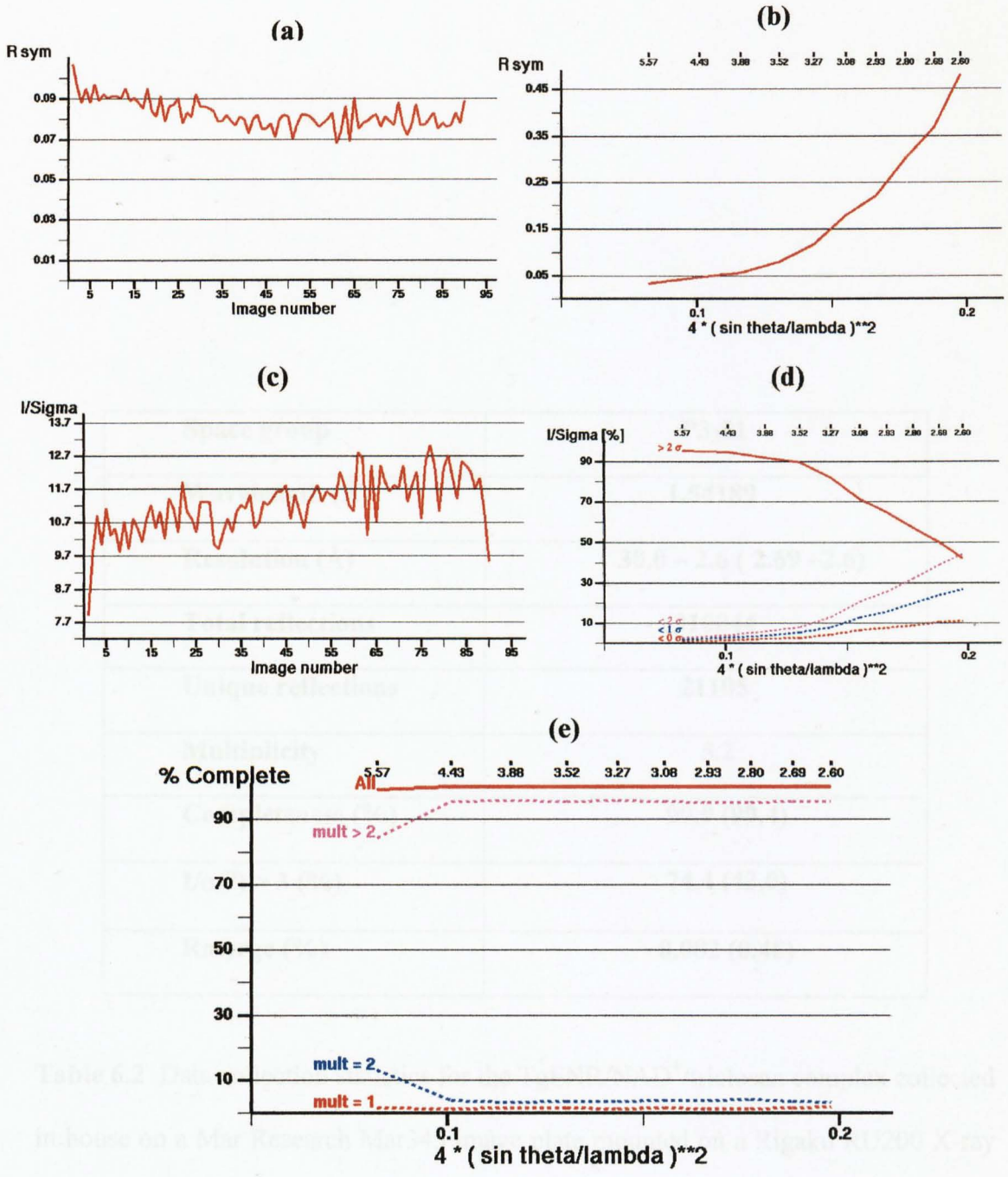


Figure 6.8. Scalepack statistics for the TgENR/NAD$^+$/triclosan complex data to 2.6Å. The graphs show (a) R_{sym} vs. image number (ϕ 0 to 95°) (b) R_{sym} vs. resolution (c) I/σI vs. image number (ϕ 0 to 95°) (d) I/σI vs. resolution (e) completeness vs. resolution.

Space group	**$P3_221$**
Wavelength (Å)	**1.54189**
Resolution (Å)	**30.0 – 2.6 (2.69 –2.6)**
Total reflections	**310044**
Unique reflections	**21105**
Multiplicity	**5.2**
Completeness (%)	**99.9 (99.4)**
$I/\sigma(I) > 3$ (%)	**74.4 (42.0)**
Rmerge (%)	**0.082 (0.48)**

Table 6.2 Data collection statistics for the TgENR/NAD^+/triclosan complex collected in house on a Mar Research Mar345 image plate mounted on a Rigaku RU200 X-ray generator to 2.6Å. Values in parentheses are for the highest resolution shell.

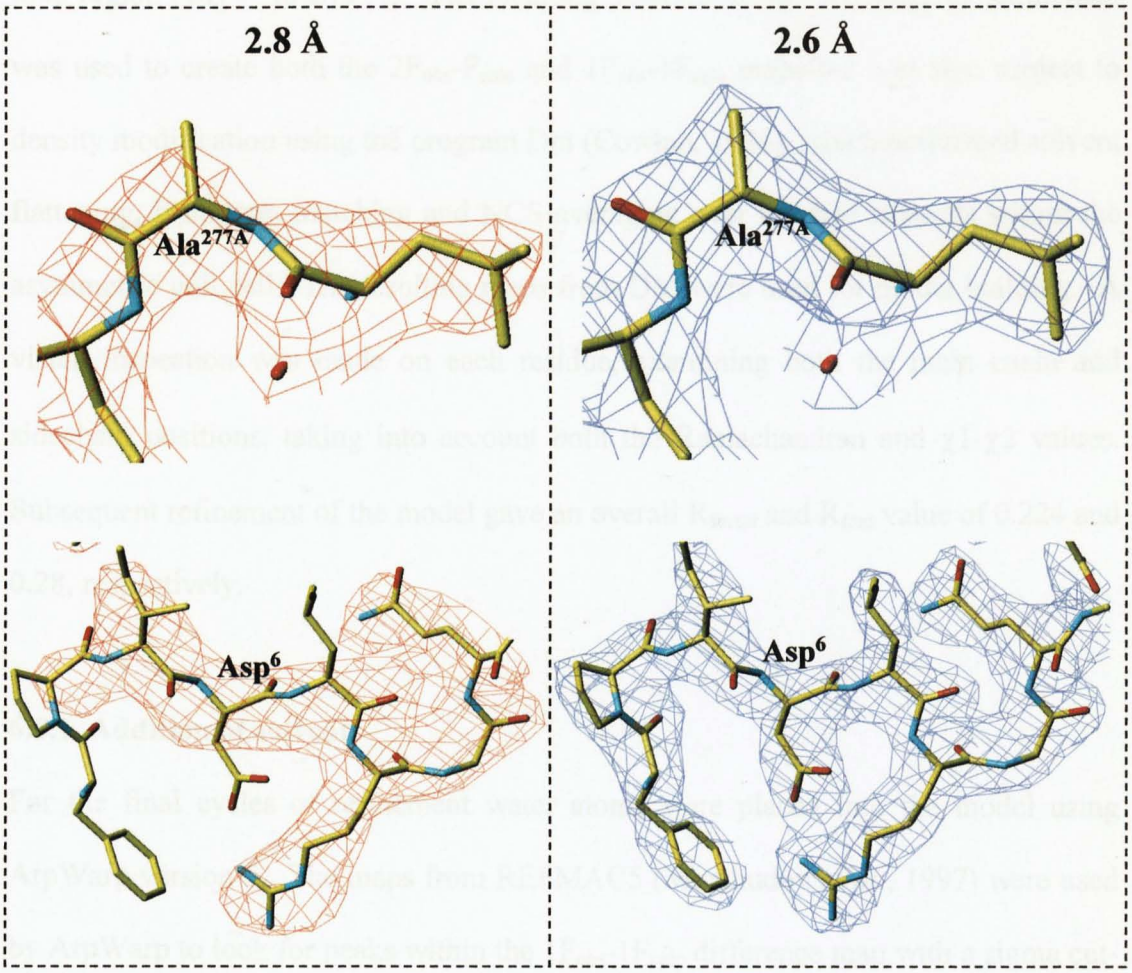


Figure 6.9 A representation of the difference in map quality for certain areas of the TgENR model. Both maps represent the $2F_{obs}$-F_{calc} maps after refinement for the 2.8Å and 2.6Å data shown as a red and blue map, respectively. Both maps are contoured at 1σ. The figure was produced using the graphics program TURBO/FRODO (Roussel *et al.*, 1990).

6.8 Model building

The increase in resolution showed areas of improvement within the TgENR model and in particular in the region of Lys^{63} to Gly^{67} and Lys^{93} to Arg^{97} and the inserted loop region (Gly^{236} to Lys^{241}). The resulting .mtz file from each stage of refinement was used to create both the $2F_{obs}$-F_{calc} and $1F_{obs}$-$1F_{calc}$ maps but was also subject to density modification using the program Dm (Cowtan, 1994), which performed solvent flattening, histogram matching and NCS averaging over the two subunits within the asymmetric unit cell. The resulting maps from DM were used for model building. A visual inspection was made on each residue, examining both the main chain and sidechain positions, taking into account both the Ramachandran and $\chi 1$-$\chi 2$ values. Subsequent refinement of the model gave an overall R_{factor} and R_{free} value of 0.224 and 0.28, respectively.

6.8.1 Addition of solvent

For the final cycles of refinement water atoms were placed into the model using ArpWarp version 5. The maps from REFMAC5 (Murshudov *et al.*, 1997) were used by ArpWarp to look for peaks within the $1F_{obs}$-$1F_{calc}$ difference map with a sigma cut-off of $> 2.5\ \sigma$ and which lay within 2.2Å to 3.3Å of a protein N or O atom. All water atoms closer than 2.2Å were merged to a single water before refinement in REFMAC5, with the maps from REFMAC5 again being used for ArpWarp. This was continued for 10 cycles. Each of the individual water molecules was then visually inspected to check it was placed within a discrete spheroid of density and within hydrogen bonding distance to neighbouring atoms. All water molecules with a B-factor over $60Å^2$ were removed. A total of 39 water atoms were added to the TgENR model resulting in a reduction of the R_{factor} and R_{free} to 0.216 and 0.279, respectively.

CHAPTER 7

Analysis of the structure of the *Toxoplasma gondii* ENR/NAD^+/triclosan complex

7.1 Introduction

The following chapter describes the final model of *Toxoplasma gondii* enoyl ACP reductase (TgENR) and the mode of binding triclosan and NAD^+.

7.2 Model stereochemistry

The final model of the TgENR/NAD^+/triclosan complex comprises 606 residues with Pro^4 to Pro^{306} in the A subunit and Phe^3 to Gly^{239} and Lys^{241} to Pro^{306} in the B subunit. Two residues in the A subunit were truncated due to weak or no density (Lys^{64} was truncated to Cβ and Lys^{94} to Cδ). The model has an R_{factor} of 0.22 (R_{free} of 0.28) for all data within the resolution range 20.0 Å to 2.6 Å (**Table 7.1**). Analysis of the TgENR model geometry was carried out using the program PROCHECK (Laskowski *et al.*, 1993) and showed a tight geometry with most values being inside or better than the range expected for a model at this resolution (**Figure 7.1 a and 7.2 a**). Examination of the χ1-χ2 plots for all residue types showed the sidechains to be in favourable conformations for the two monomers in the asymmetric unit (**Figure 7.1 b and 7.2b**). The bond angle and length r.m.s.d from ideal geometry as defined by REFMAC5 (Murshudov *et al.*, 1997) are 0.015Å and 1.7°, respectively. The Ramachandran plot for the TgENR dimer showed 87% of residues to be in the most favoured regions and 13% in additionally allowed regions, with no residues in the generously or disallowed regions (**Figure 7.3 a** and **b**). A representative part of the electron density map for the final refined TgENR model is presented in **Figure 7.4**.

7.3 Overall fold

The TgENR monomer has overall dimensions of 45 x 50 x 60Å and is formed from nine α-helices (α1-α9) comprising 108 residues (~36 %), seven β-sheets (β1-β7)

Refinement statistics and model quality	
Total non-H atoms	4678
Water sites	39
Average B-value ($Å^2$)	30
Resolution limits (Å)	20-2.6
Number of unique reflections	21105
R-value	0.22
Rfree Value	0.28
r.m.s. bond deviation (Å)	0.015
r.m.s. angle deviation (deg.)	1.7

Table 7.1. A table to show the refinement statistics for the TgENR/NAD^+/triclosan complex to 2.6Å resolution.

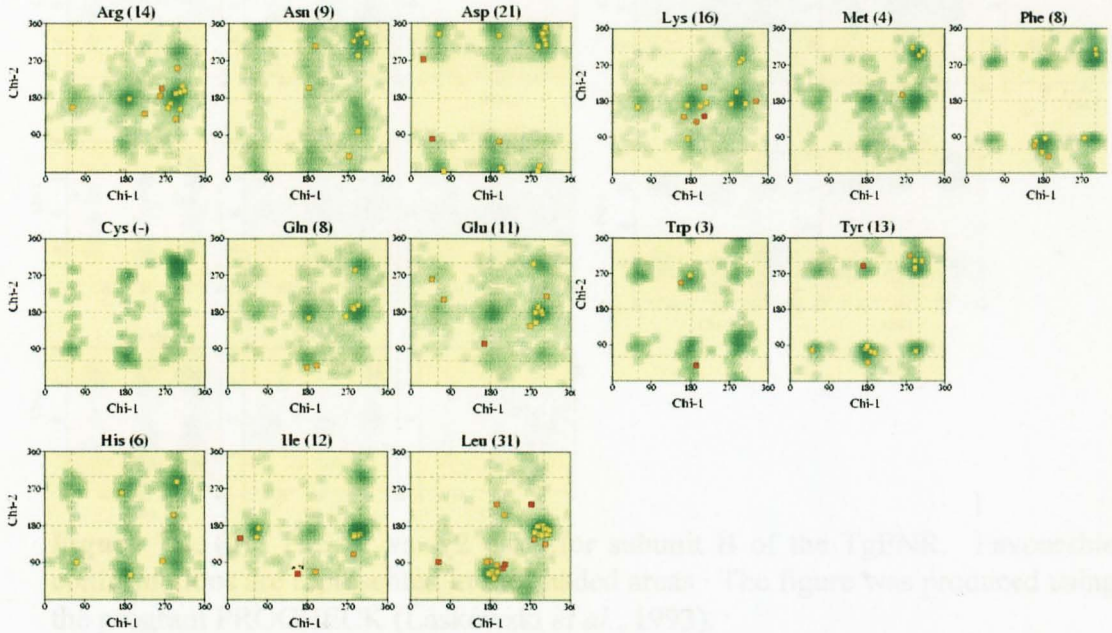


Figure 7.1. (a) The $\chi 1$ vs. $\chi 2$ plots for subunit A of the TgENR. Favourable conformations are represented in the shaded areas. The figure was produced using the program PROCHECK (Laskowski *et al.*, 1993).

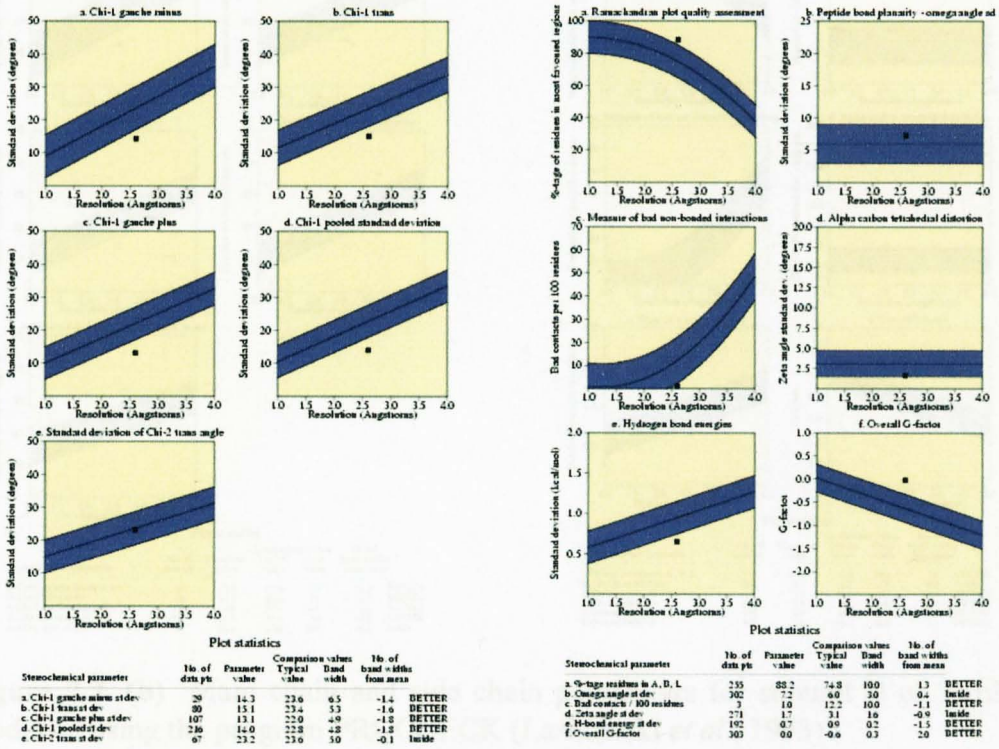


Figure 7.1. (b) Main chain and side chain parameters for subunit A, produced using the program PROCHECK (Laskowski *et al.*, 1993).

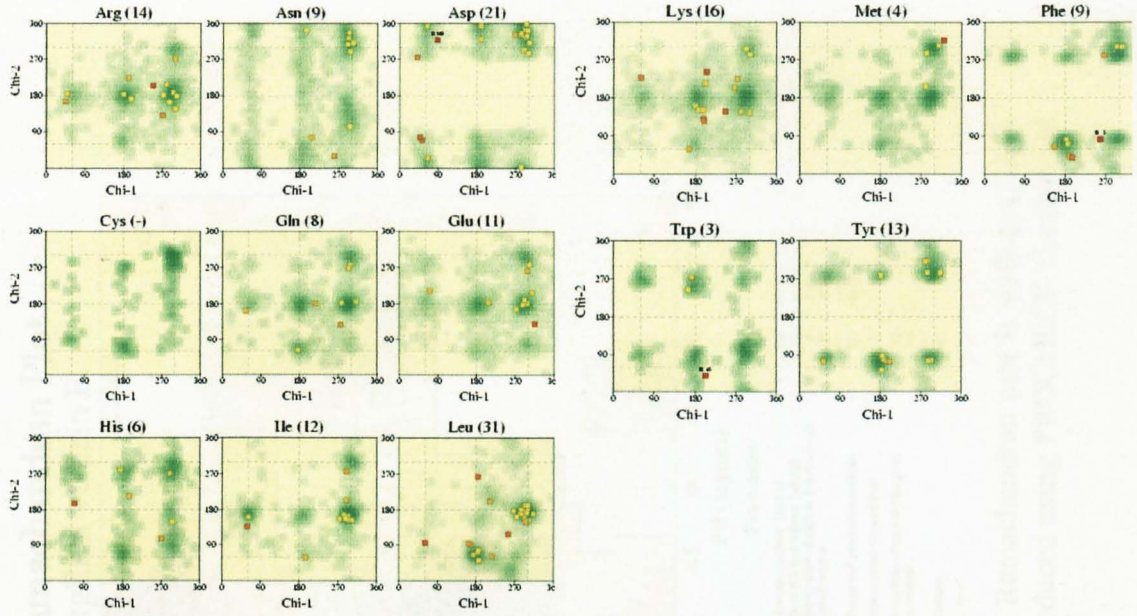


Figure 7.2. (a) The $\chi 1$ vs. $\chi 2$ plots for subunit B of the TgENR. Favourable conformations are represented in the shaded areas. The figure was produced using the program PROCHECK (Laskowski *et al.*, 1993).

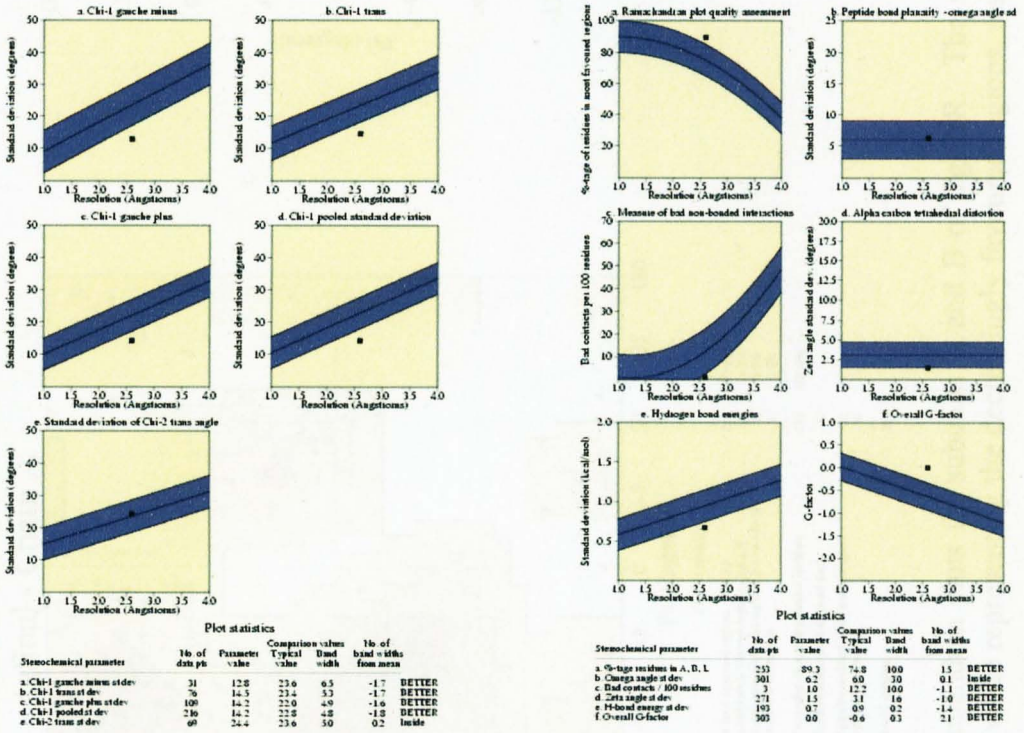


Figure 7.2. (b) Main chain and side chain parameters for subunit B of TgENR, produced using the program PROCHECK (Laskowski *et al.*, 1993).

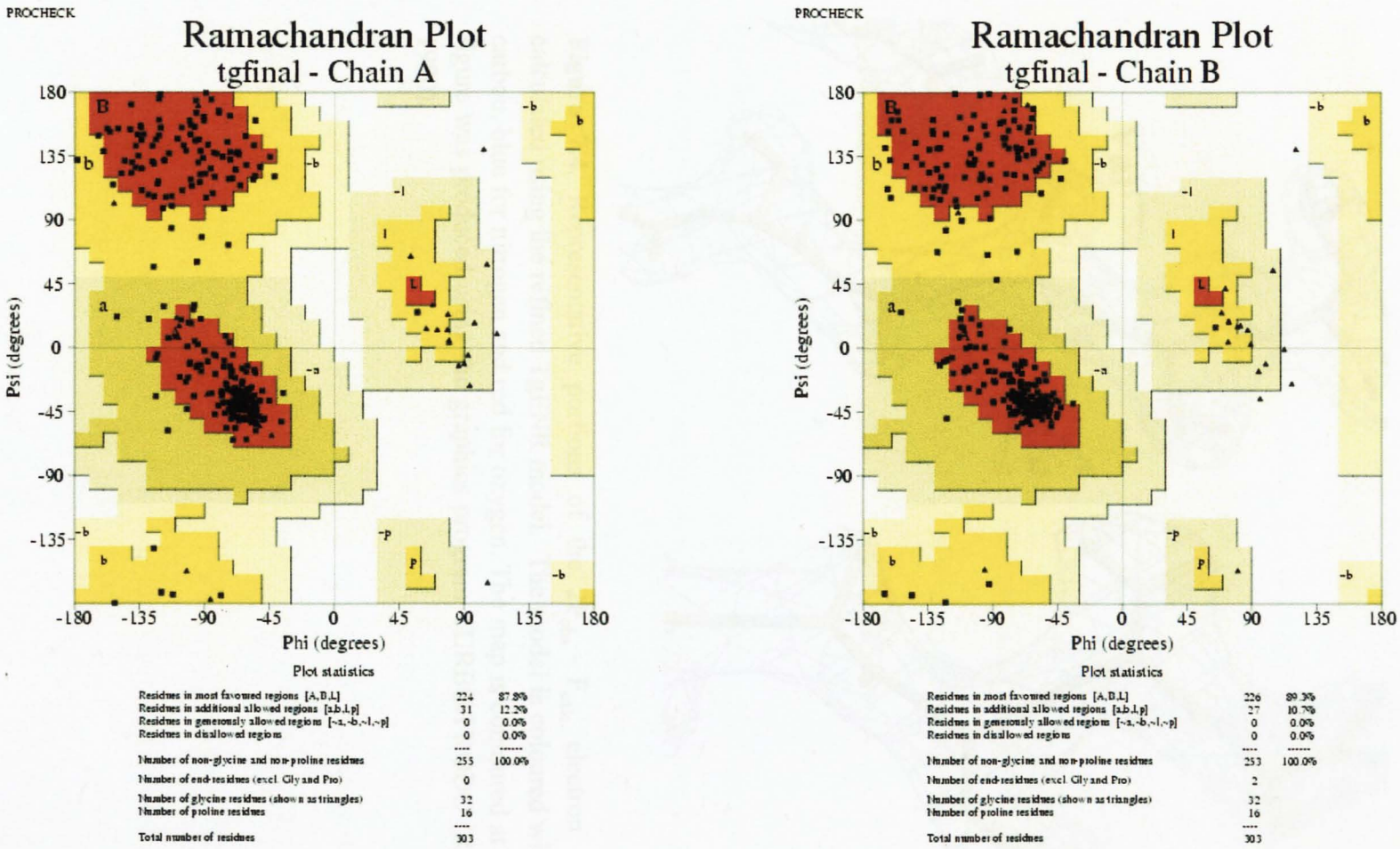


Figure 7.3. Ramachandran plots for subunits A and B of TgENR. The shading of the Ramachandran plot is with red, dark yellow, yellow and white representing the decreasingly favoured regions, respectively. Produced using PROCHECK (Laskowski *et al.*, 1993).

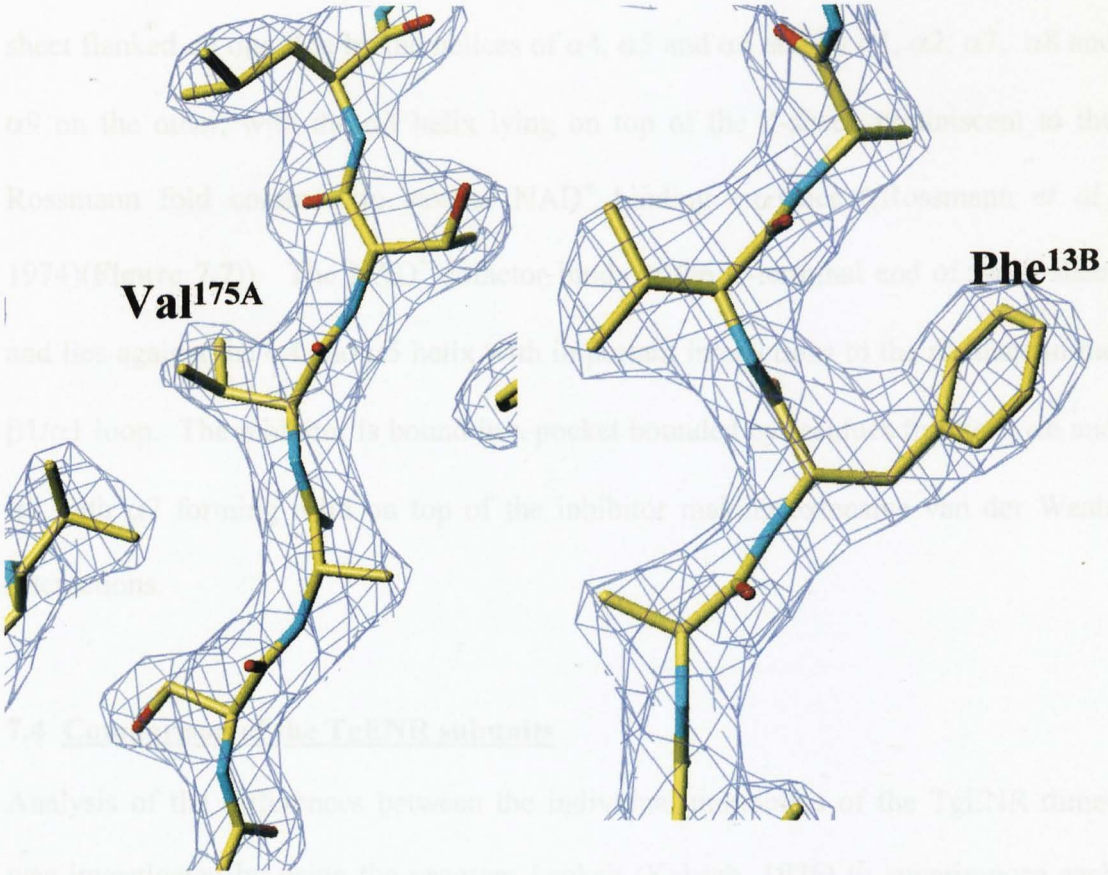


Figure 7.4. Representative portions of the $2F_{obs}$ - F_{calc} electron density map calculated using the refined TgENR model. The model is coloured with yellow for carbon, blue for nitrogen and red for oxygen. The map is contoured at 1.5σ and the figure was produced using the graphics program TURBO/FRODO (Roussel *et al.*, 1990).

formed by 43 residues (~ 14 %), six 3_{10} helices formed by 18 residues (~ 6 %) and a number of loops ((defined by the program PROMOTIF (Hutchinson and Thornton, 1996)), (**Figure 7.5** and **7.6**). The secondary structure elements form a parallel β-sheet flanked on one side by the helices of α4, α5 and α6 and by α1, α2, α7, α8 and α9 on the other, with the α3 helix lying on top of the β-sheet, reminiscent to the Rossmann fold common to several NAD^+ binding enzymes ((Rossmann *et al.*, 1974)(**Figure 7.7**)). The NAD^+ cofactor binds at the C-terminal end of the β-sheet and lies against the α4 and α5 helix with important interactions to the residues in the β1/α1 loop. The inhibitor is bound in a pocket bounded by residues from α5, α6 and α8 with α7 forming a lid on top of the inhibitor making extensive van der Waals interactions.

7.4 Comparison of the TgENR subunits

Analysis of the differences between the individual monomers of the TgENR dimer was investigated by using the program Lsqkab (Kabsch, 1976) to superimpose each monomer onto the other and calculate an r.m.s.d of the Cα atoms. An overall value of 0.4Å was obtained after overlaying monomer A onto B, showing three areas of difference at Pro^{90} to Asp^{103}, Ser^{56} to Ser^{69} and Ala^{234} to Ser^{243} (**Figure 7.8**). Interestingly these areas correspond to regions of the sequence that represent insertions when compared to the bacterial family of ENR enzymes. Removal of these areas and a recalculation of the r.m.s.d value for the Cα atoms gave an average value of 0.2Å revealing the close similarity in the conformation of the core of the molecule.

Figure 7.5. Sequence of TgENR with the secondary structural elements shown as blue cylinders and red arrows for α-helices and β-strands, respectively. Those residues not defined by density within the TgENR crystal structure are coloured purple. The figure was produced using the program Alscript (Barton, G.J., 1993).

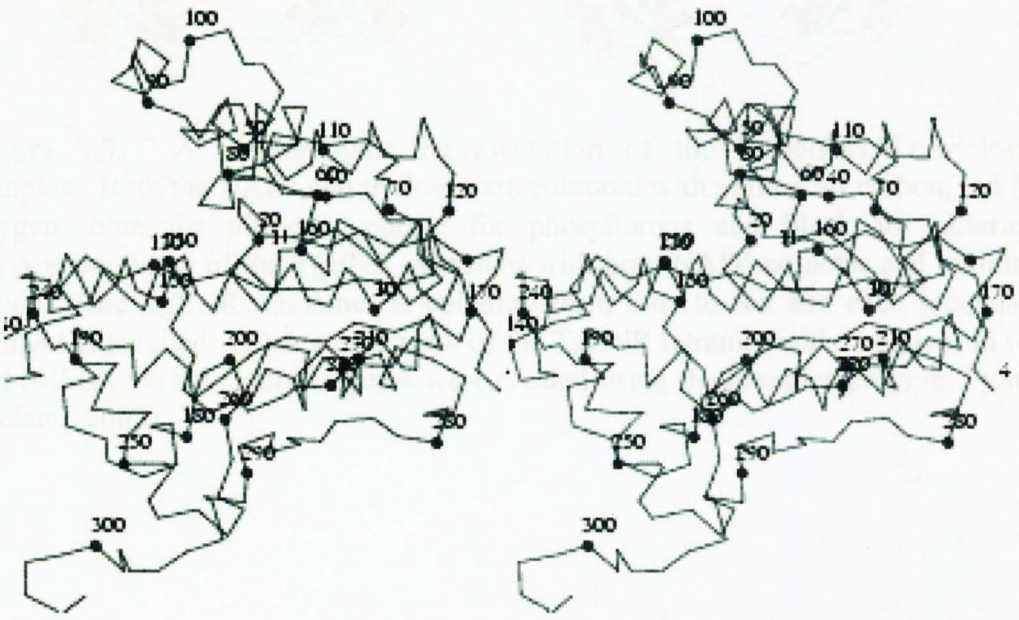


Figure 7.6. A stereo image of the backbone representation of the A subunit of TgENR with every twenty residues being numbered. The figure was produced using the program MOLSCRIPT (Kraulis, 1991).

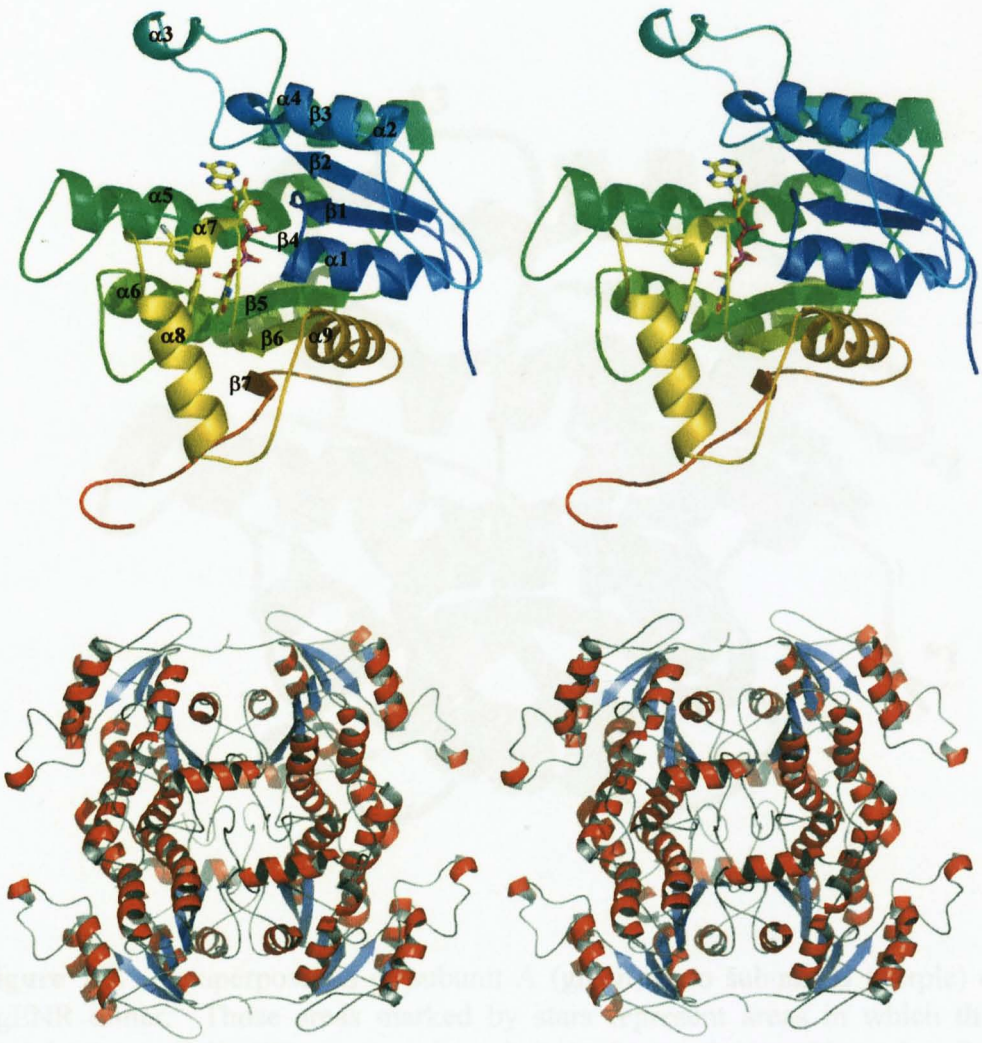


Figure 7.7. A diagrammatic representation of the TgENR/NAD^+/triclosan complex. Both the NAD^+ and triclosan are coloured with yellow for carbon, red for oxygen, blue for nitrogen, purple for phosphorous and black for chlorine. **(a)** A stereo view of the TgENR monomer with both NAD^+ cofactor and inhibitor bound. The TgENR monomer is coloured from blue to red and each secondary element is labelled. **(b)** A stereo view of the TgENR tetramer with α-helices in red and β-sheets in blue. Both figures were created using the graphics program Pymol (Delano, 2002).

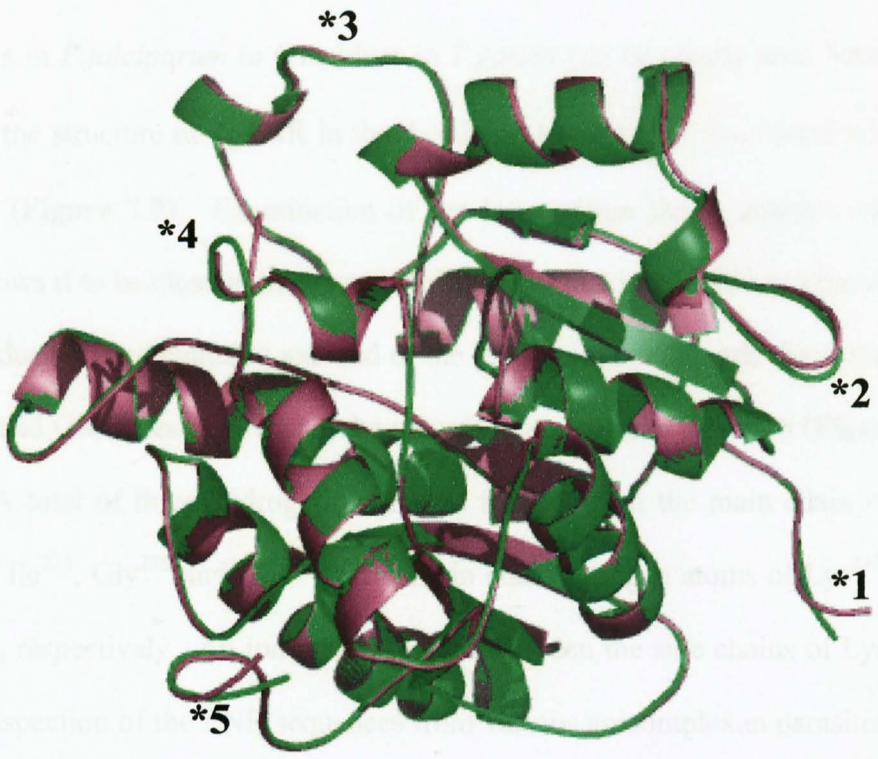


Figure 7.8. A superposition of subunit A (green) onto subunit B (purple) of the TgENR dimer. Those areas marked by stars represent areas in which the two models show slight differences, where 1 = the N terminal residues, 2 = Ser^{56} to Ser^{69}, 3 = Pro^{90} to Asp^{103}, 4 = Ala^{234} to Ser^{243} and 5 = C terminal region. The only significant region of difference is at Ala^{234} to Ser^{243} which is in the region of the insert characteristic of the apicomplexan ENR family. The figure was produced using the graphics program Pymol (Delano, 2002).

7.5 Inserted loop

The insert that is characteristic of apicomplexan ENRs and which varies in size from 42 residues in *P.falciparum* to 6 residues in *T.gondii* can be clearly seen between α7 and α8 in the structure of TgENR in the A subunit but is partly disordered within the B subunit (**Figure 7.9**). Examination of the loop within the A subunit (Gly^{236} to Lys^{241}) shows it to be close to the bound inhibitor but making no direct contacts. Two lysine residues at the beginning and end of the loop point out towards the solvent with the Ser^{238} and Glu^{240} residues facing the solvent in an opposite direction (**Figure 7.9 a and b**). A total of three hydrogen bonds are formed from the main chain carbonyl oxygen of Ile^{235}, Gly^{236} and Glu^{240} to the main chain nitrogen atoms of Lys^{237}, Ser^{238} and Ser^{242}, respectively. An ion pair is formed between the side chains of Lys^{241} and Asp^{249}. Inspection of the ENR sequences from various apicomplexan parasites shows apparent conservation of Lys^{237} and of an SGE motif, except in *Eimeria tenella* ENR (**Figure 7.10**). However with the exception of Lys^{237} the difference in insert size makes it difficult to envisage conservation of the position of the SGE motif within the apicomplexan family and the sequence similarity maybe chance resemblance. In the structure of PfENR this loop region is disordered.

7.6 Subunit interfaces and quaternary structure

The two subunits in the asymmetric unit cell of the crystal interact with its symmetry related molecule around a crystallographic two fold axis to create a tetramer identical to that seen in other ENR enzymes (Rafferty *et al.*, 1995 & Baldock *et al.*, 1998). The solvent-accessible area surface area for the TgENR monomer and tetramer was calculated to be approximately 15,500Å^2 and 42,000Å^2, respectively, using the program access (Roussel *et al.*, 1990). There are three interfaces within the TgENR

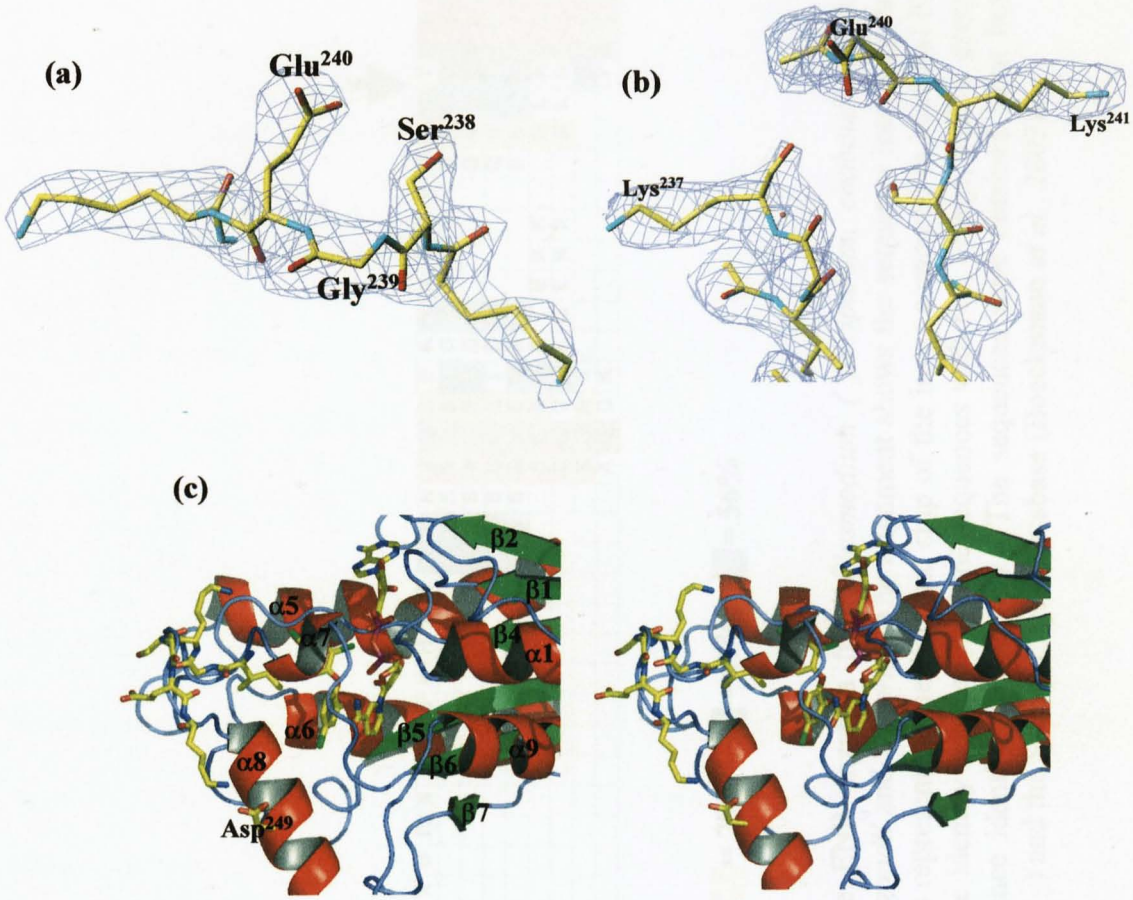


Figure 7.9. a, b and **c**. The inserted loop region of subunit A **(a)** and subunit B **(b)** of TgENR and the corresponding electron density in the final $2F_{obs}$-F_{calc} map contoured at 1σ. **(c)** A stereo view of the loop region in the A subunit of TgENR. All residues involved in forming the loop are shown along with Asp^{249} which forms an ion pair to Lys^{241}. All figures are coloured with yellow for carbon, red for oxygen and blue for nitrogen. Both figures (a) and (b) were produced in the graphics program TURBO-FRODO, figure (c) was produced using Pymol (Delano, 2002).

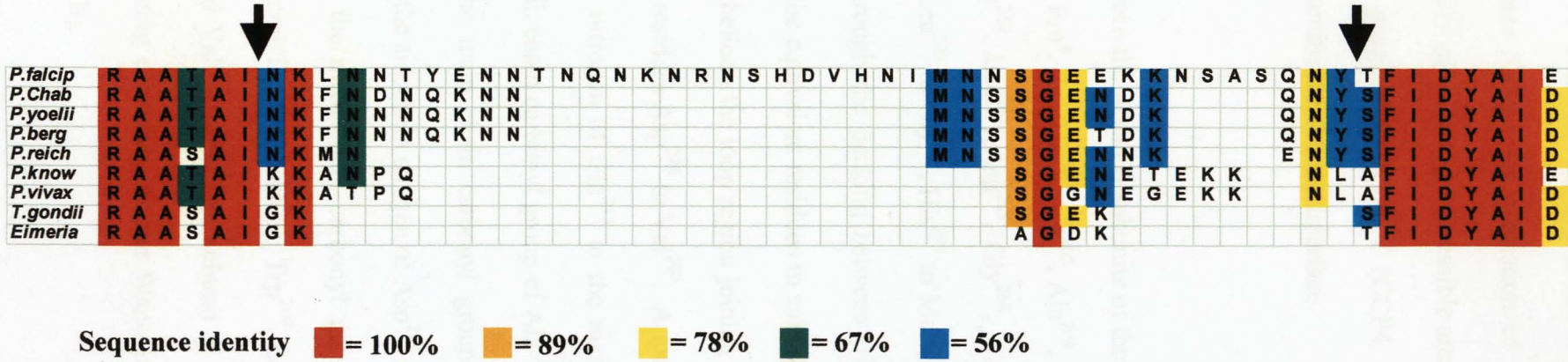


Figure 7.10. A sequence alignment of the ENR enzymes from *Plasmodium (falciparum*, *chabaudi*, *yoelii*, *berghei*, *reichenowi*, *knowlesi* and *vivax) Toxoplasma gondii* and *Eimeria*. The alignment shows the sequence inserts common to the apicomplexan parasites with two black arrows representing the start and stop of the insert compared to bacterial ENRs. The coloured boxes represent the level of sequence identity between the nine sequences with red, orange, yellow green and blue representing 100, 89, 78, 67 and 56% sequence identity, respectively. The sequences were obtained from both BLAST searches of the PlasmoDB (www.PlasmoDB.org) and the Swissprot gene database (Boeckmann *et al.*, 2003).

tetramer with contacts between subunits A-B and C-D being found on the P axis, between subunits A-C and B-D around the Q axis and between subunits A-D and B-C around the R axis (**Figure 7.11**). On formation of the tetramer approximately 2400Å^2, 1800Å^2 and 1400Å^2 of the solvent accessible area is buried by the P, Q and R axis, respectively. The program CONTACT (CCP4, 1996) together with visual inspection was used to examine the subunit interface.

7.6.1 The P axis

The P-axis interface between the A and B subunits of the TgENR dimer and tetramer, is formed by 40 residues: Pro^4, Ile^5, His^{30}, Arg^{205}, Ala^{208}, Trp^{209}, Gly^{212}, Gln^{213}, Gly^{216} to Arg^{218}, Asn^{254} to Arg^{259}, Leu^{261}, Asp^{265}, Gly^{268}, Ala^{269}, Leu^{271}, Phe^{272}, Leu^{277}, Arg^{279}, Ala^{280} to Val^{284}, Leu^{286} to Gly^{291}, His^{293} to Met^{295} and Val^{299} (**Figure 7.12 a**). The P-axis is formed through the packing between β7 and α9 and its preceding residues of subunit A to the equivalent residues in subunit B with additional contacts by the packing of the α6 helices, the loop region joining α9 to β7 and the C terminal loop region composed of residues Asp^{289} to Val^{299}. A total of six hydrogen bonds are formed by the sidechain nitrogen of Gln^{213} to the main chain carbonyl groups of Ala^{255} and Pro^{256}, the main chain carbonyl group of Ala^{280} to the main chain nitrogen of Asn^{290} and Asp^{289}, the main chain carbonyl group of Ser^{282} to the sidechain nitrogen of His^{293} and by the sidechain oxygen of Asp^{289} to a water molecule which in turn hydrogen bonds to the main chain carbonyl group of Ser^{282}. Hydrophobic packing is also present involving the residues: Trp^{209}, Leu^{257}, Gly^{268}, Phe^{272}, Leu^{277}, Arg^{279}, Val^{281}, Leu^{286} and Val^{288}. Phe^{272} of subunit A packs against its equivalent residue in subunit B forming extensive van der Waals interactions and π-π stacking interactions (**Figure 7.12 b**).

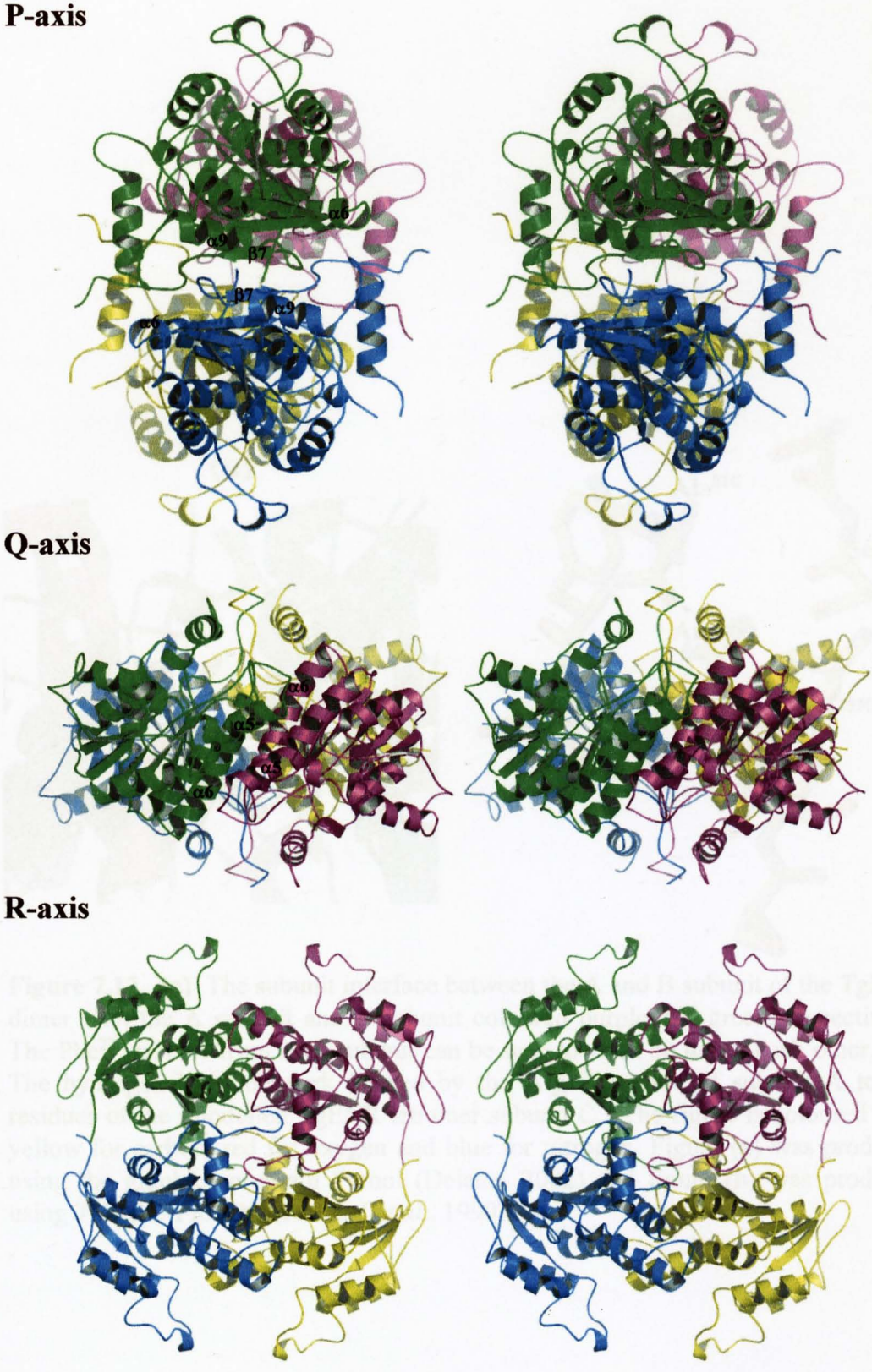


Figure 7.11. Three views of the modelled TgENR tetramer viewed down each of the three molecular two fold axis. The axis are coloured with subunit A in green, B in blue, C in purple and D in yellow. The figure was produced using the graphics program Pymol (Delano, 2002).

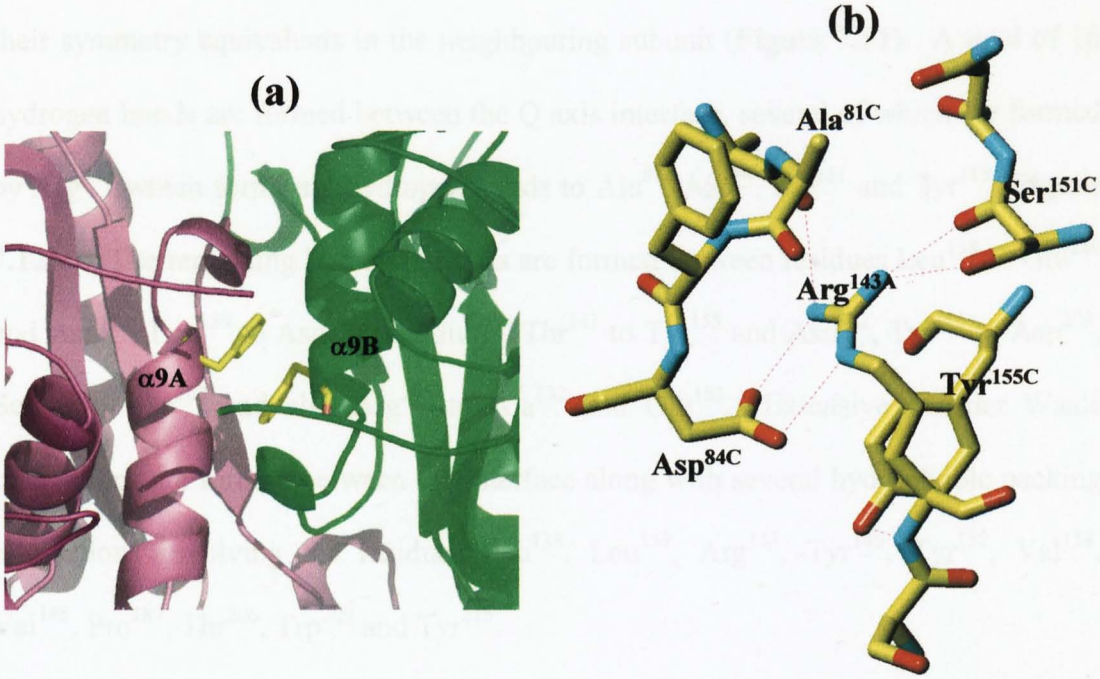


Figure 7.12. **(a)** The subunit interface between the A and B subunit of the TgENR dimer with the A subunit and B subunit coloured purple and green, respectively. The Phe^{272} residue from each subunit can be seen to stack on top of each other. **(b)** The hydrogen bond network formed by the Arg^{143} residue of subunit A to the residues of the modelled TgENR tetramer subunit C. The figure is coloured with yellow for carbon, red for oxygen and blue for nitrogen. Figure **(a)** was produced using the graphics program Pymol (Delano, 2002) and figure **(b)** was produced using TURBO-FRODO (Roussel *et al.*, 1990).

7.6.2 The Q axis

The Q axis interface is formed by 44 residues: Ala^{81} to Asp^{84}, Ile^{107}, Pro^{137} to Arg^{143}, Tyr^{146}, Leu^{147}, Ser^{150} to Asn^{152}, Ala^{154} to Ser^{156}, Val^{158}, Ser^{159}, Leu^{161}, Gln^{162}, Ala^{181} to Val^{185}, Pro^{187}, Ser^{194}, Ser^{195}, Ala^{198}, Ala^{199}, Ser^{202}, Asp^{203}, Arg^{205} to Leu^{207}, Trp^{209} to Ala^{211}, Lys^{214} and Tyr^{215}. The Q axis interface is dominated by a four helix bundle formed by helices α5 and α6 and their preceding loop regions which pack against their symmetry equivalents in the neighbouring subunit (**Figure 7.11**). A total of 16 hydrogen bonds are formed between the Q axis interface, several of which are formed by Arg^{143} which forms six hydrogen bonds to Ala^{81}, Asp^{84}, Ser^{151} and Tyr^{155} (**Figure 7.12 c**). The remaining hydrogen bonds are formed between residues Leu^{138} to Glu^{210} and Asn^{162}, Leu^{139} to Asn^{162} and Glu^{210}, Thr^{141} to Tyr^{155} and Asn^{162}, Tyr^{146} to Asp^{203}, Ser^{195} to Asp^{203} and also Arg^{205} to Ala^{182} and Glu^{183}. Extensive van der Waals interactions are formed between the interface along with several hydrophobic packing interactions involving the residues Leu^{138}, Leu^{139}, Arg^{143}, Tyr^{146}, Tyr^{155}, Val^{158}, Val^{185}, Pro^{187}, Thr^{206}, Trp^{209} and Tyr^{215}.

7.6.3 The R axis

The smallest of the interfaces is that around the R-axis which involves 29 residues (Tyr^{179}, Leu^{180}, Glu^{183} to Pro^{187}, Lys^{241}, Tyr^{246}, Ala^{247}, Asp^{249} to Ser^{251}, Asn^{253}, Asn^{254} and His^{293} to Pro^{306}). The main interactions are formed between helix α8 and the N terminal loop region and their symmetry related residues. A total of 5 hydrogen bonds are formed between Val^{185} to Ala^{294} and Gly^{296}, Tyr^{250} to Ala^{298}, Asn^{253} to Ser^{303} and Arg^{302} to Met^{304}. Hydrophobic packing interactions are formed by 6 residues: Arg^{184}, Pro^{187}, Tyr^{246}, Tyr^{250}, Met^{304} and Pro^{305}.

7.7 Cofactor and inhibitor binding sites

7.7.1 Triclosan binding site

The electron density for the triclosan inhibitor within the TgENR/NAD^+/triclosan complex was of very good quality and allowed for its unambiguous placement into the active site of both subunits (**Figure 7.13a**). The 4-chloro phenoxy ring of triclosan (ring A) interacts in a face to face manner with the NAD^+ nicotinamide unit allowing π-π stacking interactions and the formation of van der Waals interactions with the side chains of Tyr^{179}, Tyr^{189}, Pro^{226}, Phe^{243}, Ala^{232}, Ile^{244} and Ile^{235}. Furthermore the triclosan inhibitor forms one hydrogen bond to Tyr^{189}. The 2,4-dichlorophenoxy ring of triclosan (ring B) is located within a pocket bounded by the peptide backbone of residues Leu^{128} to Gly^{131}, the pyrophosphate and nicotinamide moieties of NAD^+ and by the side chains of Val^{134}, Met^{193}, Ala^{231} and Ile^{235} (**Figure 7.13b**).

7.7.2 NAD^+ binding site

Density for the entire nicotinamide adenine dinucleotide (NAD^+) cofactor could be clearly seen for both subunits permitting its unambiguous identification (**Figure 7.14a**). The adenine ring binds within a pocket formed by the main chain atoms of residues: Gly^{16} to Ala^{18}, Leu^{79} to Ala^{82} and Leu^{128} to Asn^{130} and the side chains of residues: Trp^{43}, Leu^{128}, Asn^{130}, Asp^{80}, and Asn^{152}. One face of the adenine ring forms π–π stacking interactions with Trp^{43}, with Leu^{128} making van der Waals interactions with the other. A hydrogen bond network allows for the binding and positioning of the NAD^+ substrate into the active site (**Figure 7.14 b**). In both subunits the adenine ring is hydrogen bonded by the carbonyl of Asp^{80}. The adenine ribose forms van der Waals interactions with the main chain of Gly^{16} to Ala^{18} and the side chain residues of Leu^{128} and Trp^{43}. The pyrophosphate moiety makes interactions with the main chain

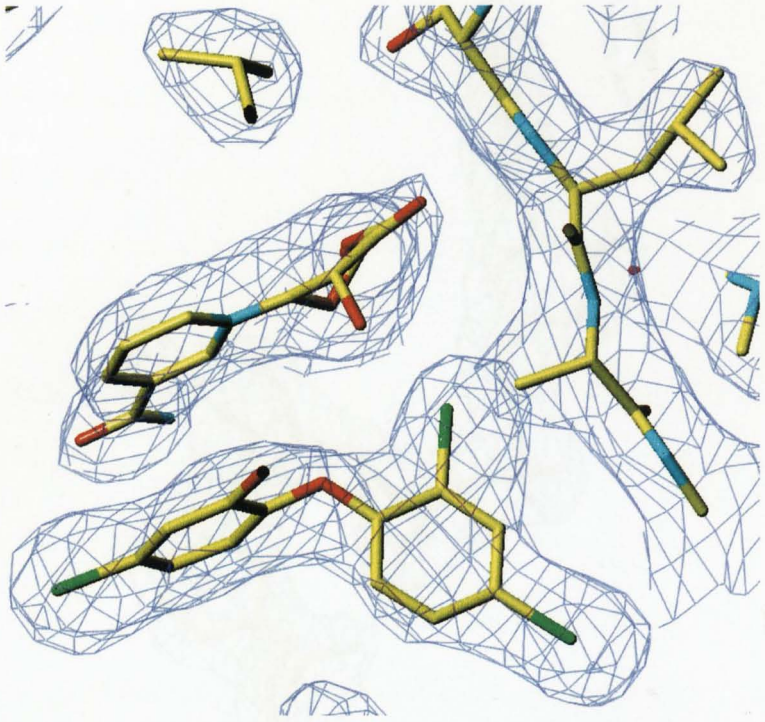


Figure 7.13. (a) Triclosan bound within the TgENR active site and its corresponding electron density $2F_{obs}$-F_{calc} map contoured at 1.0σ. The NAD^+ cofactor can also be seen packing in a face to face manner with the triclosan inhibitor. The model is coloured with yellow for carbon, blue for nitrogen, red for oxygen and green for chlorine. This figure is based on subunit A of the TgENR dimer. The figure was produced using the program TURBO-FRODO (Roussel *et al.*, 1990) .

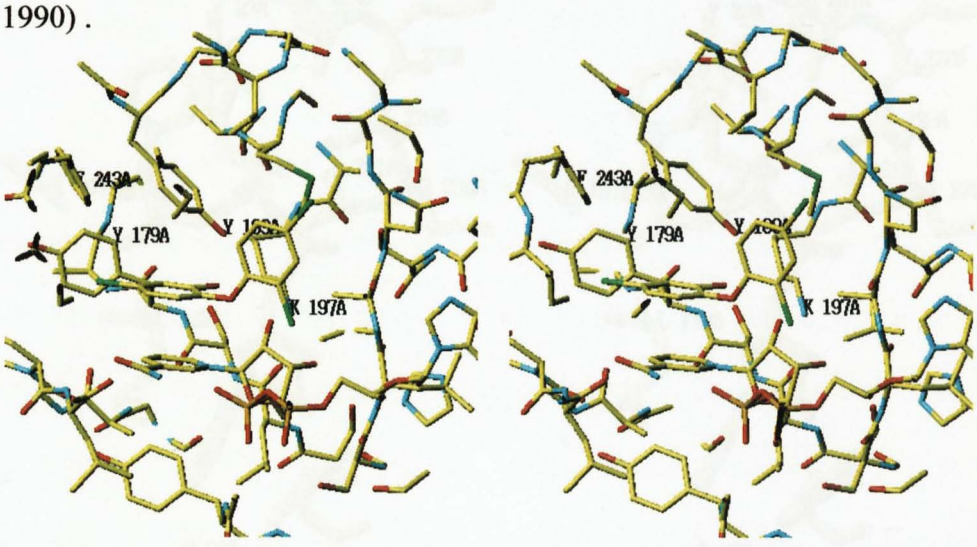


Figure 7.13. (b) A stereo diagram of the triclosan binding site of subunit A of TgENR with the active site residues Tyr^{179}, Tyr^{189}, Lys^{197} and Phe^{243} labelled. The figure is coloured with yellow for carbon, red for oxygen, orange for phosphorous, blue for nitrogen and green for the chlorine atoms of triclosan and the sulphur atom of the active site Met. The figure was produced using the program TURBO-FRODO (Roussel *et al.*, 1990).

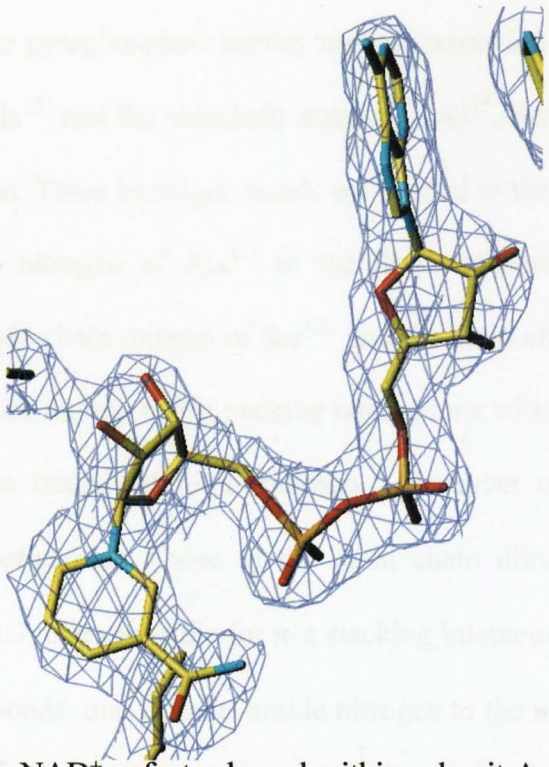


Figure 7.14. (a) The NAD^+ cofactor bound within subunit A of the TgENR active site and its corresponding $2F_{obs}$-F_{calc} electron density map contoured at 1.0σ. The model is coloured with yellow for carbon, blue for nitrogen, red for oxygen and orange for phosphorous. The figure was produced using the program TURBO-FRODO (Roussel *et al.*, 1990).

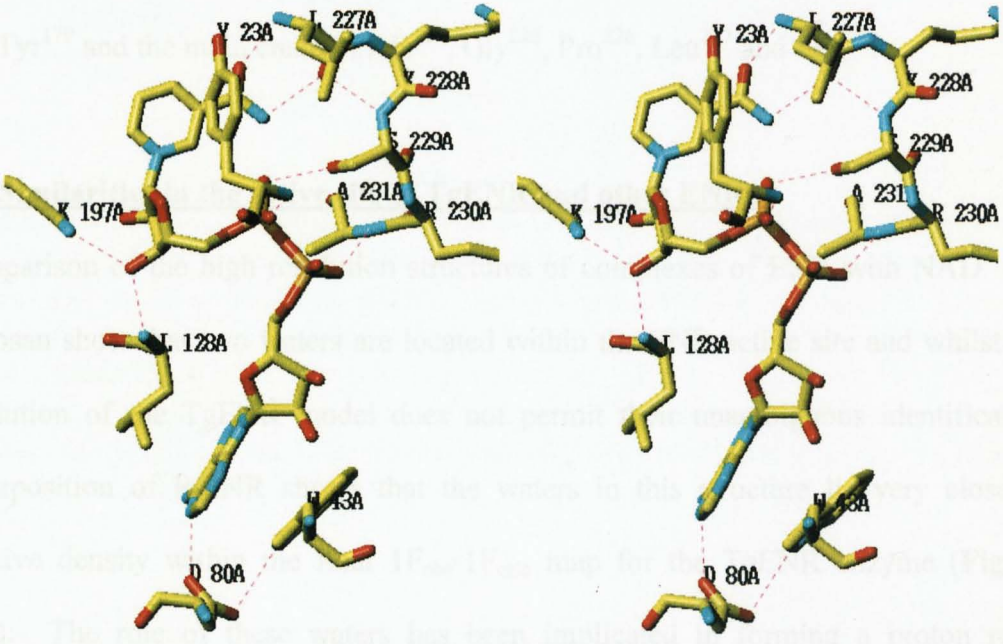


Figure 7.14. (b) A stereo diagram of the residues responsible for forming a hydrogen bonding network to the NAD^+ cofactor in subunit A of TgENR. All the residues are labelled and coloured with yellow for carbon, blue for nitrogen, red for oxygen and orange for phosphorous. The hydrogen bonds are represented as red dashed lines. The figure was produced using the program TURBO-FRODO (Roussel *et al.*, 1990).

Leu[128] and Trp[43]. The pyrophosphate moiety makes interactions with the main chain of Gly[22], Tyr[23] and Ala[231] and the sidechain atoms of Asp[19], Ser[229], Ala[231] and the di-chloro ring of triclosan. Three hydrogen bonds are formed to the pyrophosphate group from the main chain nitrogen of Ala[231] to the first phosphate and to the second phosphate from the side chain oxygen of Ser[129] and the main chain nitrogen of Tyr[23]. The nicotinamide ribose moiety forms packing interactions with: Tyr[23], Ser[127], Ala[129], Leu[177], Lys[197] and the bound inhibitor triclosan. Moreover one hydrogen bond is formed from the nicotinamide ribose to the main chain nitrogen of Leu[128]. The nicotinamide ring which is responsible for π-π stacking interactions with the triclosan, forms two hydrogen bonds, one from its amide nitrogen to the main chain carbonyl of the Leu[227] unit and a second from its amide oxygen of the main chain nitrogen of Leu[227]. In addition to the close packing interactions with the triclosan inhibitor the nicotinamide ring forms interactions to the side chain atoms of: Tyr[23], Ser[229], Ala[224] and Tyr[179] and the main chain of Ala[224], Gly[225], Pro[226], Leu[227] and Ser[178].

7.8 Similarities in the active site of TgENR and other ENRs

Comparison of the high resolution structures of complexes of ENR with NAD^+ and triclosan show that two waters are located within the ENR active site and whilst the resolution of the TgENR model does not permit their unambiguous identification superposition of PfENR shows that the waters in this structure lie very close to positive density within the final $1F_{obs}$-$1F_{calc}$ map for the TgENR enzyme (**Figure 7.15**). The role of these waters has been implicated in forming a proton relay mechanism to the bulk solvent ((Section 4.8)(Benach *et al.*, 1998 & Filling *et al.*, 2002))

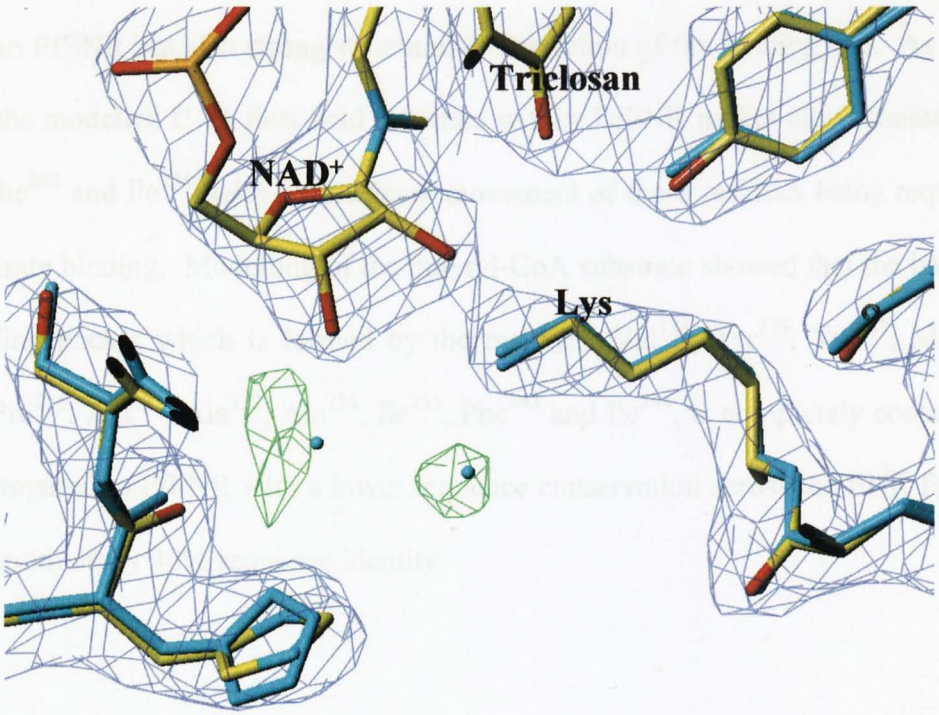


Figure 7.15. The TgENR active site with the $2F_{obs} - 1F_{calc}$ electron density map shown in purple, contoured at 1σ and the $1F_{obs} - 1F_{calc}$ electron density map contoured at 3σ coloured in green. The PfENR structure, coloured in blue is superimposed onto the TgENR with the two PfENR active site waters, shown as blue spheres being present in positive areas of the $1F_{obs} - 1F_{calc}$ map. The TgENR structure is coloured with yellow for carbon, blue for nitrogen, red for oxygen and orange for phosphorous. The figure was produced using the program TURBO-FRODO (Roussel *et al.*, 1990).

Substrate modelling studies for both the butyryl-CoA and C16 fatty acid substrate in a manner as previously described for PfENR (Section 4.8) showed that the binding pocket for the C16 fatty acid substrate which consists of the predominantly hydrophobic residues (Ala^{129}, Asn^{130}, Val^{134}, Tyr^{179}, Ala^{181}, Val^{186}, Pro^{187}, Gly^{188}, Tyr^{189}, Met^{193}, Lys^{197}, Ala^{231}, Ala^{232}, Ile^{235}, Phe^{243}) shares both 100% sequence identity to PfENR but also strong structural conservation of the binding site. As with PfENR the modelled C-16 fatty acid substrate within TgENR makes close contacts to Tyr^{179} Phe^{243} and Ile^{235}, with a significant movement of these residues being required for substrate binding. Modelling of the butyryl-CoA substrate showed that the base of the binding pocket which is formed by the residues: Val^{134}, Tyr^{179}, Tyr^{189}, Met^{193}, Lys^{197}, Pro^{226}, Ala^{231}, Ala^{232}, Ala^{234}, Ile^{235}, Phe^{243} and Ile^{244}, is completely conserved when compared to PfENR with a lower sequence conservation across the ENR family with approximately 30% sequence identity.

CHAPTER 8

Sequence and structural comparisons between the ENR family

8.1 Structural comparisons across the ENR family

In order to investigate the similarities and differences between members of the ENR family the ENR structures for the *B.napus*, *E.coli*, *T.gondii* and *P.falciparum* in complex with NAD^+ and triclosan were used for comparative studies. By superimposing all of the ENR structures onto each other (**Table 8.1**) it was possible to see that all have a very similar core structure made up of a well conserved central parallel β-sheet with strong positional conservation of several flanking α-helices in particular the helices 1, 4, 5, 6 and 9 (TgENR structure numbering) (**Figure 8.1**). The Cα atom r.m.s.d variation of each of the ENR structures when superimposed onto each other is low with the highest difference being between the EcENR to the *B.napus* and apicomplexan ENR structures. A strong similarity was noted between the two members of the apicomplexan ENR family with an average r.m.s.d value for the Cα atoms being 0.8Å. The closest members of the ENR family based on the currently available structures are between the *T.gondii* and *B.napus* ENR enzymes with an r.m.s.d value for the Cα atoms of 0.6Å. The plant and apicomplexan enzymes both contain significant inserts both before and after β3 compared to the bacterial enzymes (**Figure 8.2**). These inserts form extended loops including a small additional helix which runs parallel to α2. Analysis of the structure based sequence alignment shows the vast majority of conserved residues lie within the active site and NAD^+ binding pocket (**Figure 8.2** and **Figure 8.3**).

8.2 Surface comparisons of the ENR family

In order to investigate the surface characteristics of each of the solved ENR structures the program GRASP (Nicholls *et al.*, 1991) was used to calculate the surface potential for the apicomplexan *P.falciparum* and *T.gondii* ENRs and the representative plant

	B.napus	*E.coli*	*P.falciparum*	*T.gondii*
B.napus		1	0.7	0.6
E.coli	1		0.9	1
P.falciparum	0.7	0.9		0.8
T.gondii	0.6	1	0.8	

Table 8.1. A table to show the r.m.s.d of all Cα atoms upon superposition of the ENR structures currently available. All values were calculated using the program Lsqkab (Kabsch, 1976) and are in Å.

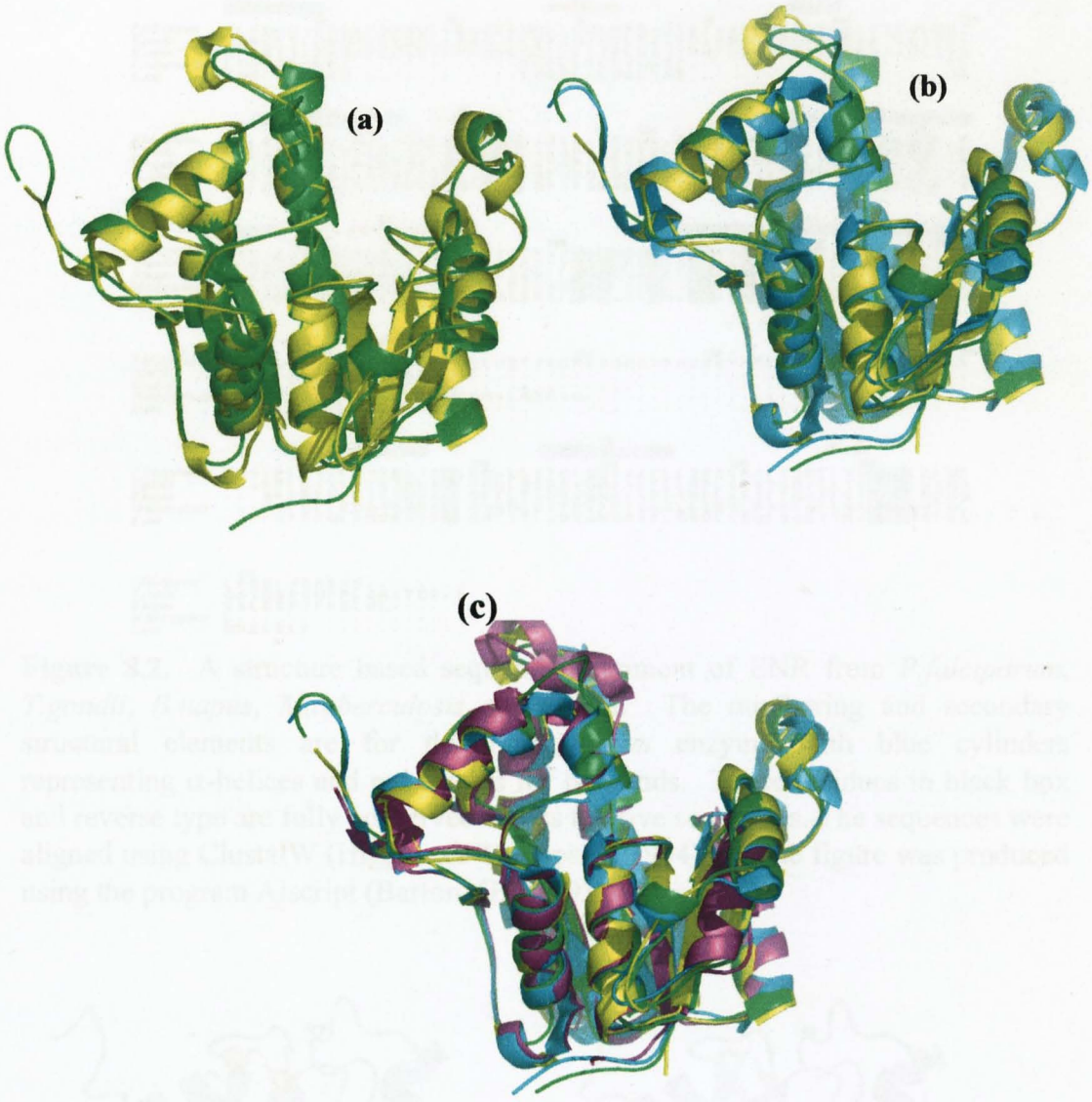


Figure 8.1. The superposition of the ENR family of enzymes with the *T.gondii*, *P.falciparum*, *B.napus* and *E.coli* ENR enzymes coloured in green, yellow, blue and magenta, respectively. (a) The superposition of the *T.gondii* and *P.falciparum* ENR enzyme. (b) The superposition of the *T.gondii*, *P.falciparum* and *B.napus* ENR enzyme. (c) The superposition of the *T.gondii*, *P.falciparum*, *B.napus* and *E.coli* ENR enzyme. All figures were produced using the graphics program Pymol (Delano, 2002).

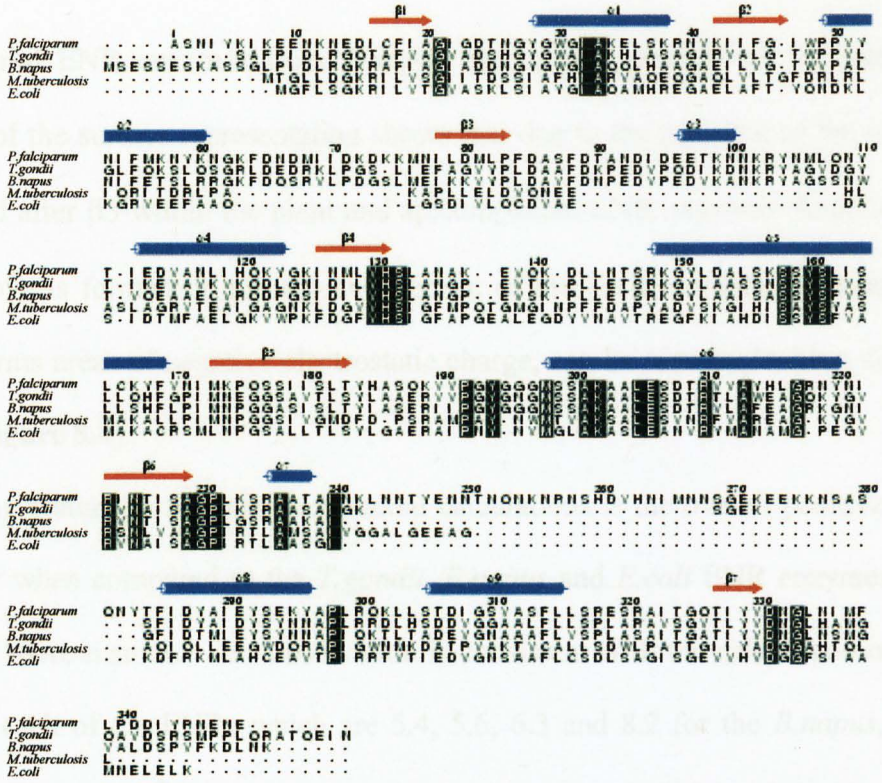


Figure 8.2. A structure based sequence alignment of ENR from *P.falciparum*, *T.gondii*, *B.napus*, *M.tuberculosis* and *E.coli*. The numbering and secondary structural elements are for the *P.falciparum* enzyme with blue cylinders representing α-helices and red arrows for β-strands. Those residues in black box and reverse type are fully conserved across the five sequences. The sequences were aligned using ClustalW (Higgins & Thompson, 1994) and the figure was produced using the program Alscript (Barton, G.J. 1993).

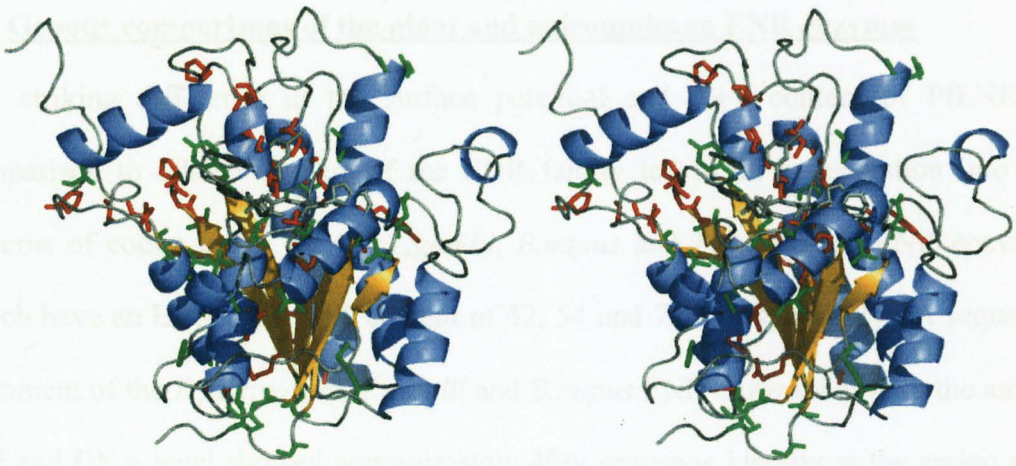


Figure 8.3. A stereo view of the TgENR enzyme with those residues fully conserved between the *B.napus*, *E.coli*, *M.tuberculosis*, *T.gondii* and *P.falciparum* ENR structures in red. Those residues for which there is strong sequence identity (>80%) are shown in green. The figure was produced using the graphics program Pymol (Delano, 2002).

and bacterial ENR enzymes from *B.napus* and *E.coli*, respectively. (**Figure 8.4**). Analysis of the surface representation shows that due to the presence of the insertion before and after β3 within the plant and apicomplexan ENR enzymes (Section 8.1) a large groove is formed on the top and bottom of the ENR tetramer. Moreover the groove forms areas of negative electrostatic charge, can be seen by looking down the Q-axis (**Figure 8.4**).

A striking feature of the surface potential calculations is the overall positive charge of PfENR when compared to the *T.gondii, B.napus* and *E.coli* ENR enzymes, all of which are more neutral. This observation is consistent with the calculation of the pI value for each of the ENRs which are 5.4, 5.6, 6.3 and 8.2 for the *B.napus*, *E.coli*, *T.gondii* and *P.falciparum* enzymes, respectively. One possible explanation may be that in order to maintain both a hydrophilic exterior and a high A+T gene content a selective pressure must be placed upon those amino acids which are both hydrophilic and have a low G+C codon content which includes the positively charged lysine.

8.3 Genetic comparisons of the plant and apicomplexan ENR enzymes

The striking difference in the surface potential and A+T content of PfENR in comparison to other members of the ENR family led to an investigation into the patterns of codon usage for the *T.gondii*, *B.napus* and *P.falciparum* ENR enzymes which have an ENR gene A+T content of 42, 54 and 72%, respectively. A sequence alignment of the *P.falciparum*, *T.gondii* and *B.napus* ENR enzymes at both the amino acid and DNA level showed approximately 40% sequence identity at the amino acid level with a much larger differences between the sequences at the DNA level with only 6% codon identity (**Figure 8.5**). Analysis of the percentage of each amino acid within the different ENR sequences revealed a strong bias within *P.falciparum* for

and bacterial ENR enzymes from *B.napus* and *E.coli*, respectively. (**Figure 8.4**). Analysis of the surface representation shows that due to the presence of the insertion before and after β3 within the plant and apicomplexan ENR enzymes (Section 8.1) a large groove is formed on the top and bottom of the ENR tetramer. Moreover the groove forms areas of negative electrostatic charge, can be seen by looking down the Q-axis (**Figure 8.4**).

A striking feature of the surface potential calculations is the overall positive charge of PfENR when compared to the *T.gondii, B.napus* and *E.coli* ENR enzymes, all of which are more neutral. This observation is consistent with the calculation of the Pi value for each of the ENRs which are 5.4, 5.6, 6.3 and 8.2 for the *B.napus*, *E.coli*, *T.gondii* and *P.falciparum* enzymes, respectively. One possible explanation may be that in order to maintain both a hydrophilic exterior and a high A+T gene content a selective pressure must be placed upon those amino acids which are both hydrophilic and have their respective codon low in G+C. These residues would include lysine.

8.3 Genetic comparisons of the plant and apicomplexan ENR enzymes

The striking difference in the surface potential and A+T content of PfENR in comparison to other members of the ENR family led to an investigation into the patterns of codon usage for the *T.gondii*, *B.napus* and *P.falciparum* ENR enzymes which have an ENR gene A+T content of 42, 54 and 72%, respectively. A sequence alignment of the *P.falciparum*, *T.gondii* and *B.napus* ENR enzymes at both the amino acid and DNA level showed approximately 40% sequence identity at the amino acid level with a much larger differences between the sequences at the DNA level with only 6% codon identity (**Figure 8.5**). Analysis of the percentage of each amino acid within the different ENR sequences revealed a strong bias within *P.falciparum* for

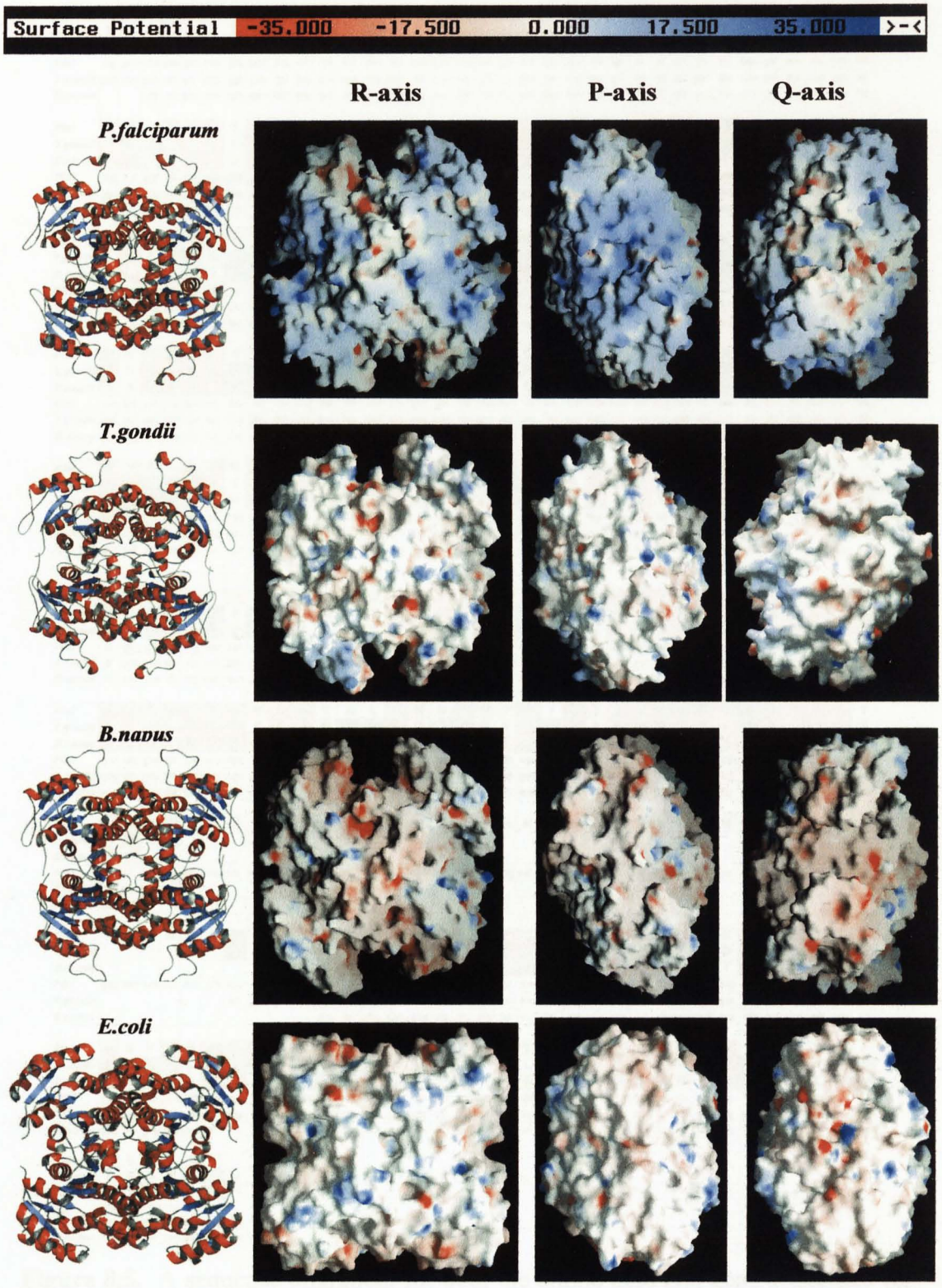


Figure 8.4. A surface potential representation for the *P.falciparum*, *T.gondii*, *B.napus* and *E.coli* ENR enzymes as seen down the P, Q and R axes. All four structures were placed on the same scale as shown at the top of the figure and coloured by electrostatic charge ranging from –35.0 in red to +35.0 in blue Kcal/[mol·electron]. The surface potential figures were produced using the program GRASP (Nicholls *et al.*, 1991). An approximate view of each of the ENR enzymes down the R-axes without a surface representation is shown on the left hand side and were created using the program Pymol (Delano, 2002).

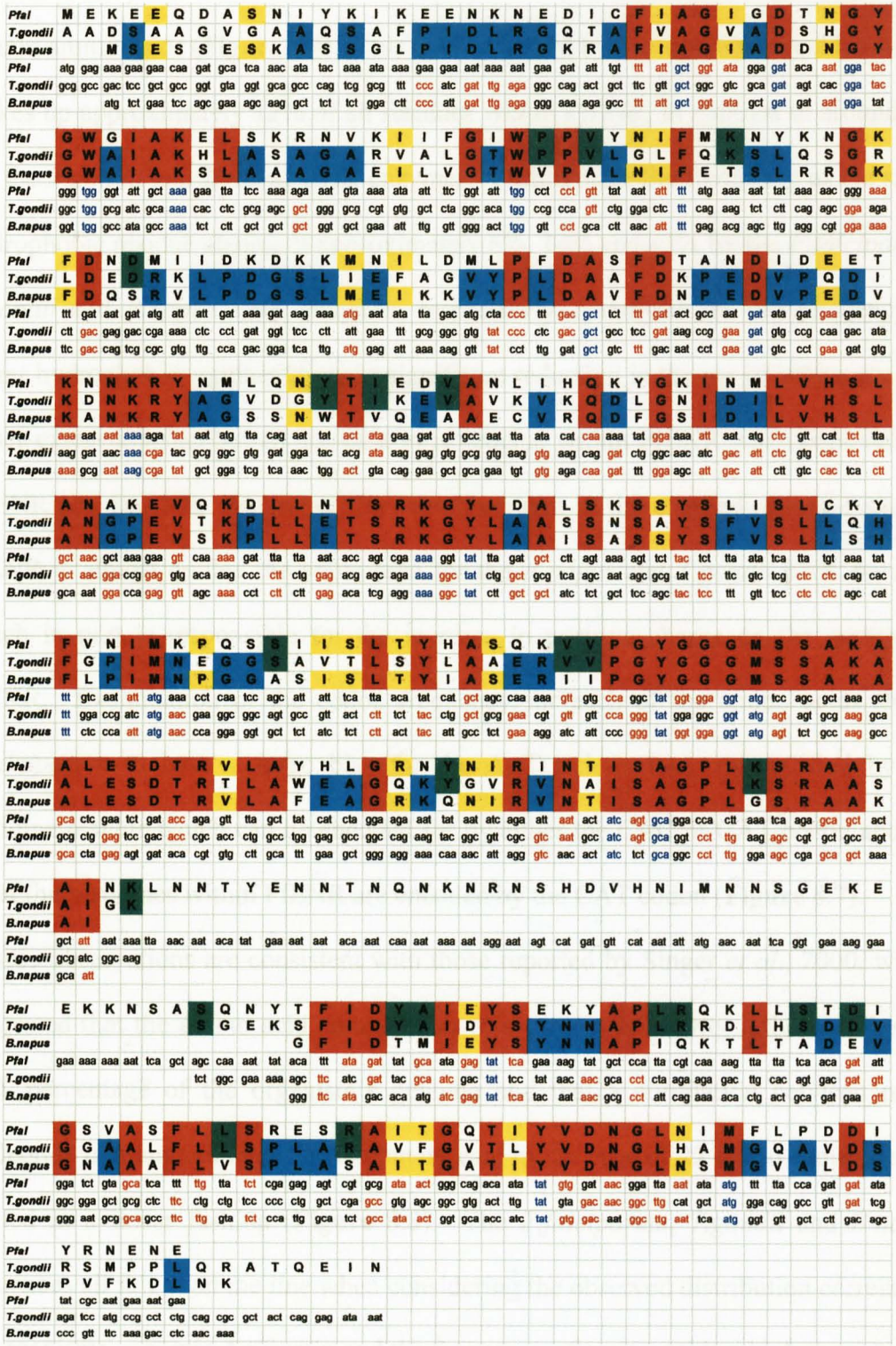


Figure 8.5. A sequence alignment for both the amino acid and gene sequences for the *P.falciparum*, *T.gondii* and *B.napus* ENR enzyme sequences. Those residues which are fully conserved are highlighted by a red box, those which are conserved between the *P.falciparum* and *T.gondii* sequences are highlighted by a green box, those residues similar between the *P.falciparum* and *B.napus* sequences are highlighted by a yellow box and those residues conserved between the *T.gondii* and *B.napus* sequences are highlighted by a blue box. Those codons which are conserved between two sequences are in red and those fully conserved are in blue.

certain amino acids (**Figure 8.6**). This is exemplified by the asparagine residue which makes up approximately 12% of all residues within PfENR, in contrast to around 4% for the *T.gondii* and *B.napus* ENR enzymes. Furthermore the frequency of the small hydrophobic residue alanine is approximately 6%, 14% and 13% for the *P.falciparum*, *T.gondii* and *B.napus* ENR enzymes, respectively. Analysis for the frequency of larger hydrophobic residues shows a significant reduction in the use of valine and to a lesser extent leucine, but an increase in isoleucine within PfENR when compared to TgENR and BnENR. Moreover the overall content of the charged residue lysine in PfENR is almost double the value seen for the *B.napus* and *T.gondii* ENR enzyme, which in part is responsible for the increase in positive charge on the surface of PfENR. Examination of the codon frequency for each of the amino acids shows *P.falciparum* to have a large bias for the A+T rich codons, exemplified by the histidine residue which can be encoded for by either the CAC or CAT codon and yet all of the histidines within PfENR are encoded by the CAT codon (**Figure 8.7**).

These observations are consistent with those reported by Singer *et al.*, 2000, which suggested that the amino acid composition of a protein is heavily influenced by the A+T percentage of the organism's genome and that a correlation between mean nucleotide content and amino acid content can, in the case of 22 tested sequences be plotted in a linear fashion (**Figure 8.8**). The *P.falciparum*, *T.gondii* and *B.napus* sequences all fall upon the line when considering the observed distribution correlating an organisms genomic A+T ratio and the frequency of amino acids which are encoded for by G+C rich (Gly, Ala, Arg and Pro) and A+T rich (Phe, Tyr, Met, Ile, Asn and Lys) codons. For the PfENR sequence the inserted loop, due to its high asparagine and lysine composition is responsible for an increase in A+T rich codons of approximately 4%.

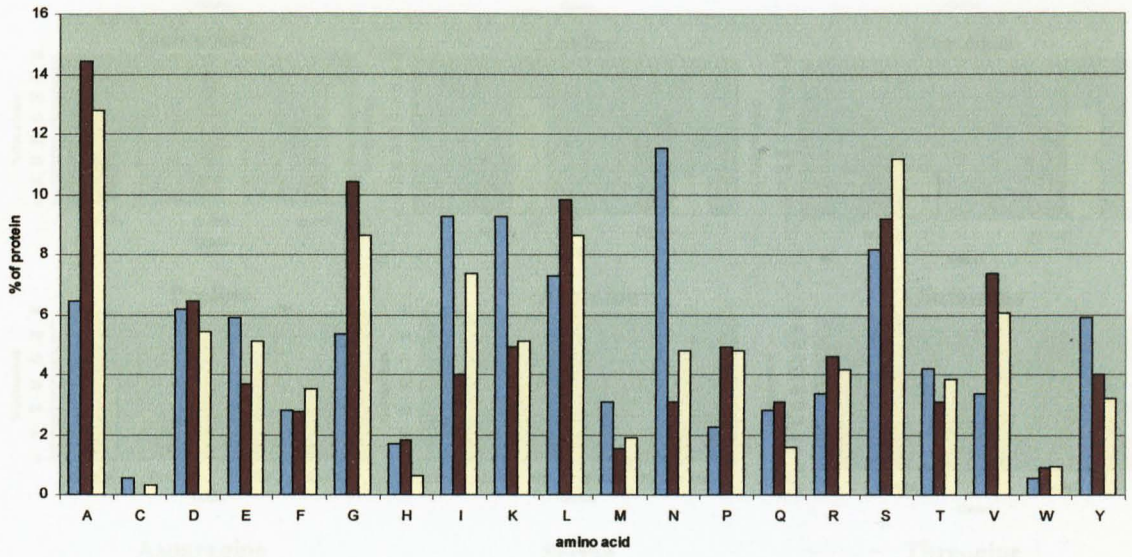


Figure 8.6. The amino acid composition of the mature ENR enzymes from *P.falciparum*, *T.gondii* and *B.napus* coloured in blue, red and yellow, respectively. The amino acids are displayed as one letter codes along the x axis with the overall protein percentage of each residue along the y-axis.

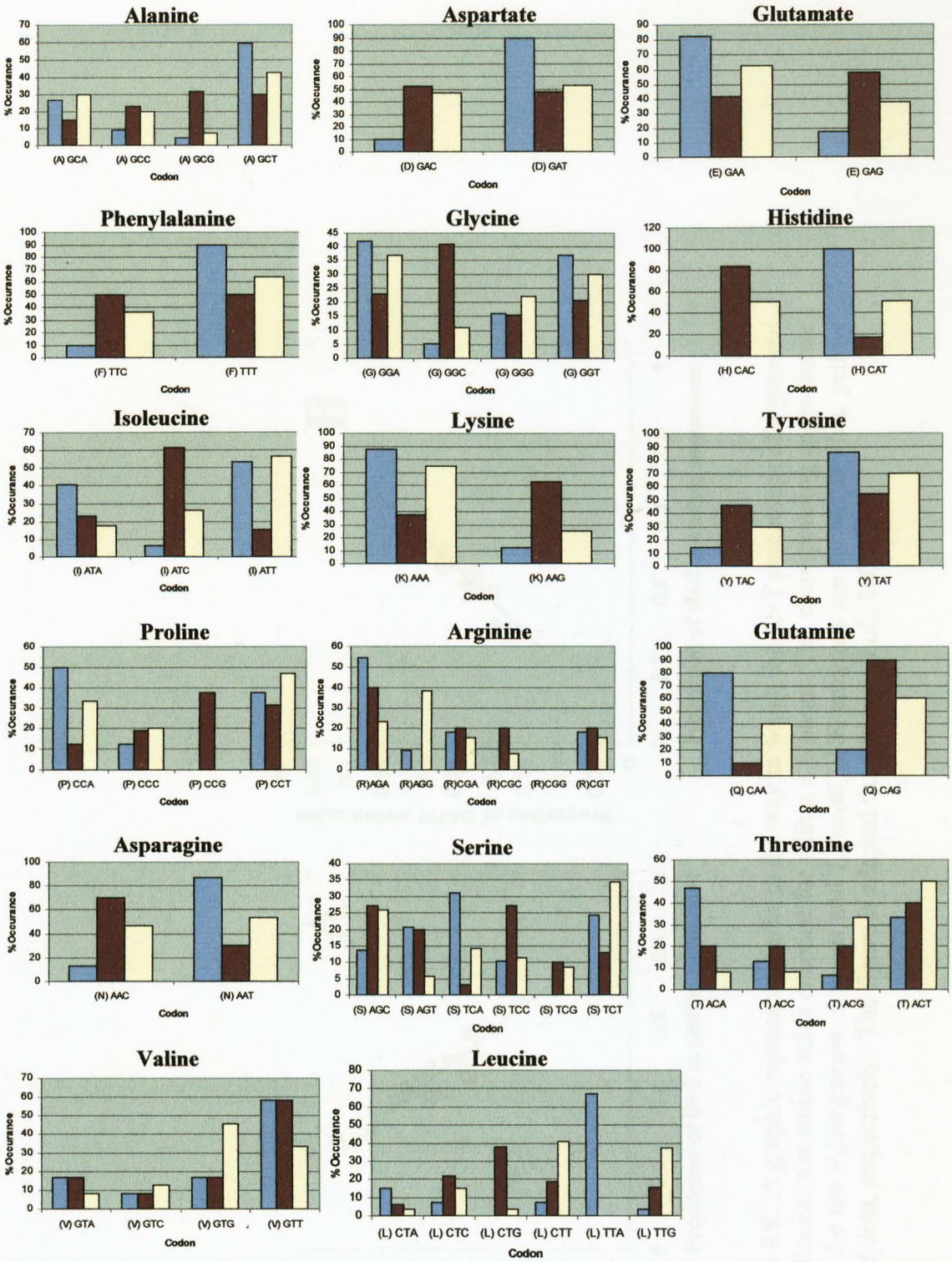


Figure 8.7. The codon frequency within the ENR enzyme gene for the *P.falciparum*, *T.gondii* and *B.napus* ENR enzyme coloured blue, red and yellow, respectively. The amino acid which each graph represents is shown above and the codon used is shown along the x-axis with the percentage occurrence along the y-axis.

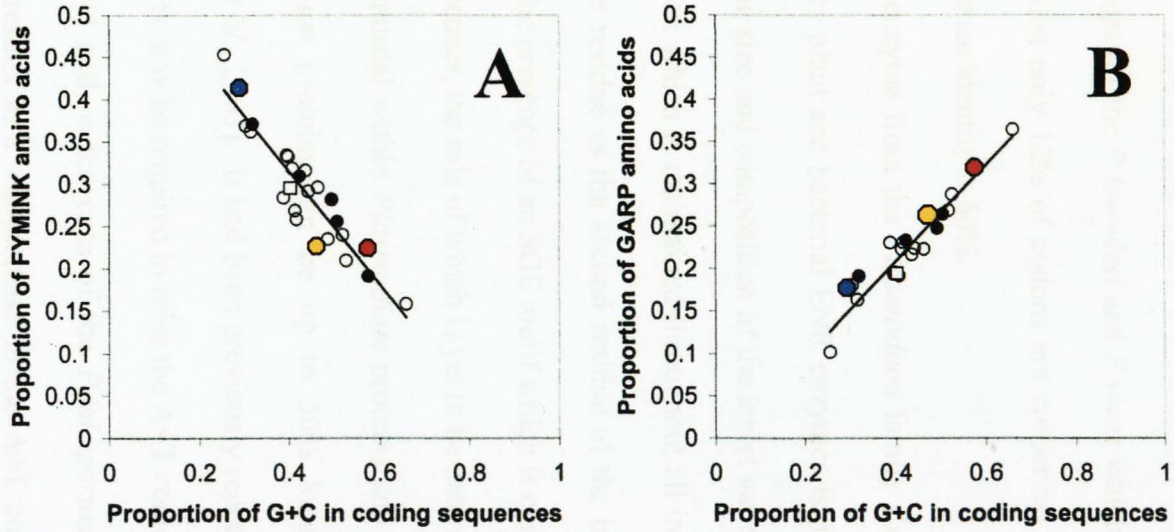


Figure 8.8. A graphic representation as to the way in which the A+T percentage of a genome can influence the amino acids encoded for within a protein. The corresponding approximate values for the *P.falciparum*, *B.napus* and *T.gondii* ENR enzymes are shown as blue, yellow and red dots, respectively. The figure is adapted from Singer *et al.*, 2000.

8.4 Comparisons between the *Plasmodium* ENR family

As a continuation of the study of the ENR family the DNA and amino acid sequences were aligned for the ENR enzymes of all the currently sequenced *Plasmodium* species (**Figure 8.9**). *P.falciparum* and *P.vivax* are responsible for human malaria, *P.berghei*, *P.Chabaudi* and *P.yoelii* are responsible for rodent malaria and *P.Knowlesi* is responsible for malaria within monkeys. The A+T content of these ENR genes is approximately 72% except for the *P.knowlesi* and *P.vivax* which are lower at 57 and 52%, respectively. Whilst only 12% of codons are conserved between species the overall amino acid sequence identity is 60%.

Analysis of the ENR enzyme from the *Plasmodium* family shows that a sequence insert with respect to the plant and bacterial ENR enzymes is maintained across the family, however both the size and composition of the insert varies (**Figure 8.9**). One characteristic of the insert which is maintained throughout all insert sequences is the conservation of a lysine residue as the second residue of the insert. An additional feature of this insert is the presence of an SGE motif which is conserved in all but one of the *Plasmodium* sequences, the role of which is yet to be determined. The presence of such inserts is not unusual within *Plasmodium* proteins and a comparative study showed that *Plasmodium* proteins can be up to 50% longer than their yeast orthologues (Aravind *et al.*, 2003). It had been previously reported that the presence of low complexity inserts may be required to raise the A+T content of a specific gene (Pizzi & Frontali, 2001) and in the case of the *P.falciparum* ENR sequence the presence of the insert results in a 4% increase in the A+T content. However the *Plasmodium chabaudi*, *yoelii* and *beghei* which all have approximately the same ENR gene A+T content as *P.falciparum* contain a much smaller insert. The role for these low complexity inserts have in some cases been assigned to a catalytic or important

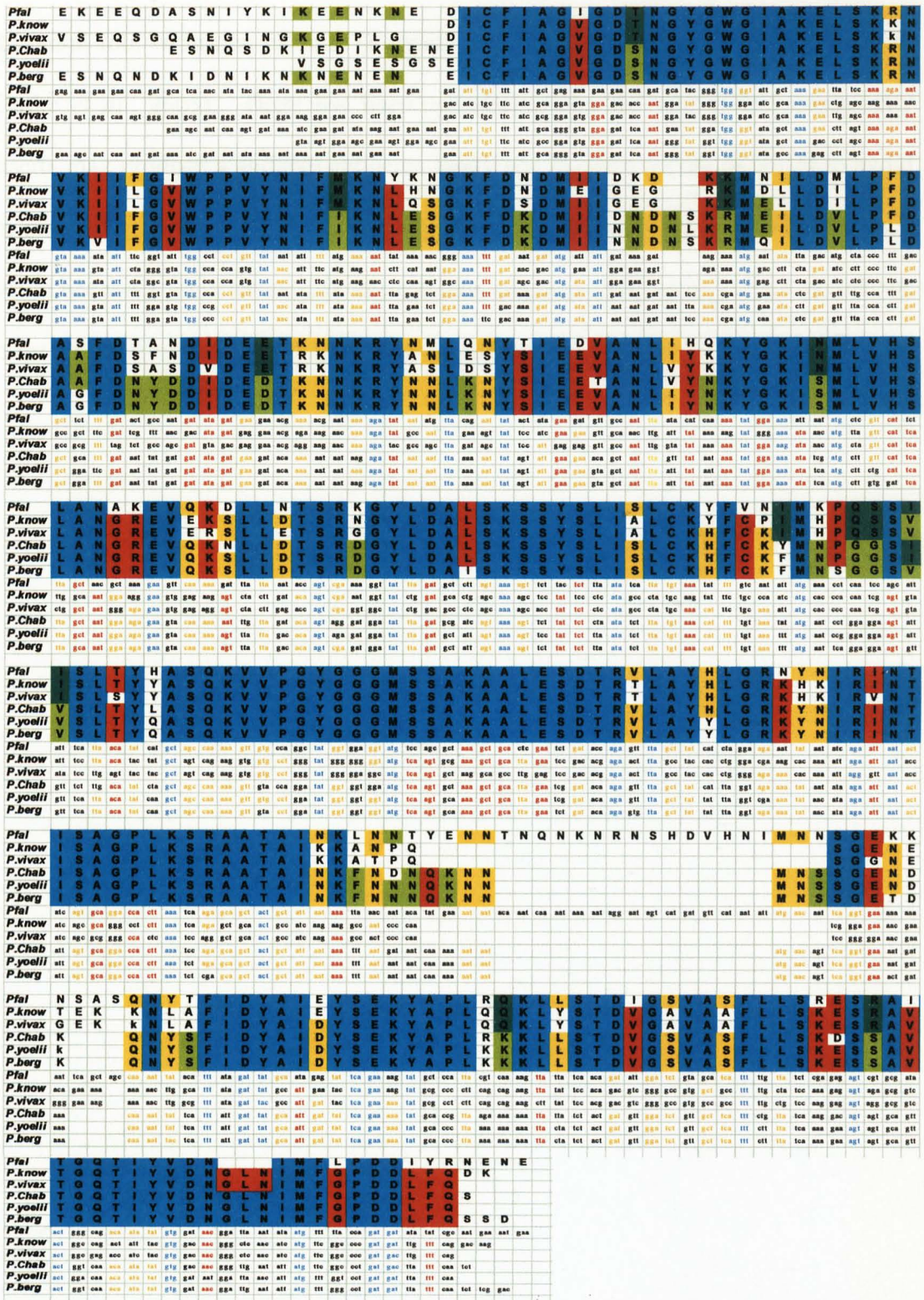


Figure 8.9. A sequence alignment for both the amino acid and gene sequences for the *Plasmodium: falciparum*, *knowlesi, vivax, chabaudi, yoelii* and *berghei* ENR enzyme sequences. Those residues which are fully conserved are highlighted by a blue box, those which are conserved between five sequences, four sequences and three sequences are highlighted by a red, yellow and green box, respectively. Those codons which are conserved between all sequences, five sequences and four sequences are coloured in blue, red and yellow, respectively.

structural role within an enzyme (Kappes *et al.*, 1995, Yuvaniyama *et al.*, 2003 & Clarke *et al.*, 2003). There has been speculation that the role of low complexity inserts may produce a variable antigenic landscape in order for the parasites to avoid an immune response (Anders, 1986). PfENR is however intracellular and as such would not come into contact with any immune response making the role of the PfENR insert with respect to an immune response unlikely.

CHAPTER 9

Inhibitor binding and modelling studies for the apicomplexan *T.gondii* and *P.falciparum* ENR enzyme

9.1 ENR is the target for the development of novel antiparasitic agents

Sequence analysis and modelling studies (section 3.2) along with the subsequent structure determination of *P.falciparum* and *T.gondii* ENR in complex with triclosan suggested that these enzymes would be strongly inhibited by this compound and that it might act to inhibit parasite growth *in vivo*. Therefore, in collaboration with the group of Prof Rima Mcleod at the Department of Visual Sciences at the University of Chicago, the ability of triclosan to inhibit ENR and to retard the growth of these parasites was tested. The results from these inhibition assays showed that triclosan acts as an ENR inhibitor and arrests the growth of *P.falciparum* and *T.gondii* in culture with an IC_{50} of 150ng/ml and 62ng/ml, respectively (McLeod *et al.*, 2001, **Figure 9.1**). Subsequently this finding has been reported for both TgENR and PfENR by several groups (Kapoor *et al.*, 2001 & Surolia *et al.*, 2001).

9.2 Drug delivery within apicomplexan parasites

Although triclosan has been shown to be a good inhibitor of the apicomplexan *T.gondii* and *P.falciparum* ENR enzymes consideration in the development of new novel antiparasitic drugs must be given to the problems in delivering the drugs to their appropriate target. In targeting pathways unique to the apicoplast the drug must not only gain entry into host cells, but cross the parasitophorous vacuole and the four membranes surrounding the apicoplast organelle. This is made more difficult for the *T.gondii* bradyzoites which reside in cysts composed of both host and parasite constituents.

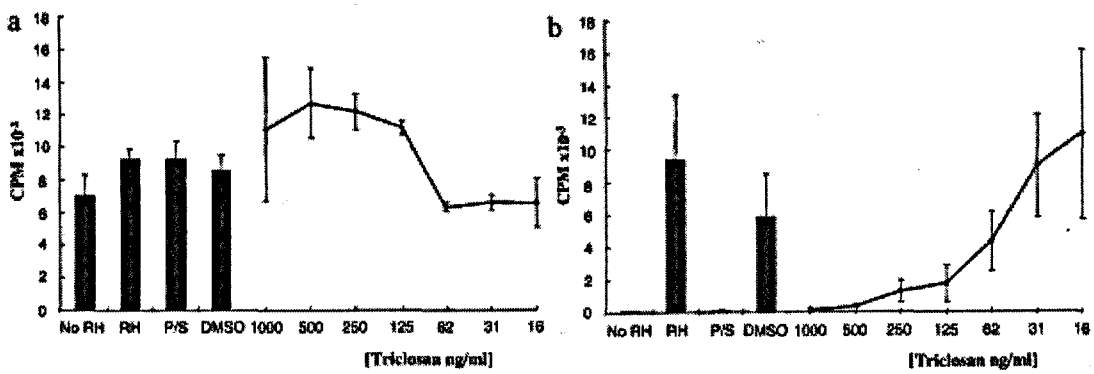


Figure 9.1. Inhibition of the apicomplexan parasite *T.gondii* by triclosan. (a) The host cells show little change in thymidine uptake after the addition of triclosan. (b) The uptake of uracil by the *T.gondii* parasite is reduced after the addition of triclosan, 4 days after the initial infection. The IC_{50} was approximately 62ng/ml. The effect of triclosan was increased over time such that after 125 ng/ml was added to a culture of *T.gondii*, percentage inhibition on day1 was 20% and 72% on day 4. For pyrimethamine/sulfadiazine these values for day 1 and 4 were 63% and 100% respectively.

Abbreviations: Rh, RH strain of *T.gondii* within fibroblasts; No RH is control with fibroblasts alone; DMSO is the fibroblast cells alone but with the highest DMSO concentration used for triclosan administration; P/S, a positive control of fibroblast cells with *T.gondii* and pyrimethamine and sulfadizine; CPM, counts per minute.

Inhibition levels were determined by measuring the the CPM (counts per minute) of radio labelled thymidine by host cells and uracil by the *T.gondii* parasite. Inhibition of either host cells or parasites would result in a decrease in radio labelled thymidine or uracil uptake and as such result in a decrease in counts per minute. The figure is taken from McLeod *et al.*, 2001.

9.2.1 Poly arginine tails

Studies on the nuclear transcription activator protein (Tat) from the HIV type 1 virus showed it was able to cross the plasma membranes of cells in an energy dependent fashion. Of particular interest was a region of highly basic amino acids (RKKRRQRRR) which assisted transport across membranes (Wender *et al.*, 2000). Furthermore it has been shown that polymers of six or more arginine amino acids were more efficient at crossing membranes than polymers of histidine, lysine or ornithine. Cellular uptake of these polymers was not inhibited by low temperatures but was inhibited by preincubation of the cells with sodium azide, indicative of an energy dependent transport system (Mitchell. 2000). Analysis of the mode of triclosan binding in TgENR suggested that attachment of a linker arginine via the phenolic OH would prevent inhibition of the enzyme when the linker is attached but should the linker be removed then triclosan would be free to bind. The triclosan inhibitor was attached to the arginine linker by the hydroxyl group forming a hydrolysable ester linkage or a non releasable ether linkage (**Figure 9.2d, e**). Furthermore studies carried out by the collaborating laboratory of Prof Rima McLeod showed that the addition of a releasable eight residue arginine linker to the antimicrobial inhibitor triclosan (Tr8) improved the speed of delivery and enabled triclosan to enter both the extracellular and intracellular *T.gondii* tachyzoites and bradyzoites (Samuel *et al.*, 2003, **Figure 9.2a, b & c**). One limiting factor in the use of several ENR inhibitors is that their ability to mimic the lipophilic substrate makes them inherently hydrophobic. This is true with triclosan which must be dissolved in a solvent such as DMSO before addition to samples, severely limiting its medical applications. The addition of a releasable arginine linker not only allows for the

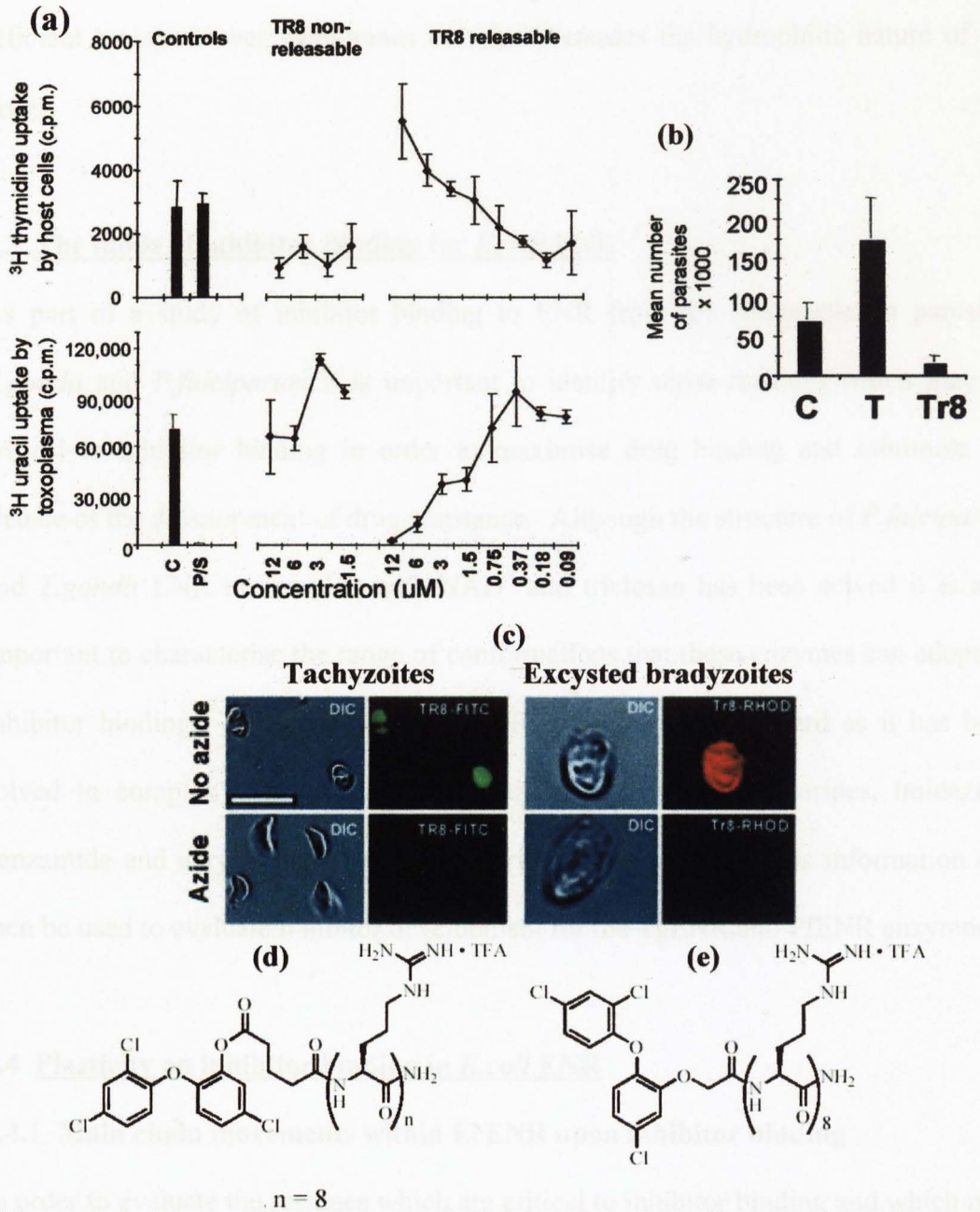


Figure 9.2. (a) An R8 conjugate of triclosan linked in either a releasable or non releasable manner having little or no toxic effect on host cell replicataion as measured by ^{3}H thymidine uptake. The releasable triclosan does however have an effect on the ^{3}H uracil uptake by toxoplasma cells. For both graphs the control C is *T.gondii*-infected host cells in medium and P/S represents *T.gondii* host cells treated with 0.1μg/ml pyrimethamine and sulfadiazine 25μg/ml which inhibits parasite growth. **(b)** The administration of releasable r8-triclosan (Tr8) and triclosan (T) "in vivo" to *T.gondii* tachyzoites. The releasable r8-triclosan but not triclosan alone reduces parasite numbers significantly 5 days after challenge. **(c)** Fluorescein and DIC microscopy images showing the uptake of labelled Tr8 by *T. gondii* tachyzoites and excysted bradyzoites and the inhibition of uptake of arginine triclosan conjugates by azide indicative of an energy dependant transport system (Scale bar = 10 μm.). **(d)** and **(e)** Representation of triclosan bound to an eight residue arginine linker by a hydrolysable ester linkage **(d)** and by a non releasable ether bond **(e)**.

efficient transport over membranes but also increases the hydrophilic nature of the drug.

9.3 The mode of inhibitor binding for *E.coli* ENR

As part of a study of inhibitor binding to ENR from the apicomplexan parasites *T.gondii* and *P.falciparum* it is important to identify those residues which may be critical to inhibitor binding in order to maximise drug binding and minimise the chance of the development of drug resistance. Although the structure of *P.falciparum* and *T.gondii* ENR in complex with NAD^+ and triclosan has been solved it is also important to characterise the range of conformations that these enzymes can adopt on inhibitor binding. The structure of EcENR is useful in this regard as it has been solved in complex with a range of inhibitors (triclosan, diazaborines, imidazole, benzamide and acrylamide based aminopyridine derivatives). This information can then be used to evaluate inhibitor development for the TgENR and PfENR enzymes.

9.4 Plasticity on inhibitor binding in *E.coli* ENR

9.4.1 Main chain movements within EcENR upon inhibitor binding

In order to evaluate the residues which are critical to inhibitor binding and which may adopt different conformations in EcENR, all of the deposited pdb files that were derived from the analysis of crystals formed by co-crystallisation of the ENR enzymes with inhibitors were compared. Restricting the study to these structures should ensure that any conformational change which may occur on inhibitor binding would be allowed to develop fully. Of the EcENR inhibitor complexes only those of the diazaborines, triclosan and compound 29 were solved through co-crystallisation (Baldock *et al.*, 1996, Levy *et al.*, 1999 & Seefeld *et al.*, 2002). In order to investigate

significant main chain movements within EcENR upon inhibitor binding the program Escet (Schneider, 2002) was used to compare the position of the Cα atoms between the EcENR/NAD^+ complex and any co-crystal inhibitor complex (**Figure 9.3**). Comparisons between the EcENR/NAD^+ complex and the various EcENR inhibitor complexes showed three significant regions of difference between the various complexes (41 to 57, 190 to 223 and 249 to 259). Of these regions of flexibility that between residues 190 to 223, which corresponds to helix α6 and its flanking residues shows the greatest movement. Comparison of the region 190 to 223 within the various complexes of EcENR shows it to adopt a variety of conformations depending on the nature of the inhibitor bound. In the structure of EcENR/NAD^+ this region forms a disordered loop with no interpretable density for residues 196 to 205, however on binding of the diazaborine family of drugs this region adopts either an ordered loop (Baldock *et al.*, 1996) or an helical fold (Levy *et al.*, 2001) as seen for either triclosan or compound 29 (**Figure 9.4**). Furthermore analysis of the position of the helix formed in the triclosan and compound 29 complexes shows it to adopt a different position in each complex in order to maximise the van der Waals interactions formed between the ENR enzyme and its bound inhibitor (**Figure 9.4** & **Figure 9.5**). This observation is consistent with those previously reported (Roujeinikova[b] *et al.*, 1999 & Qiu[a] *et al.*, 1999).

9.4.2 Sidechain movements within *E.coli* ENR upon inhibitor binding

Analysis of EcENR in complex with triclosan, thieno-diazaborine and compound 29 showed that significant changes in the conformation of the sidechains was restricted to residues 190-223 which form part of the protein which displayed the greatest flexibility in inhibitor binding (**Figure 9.6**). In EcENR inhibitors are bound within a

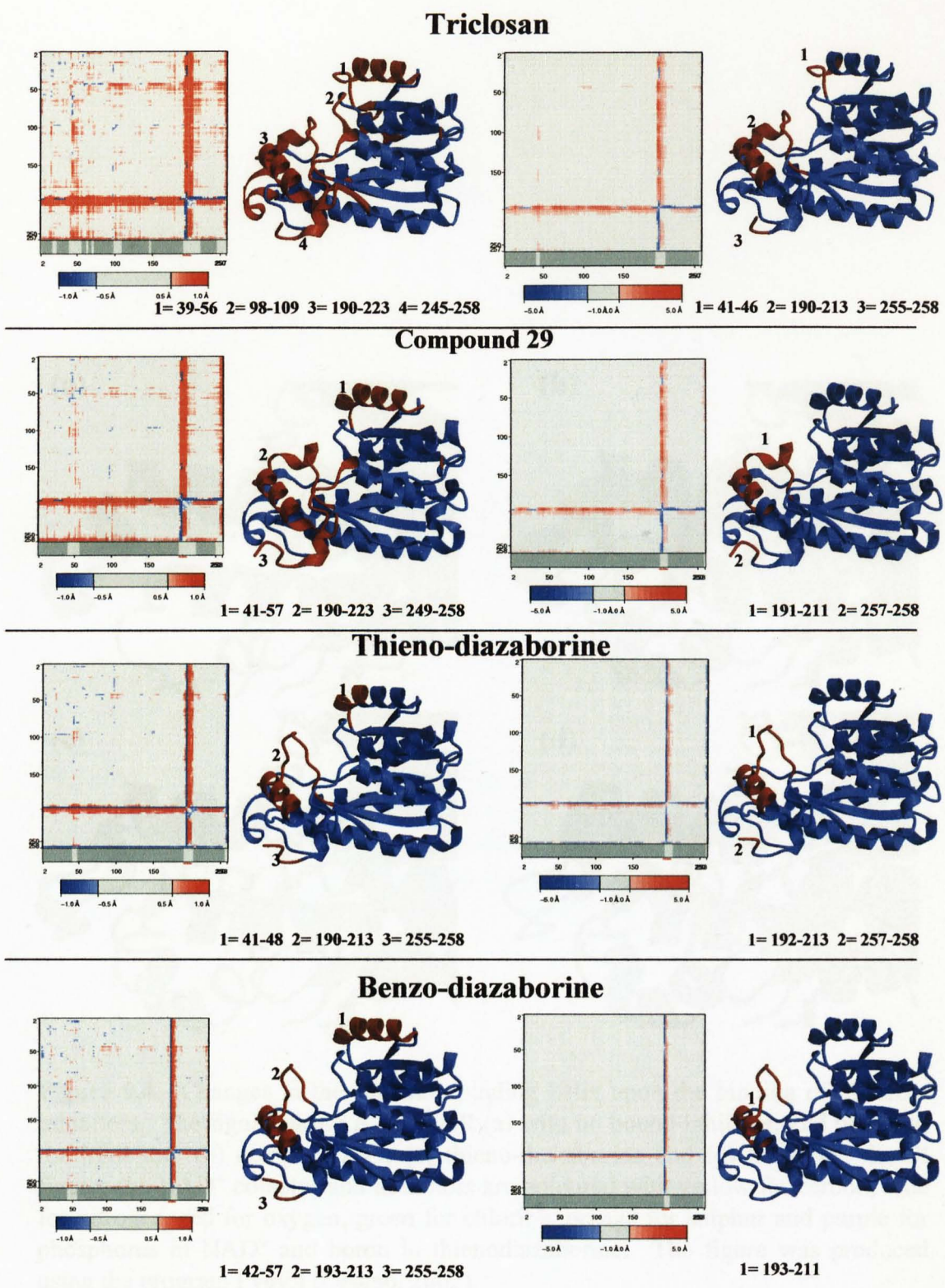


Figure 9.3. Analysis of mainchain movement as a result of inhibitor binding between the EcENR/NAD$^+$ complex and the EcENR/NAD$^+$/inhibitor complexes with (a) triclosan (b) compound 29 (c) thieno diazaborine and (d) benzo diazaborine using the program Escet (Schneider, 2002). For each complex a difference matrix shows movements within the range of 0.5 – 1.0Å (left) and 1.0 –5.0Å (right). The EcENR structure is displayed at the right of the Escet plot and coloured red for areas that undergo conformational change upon inhibitor binding.

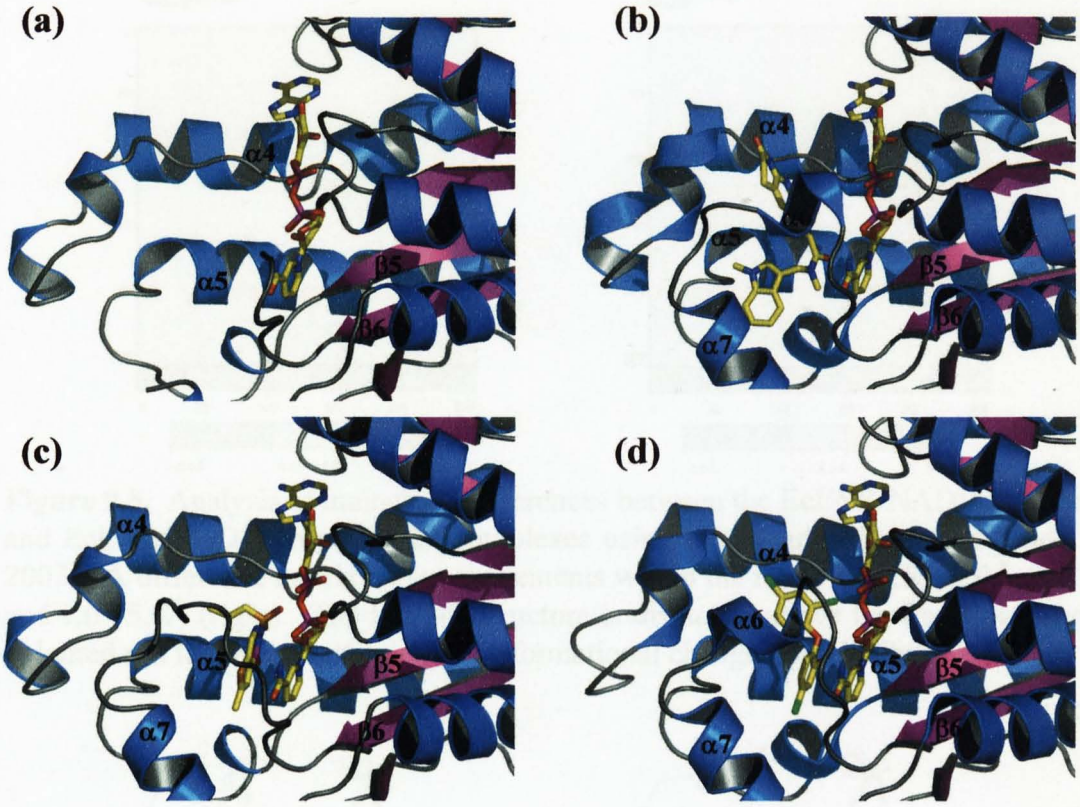


Figure 9.4. Changes in the substrate binding helix upon the binding of different inhibitors. The figure shows *E.coli* ENR (a) with no bound inhibitor, and bound to the inhibitors (b) compound 29, (c) thieno-diazaborine and (d) triclosan. In all figures the NAD^+ cofactor and inhibitors are coloured with yellow for carbon, blue for nitrogen, red for oxygen, green for chlorine, orange for sulphur and purple for phosphorus in NAD^+ and boron in thienodiazaborine. The figure was produced using the program Pymol (Delano, 2002).

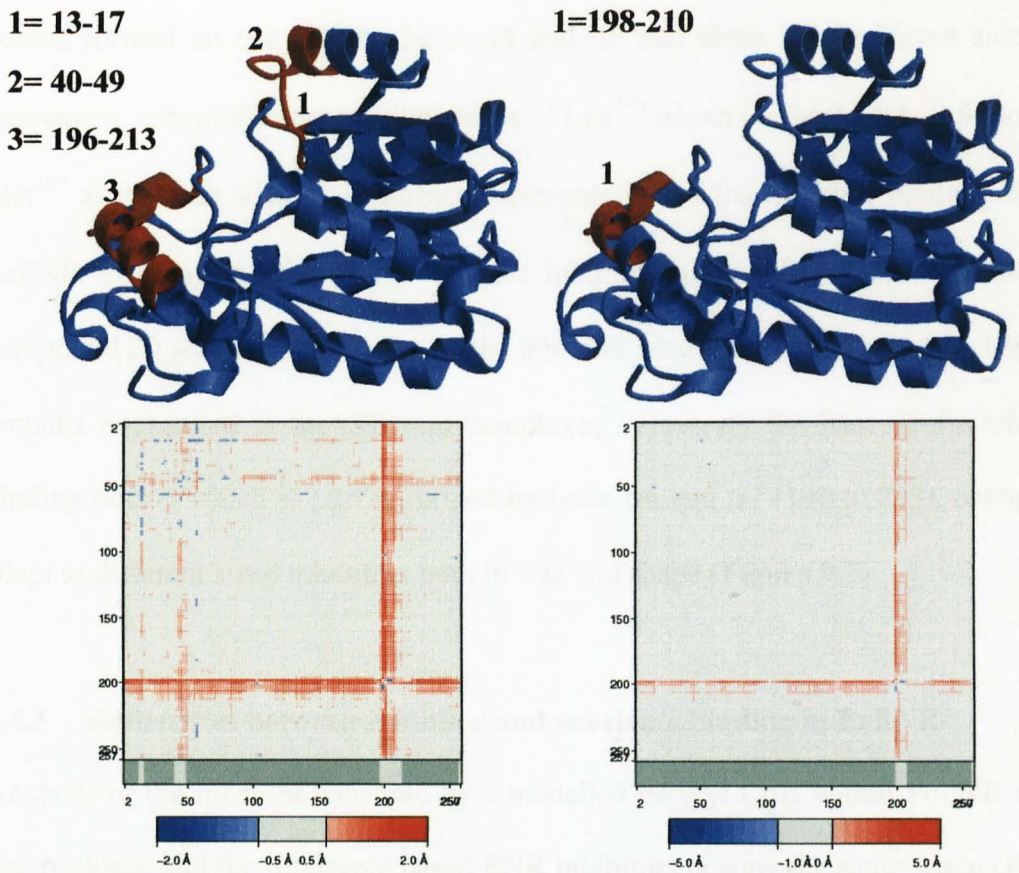


Figure 9.5. Analysis of mainchain differences between the EcENR/NAD$^+$/triclosan and EcENR/NAD$^+$/compound 29 complexes using the program Escet (Schneider, 2002). A difference matrix shows movements within the range of 0.5 – 2.0Å (left) and 1.0 –5.0Å (right). The EcENR structure is displayed above the Escet plot and coloured red for areas that undergo conformational change upon inhibitor binding.

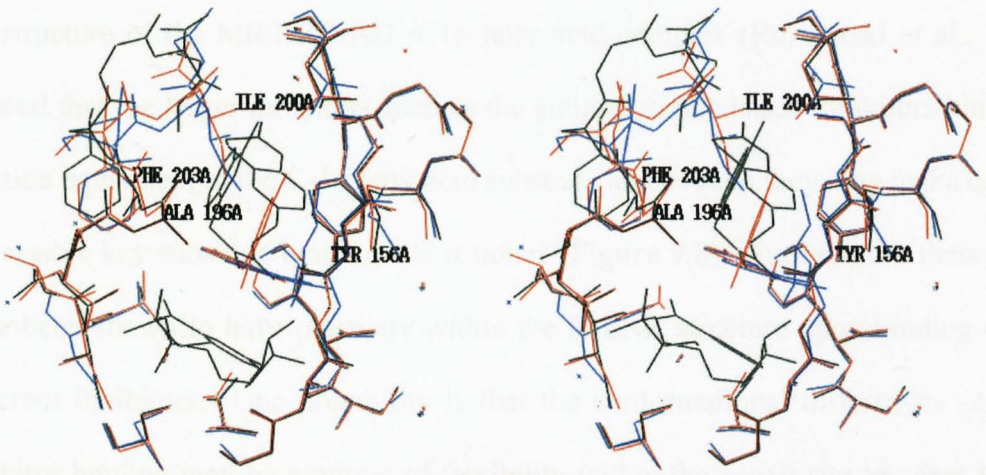


Figure 9.6. A stereo view of the EcENR active site in complex with triclosan (red), compound 29 (black) and thieno diazaborine (blue). For simplicity the NAD$^+$ cofactor, which occupies the same position in all complexes is only shown for the compound 29 complex and is coloured black. The figure was produced in TURBO/FRODO (Roussel *et al.*, 1990).

pocket formed on one side by helix α4 and α5 and show no significant sidechain movement, except for very small changes of Leu^{100} which lies in the β4 - α4 loop and Met^{159} at the start of α5. A surface representation of the inhibitor binding site for EcENR in complex with NAD^+ and the inhibitors thienodizaborine, triclosan and compound 29 showed that the top of the inhibitor binding pocket formed by helix α4 remains unchanged in the different complexes. However the base of the inhibitor binding pocket which in part is formed by helix α6 and α7 (190 to 223), changes to adapt to different sized inhibitors both in size and shape (**Figure 9.7**).

9.4.3 Similarities between inhibitor and substrate binding in EcENR

Analysis of the mode of substrate by a modelled butyryl-CoA within EcENR shows that triclosan and the imidazole based ENR inhibitors occupy the same space (**Figure 9.8**). However the other inhibitors occupy additional space at the top of the substrate binding pocket formed by helix α6 and α7 in EcENR (**Figure 9.8**). Interestingly, modelling of the larger C-16 fatty acid in EcENR based on the observed position in the structure of the MtENR/NAD^+/C16 fatty acid complex (Rozwarski et al., 1999) showed that the larger inhibitors such as the aminopyridine based inhibitors bind in a position equivalent to the C-16 fatty acid substrate and as such they may make contact to possible key substrate binding site residues (**Figure 9.8**). Furthermore these areas have been shown to have plasticity within the EcENR structure upon binding of the different inhibitors. One possibility is that the conformational differences seen on inhibitor binding may be a mimic of flexibility within the active site to adapt to and handle different length substrates. However analysis of the structure of MtENR/NAD^+ reveals that the substrate binding helix (196 to 219), which is at an equivalent position as α6 in EcENR is formed in the absence of an inhibitor or

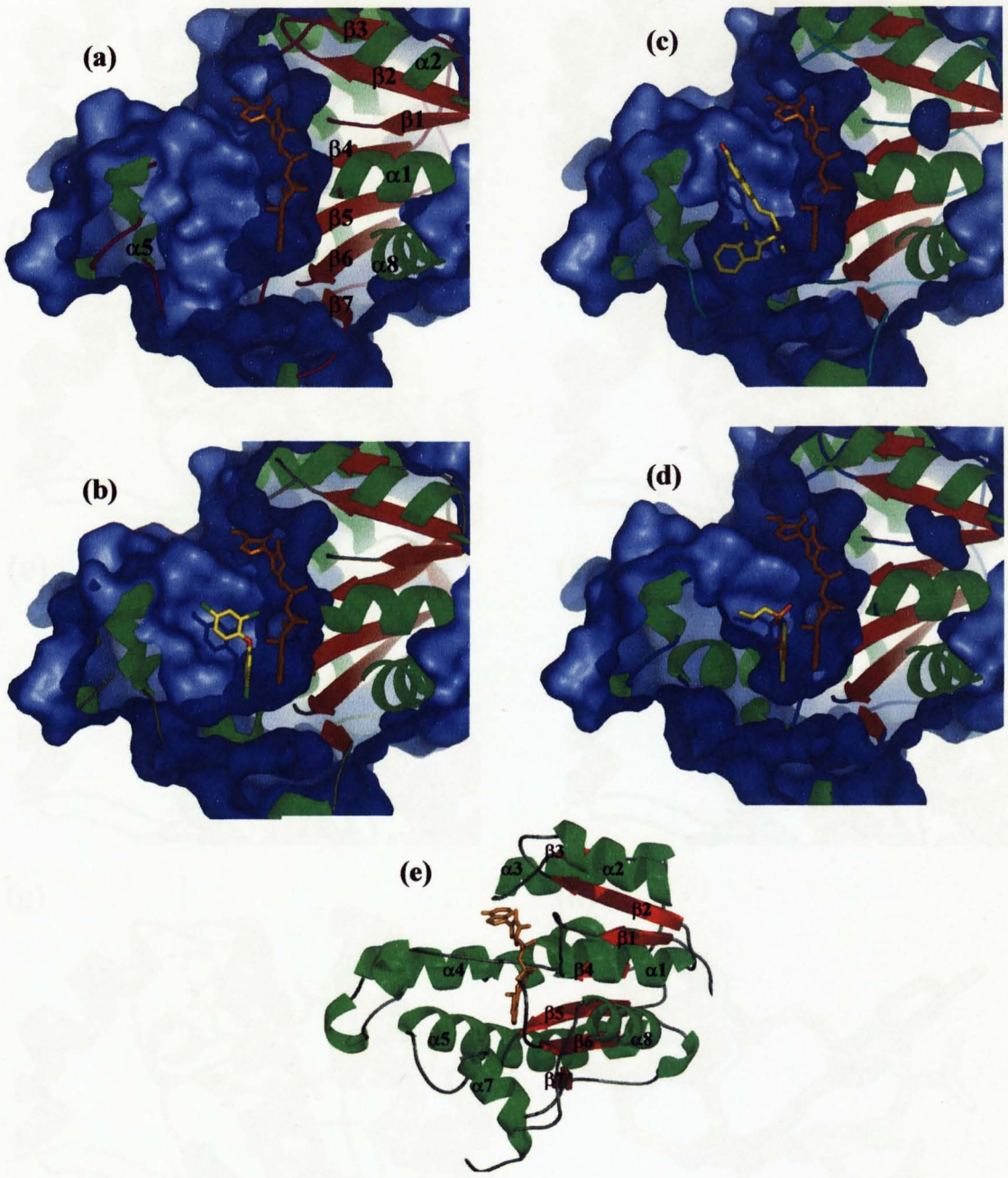


Figure 9.7. A surface representation of the *E.coli* ENR active with (a) no inhibitor bound, (b) compound 29 bound, (c) triclosan bound and (d) thieno-diazaborine bound. All inhibitors are coloured with yellow for carbon, blue for nitrogen, red for oxygen, green for chlorine and orange for boron and the NAD^+ cofactor is shown in orange. The secondary structural elements are labelled where visible for the EcENR/NAD^+ complex. (e) The EcENR/NAD^+ complex in an equivalent view as figure a, b, c and d but without a surface representation. All of the secondary elements are numbered except for α6 which is not present within the EcENR/NAD^+ complex but is present in many EcENR inhibitor complexes. The figure was produced using the program Pymol (Delano, 2002).

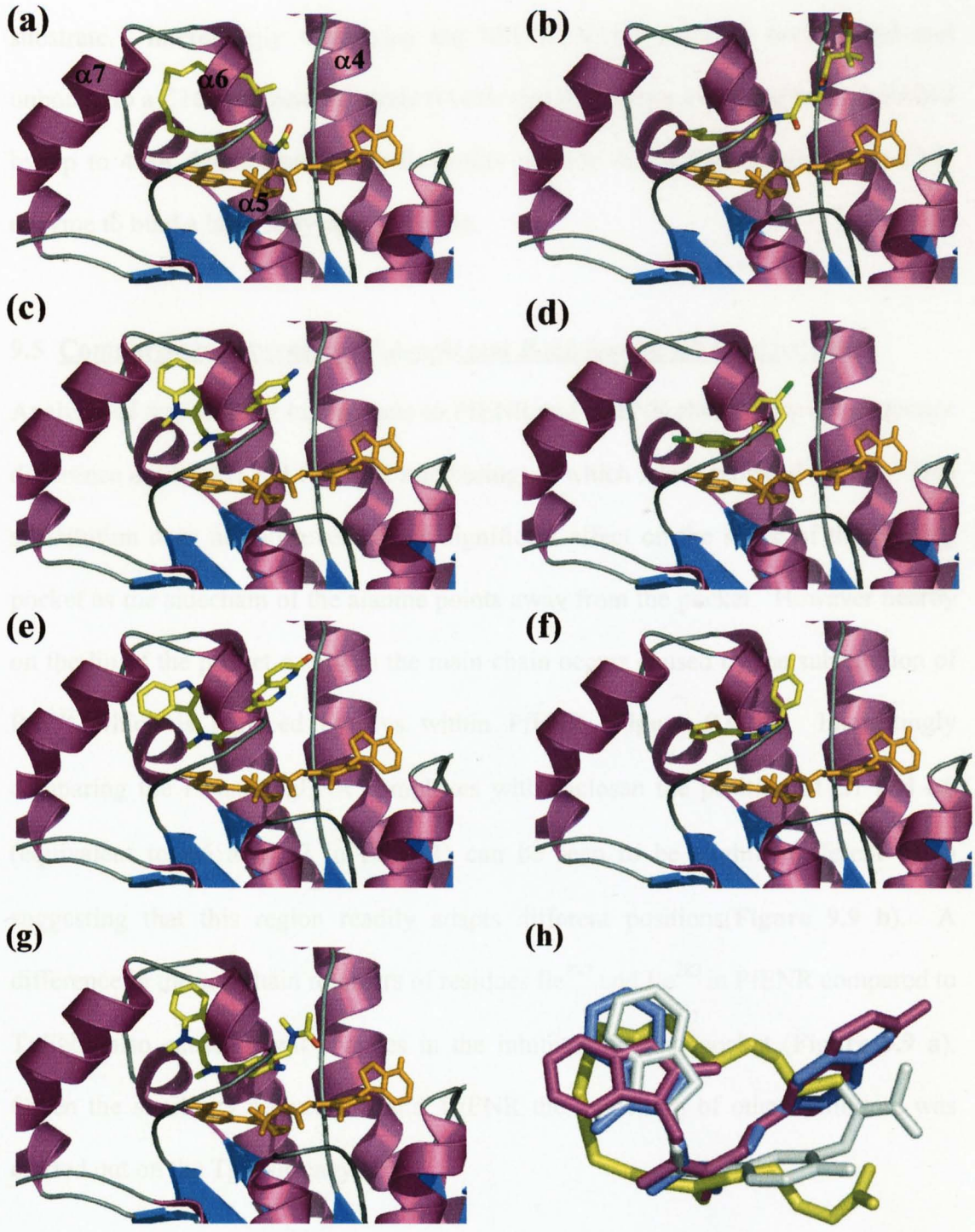


Figure 9.8. The EcENR/NAD^+/triclosan complex with various inhibitors modelled into the active site. (a) C-16 fatty acid substrate, (b) butyryl substrate, (c) acrylamide based inhibitor, (d) triclosan, (e) compound 29, (f) imidazole based inhibitor and (g) benzamide based inhibitor. (h) The overlay onto the C16 fatty acid substrate coloured yellow, with compound 4, benzamide and imidazole based inhibitors coloured in purple, white and blue respectively. EcENR is coloured with magenta for α-helices, blue for β-strands and for the inhibitors and NAD^+ cofactor yellow for carbon, red for oxygen, blue for nitrogen, purple for phosphorous, orange for sulphur and green for chlorine. The figure was produced using the graphics program Pymol (Delano, 2002).

substrate. Interestingly comparing the MtENR/NAD^+ complex both bound and unbound to a C16 fatty acid substrate reveals that the substrate binding helix is shifted by up to 4.0Å demonstrating the flexibility present within this region of the ENR enzyme to bind a large fatty acid substrate.

9.5 Comparisons between the *T.gondii* and *P.falciparum* ENR active sites

Analysis of the inhibitor binding site in PfENR and TgENR shows only one sequence difference at position 131 (TgENR numbering) in which Ala is replaced by Gly. This substitution does not however have a significant effect on the shape of the binding pocket as the sidechain of the alanine points away from the pocket. However nearby on the lip of the pocket a shift in the main chain occurs caused by the substitution of Pro^{132} which is replaced by Lys within PfENR (**Figure 9.9 a**). Interestingly comparing the Pf and TgENR complexes with triclosan the position of α7 and α8 (equivalent to α6 and α7 in EcENR) can be seen to be slightly different again suggesting that this region readily adapts different positions(**Figure 9.9 b**). A difference in the sidechain rotamers of residues Ile^{237} and Ile^{283} in PfENR compared to TgENR also causes slight changes in the inhibitor binding pocket (**Figure 9.9 a**). Given the similarity between Pf and TgENR the modelling of other inhibitors was carried out on the TgENR enzyme.

9.6 Inhibitor modelling within *T.gondii* ENR

To examine how the various inhibitors can be accommodated in the TgENR enzyme the structures of the EcENR/NAD^+/inhibitor complexes were superimposed. A dominant feature of inhibitor modelling is that compound 29 and the related benzamide and acrylamide based inhibitors all form steric clashes within the binding

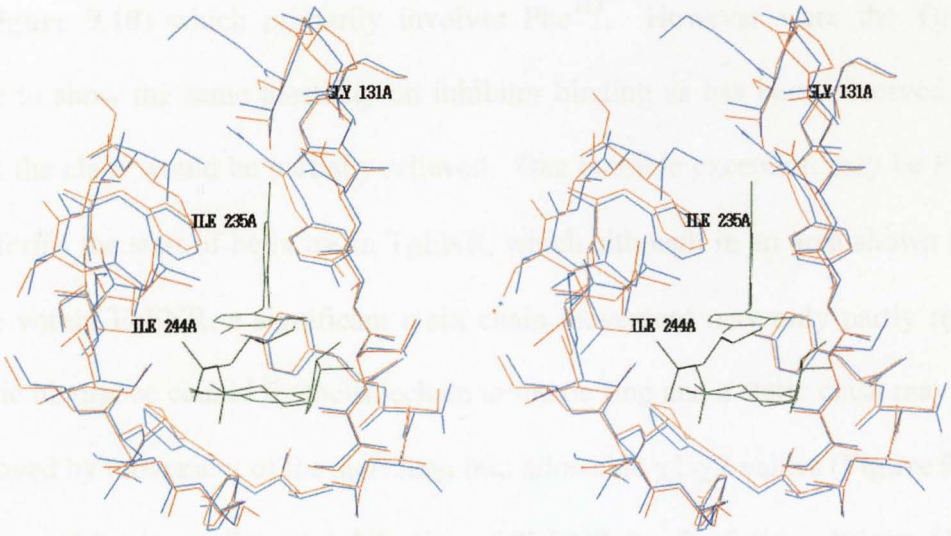


Figure 9.9. a A stereo diagram showing a comparison between the inhibitor binding site of the PfENR/NAD^+/triclosan (blue) and TgENR/NAD^+/triclosan (red) complexes. The NAD^+ and triclosan are coloured with yellow for carbon, blue for nitrogen, red for oxygen, green for chlorine and purple for phosphorous. The figure was produced using the graphics program TURBO/FRODO (Rousell *et al.*,1990).

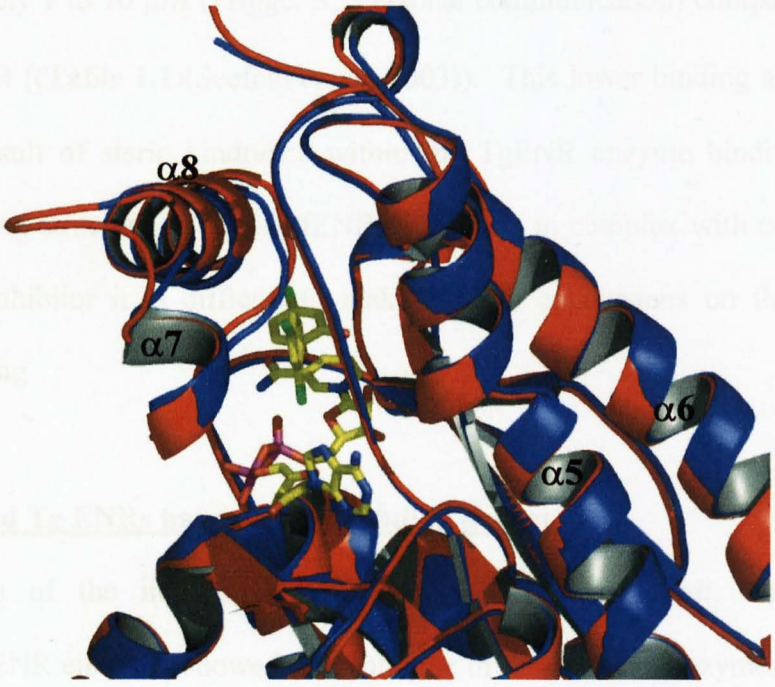


Figure 9.9. b A comparison between the inhibitor binding loop of the PfENR/NAD^+/triclosan (blue) and TgENR/NAD^+/triclosan (red) complexes. The NAD^+ and triclosan are coloured with yellow for carbon, blue for nitrogen, red for oxygen, green for chlorine and purple for phosphorous. The figure was produced using the graphics program Pymol (Delano, 2002).

site (**Figure 9.10**) which primarily involves Phe^{243}. However were the TgENR enzyme to show the same plasticity on inhibitor binding as has been observed with EcENR the clash would be partially relieved. One possible exception may be Phe^{243} which forms the start of helix α8 in TgENR, which although in an area shown to be flexible within EcENR, a significant main chain movement may only partly reduce the steric hindrance caused by the sidechain aromatic ring and a steric clash may only be removed by movement of the sidechain into allowable χ1-χ2 values (**Figure 9.11**). Preliminary kinetic studies of inhibition of TgENR by Prof Sean Prigge (Johns Hopkins Bloomberg School of Public Health) showed compound 4, which is closely related to compound 29 to be an effective inhibitor of TgENR, but with a IC_{50} value of approximately 1 to 10 μM (Prigge. S., Personal communication) compared to 0.06 μM for EcENR ((**Table 1.1**)(Seefeld *et al.*, 2003)). This lower binding affinity may be a direct result of steric hindrance within the TgENR enzyme binding site but without a crystal structure of either PfENR or TgENR in complex with compound 4 or a related inhibitor it is difficult to make reliable conclusions on the mode of inhibitor binding

9.7 The Pf and Tg ENRs have a larger binding pocket

A comparison of the inhibitor binding pockets for the *E.coli*, *T.gondii* and *P.falciparum* ENR enzymes showed a methionine in the bacterial enzymes (Met^{206} in EcENR) is substituted by an alanine residue within the apicomplexan ENR enzymes (**Figure 5.1**). Sequence analysis of the ENR family shows that the bacterial and plant ENR sequences all contain a bulky hydrophobic residue (Met, Leu or Ile) but the ENR enzymes from *T.gondii, Eimeria* and the *Plasmodium* family have a fully conserved alanine residue. This sequence change produces an increase in space at the base of

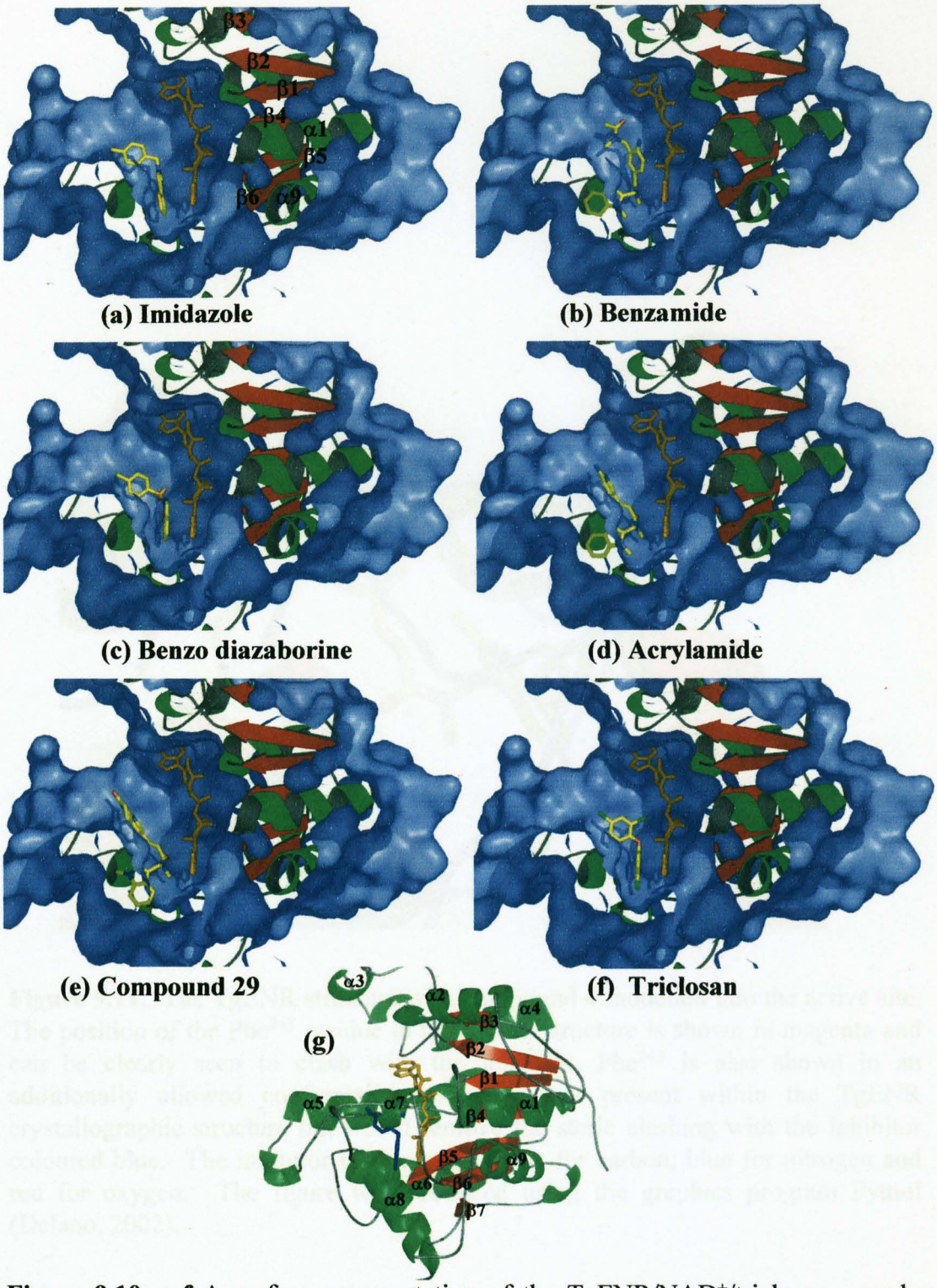


Figure 9.10. a-f A surface representation of the TgENR/NAD^+/triclosan complex active site with various ENR inhibitors modelled in. The active sites surface is represented in blue and all inhibitors are coloured with yellow for carbon, blue for nitrogen, red for oxygen, green for chlorine, orange for sulphur and purple for boron. **(g)** The TgENR/NAD^+ complex in an equivalent view as figure a-f but without a surface representation. All of the secondary elements are numbered. The NAD^+ cofactor is coloured orange. In the modelled complexes the acrylamide and benzamide based inhibitors can be seen to penetrate the surface of the enzyme binding pocket. The figure was produced using the graphics program Pymol (Delano, 2002).

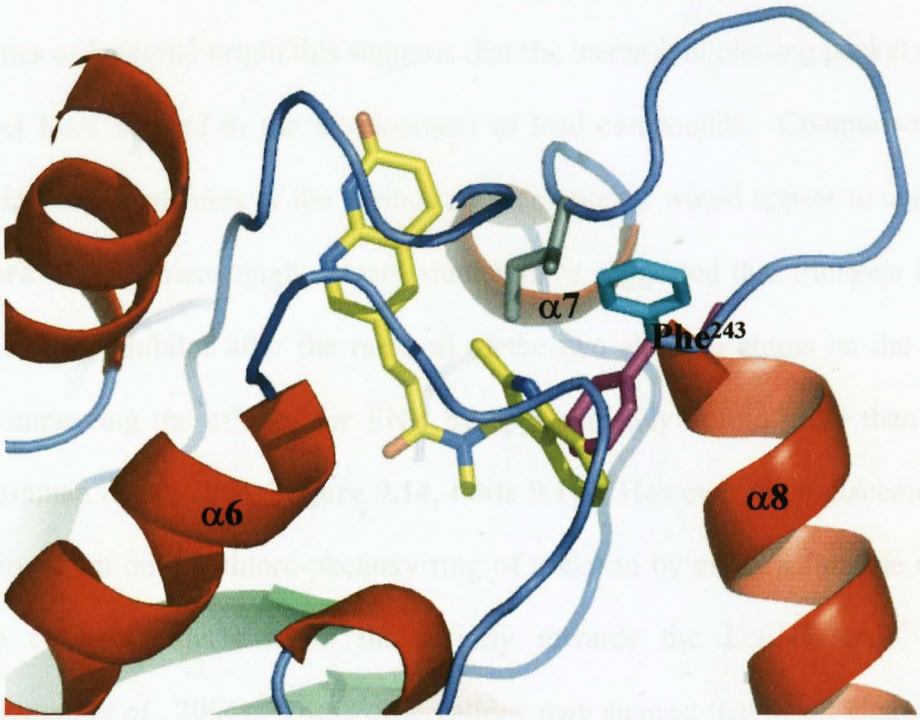


Figure 9.11. The TgENR structure with compound 4 modelled into the active site. The position of the Phe^{243} residue in the solved structure is shown in magenta and can be clearly seen to clash with the inhibitor. Phe^{243} is also shown in an additionally allowed conformation which is not present within the TgENR crystallographic structure but would remove the steric clashing with the inhibitor coloured blue. The inhibitor is coloured yellow for carbon, blue for nitrogen and red for oxygen. The figure was produced using the graphics program Pymol (Delano, 2002).

the binding pocket and reduces the van der Waals packing interactions with the triclosan inhibitor in comparison to EcENR **(Figure 9.12)**. The production of a triclosan analogue which contains an additional atom at this position may allow for the van der Waals contacts to be optimised improving inhibitor binding. As most drug screening programs against the ENR enzyme have currently been based on enzymes of bacterial origin this suggests that the increase in binding pocket space has not yet been utilised in the development of lead compounds. Comparisons of the superimposed structures of the various drug complexes would appear to confirm this **(Figure 9.13)**. Interestingly, recent studies have suggested that triclosan is a more potent ENR inhibitor after the removal of the two chlorine atoms on the di-chloro ring, improving the affinity for ENR by approximately 7 fold more than triclosan ((Sivaraman *et al.*, 2004)**(Figure 9.14, table 9.1)**). However the replacement of the chlorine atom on the chloro-phenoxy ring of triclosan by either a fluorine or methyl group caused a decrease in the affinity towards the EcENR/NAD^+ complex (Sivaraman *et al.*, 2004). These observations may suggest that small changes on the chloro-phenoxy ring of triclosan which introduce subtle changes in the electronic and steric properties may reduce the affinity of triclosan to ENR. Importantly, although there is very strong sequence conservation across the apicomplexan family of an alanine residue at the base of the triclosan binding pocket, this conservation is not maintained across the ENR family and as such an alanine residue is not essential for catalysis or substrate binding and the chance of a mutation causing drug resistance through steric hindrance may be increased.

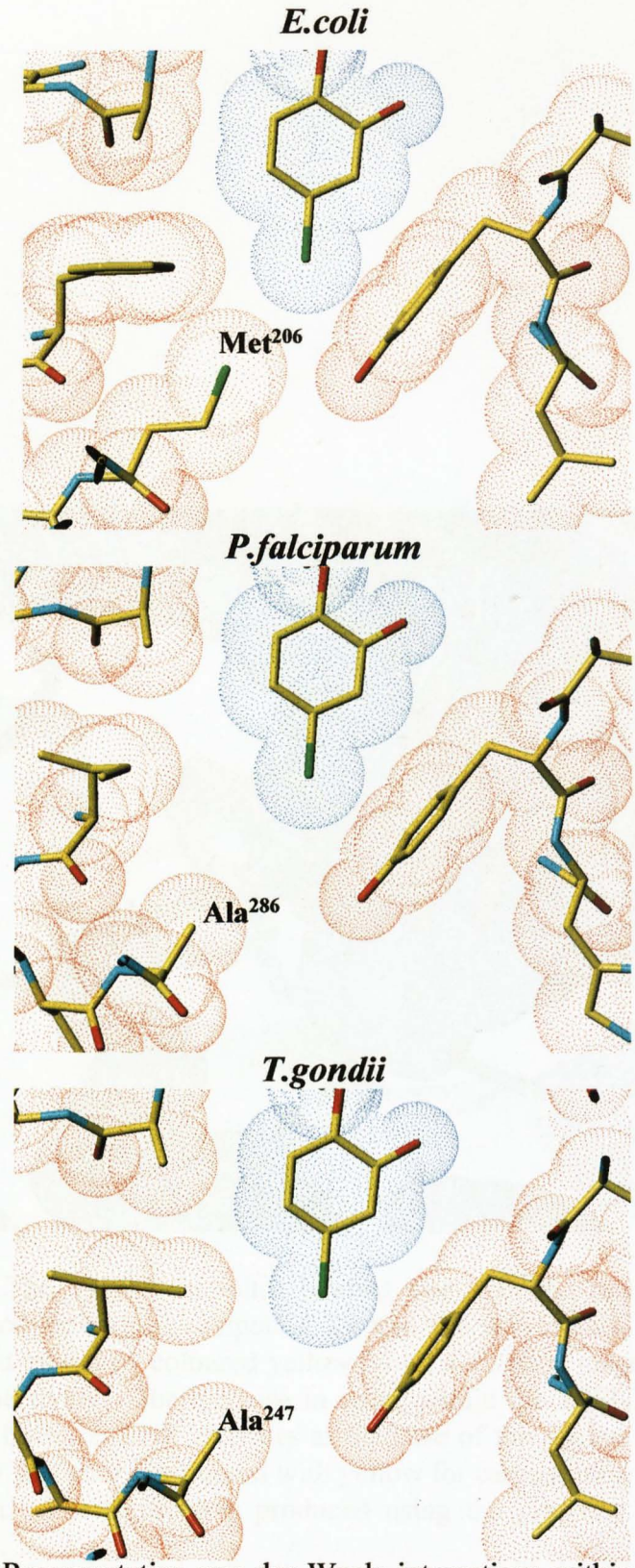


Figure 9.12. Representative van der Waals interactions within the ENR active site of *E.coli*, *P.falciparum* and *T.gondii*. The van der Waals surface of the triclosan inhibitor is shown in blue and the van der Waals surface for the ENR enzyme active site residues is shown in red. All figures are coloured with yellow for carbon, red for oxygen, blue for nitrogen and green for chlorine. The figure was produced in the graphics program TURBO/FRODO (Roussel *et al.*, 1990).

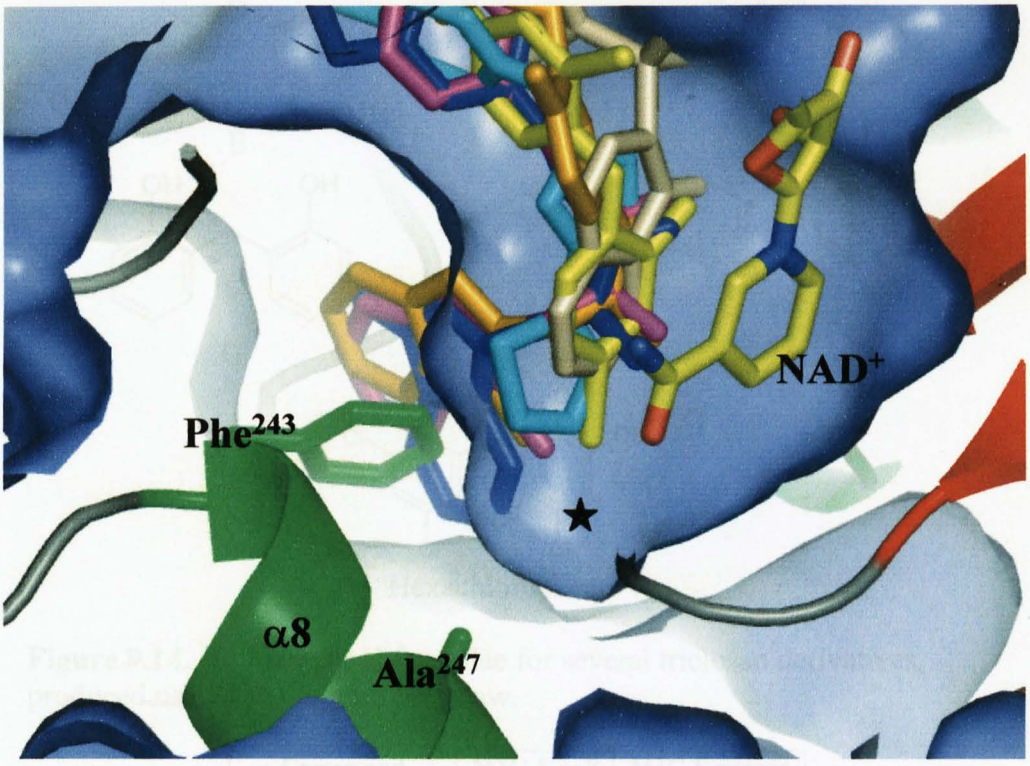


Figure 9.13. The TgENR inhibitor binding site with the modelled inhibitors triclosan, benzodiazaborine, compound 29 and the acrylamide, benzamide and imidazole based inhibitors coloured yellow, gold, dark blue, magenta, orange and light blue, respectively. The increase in space within the TgENR binding sitein contrast to the bacterial ENR enzymes at the base of the pocket is shown with a star. The NAD^+ cofactor is coloured with yellow for carbon, blue for nitrogen and red for oxygen. The figure was produced using the graphics program Pymol (Delano, 2002).

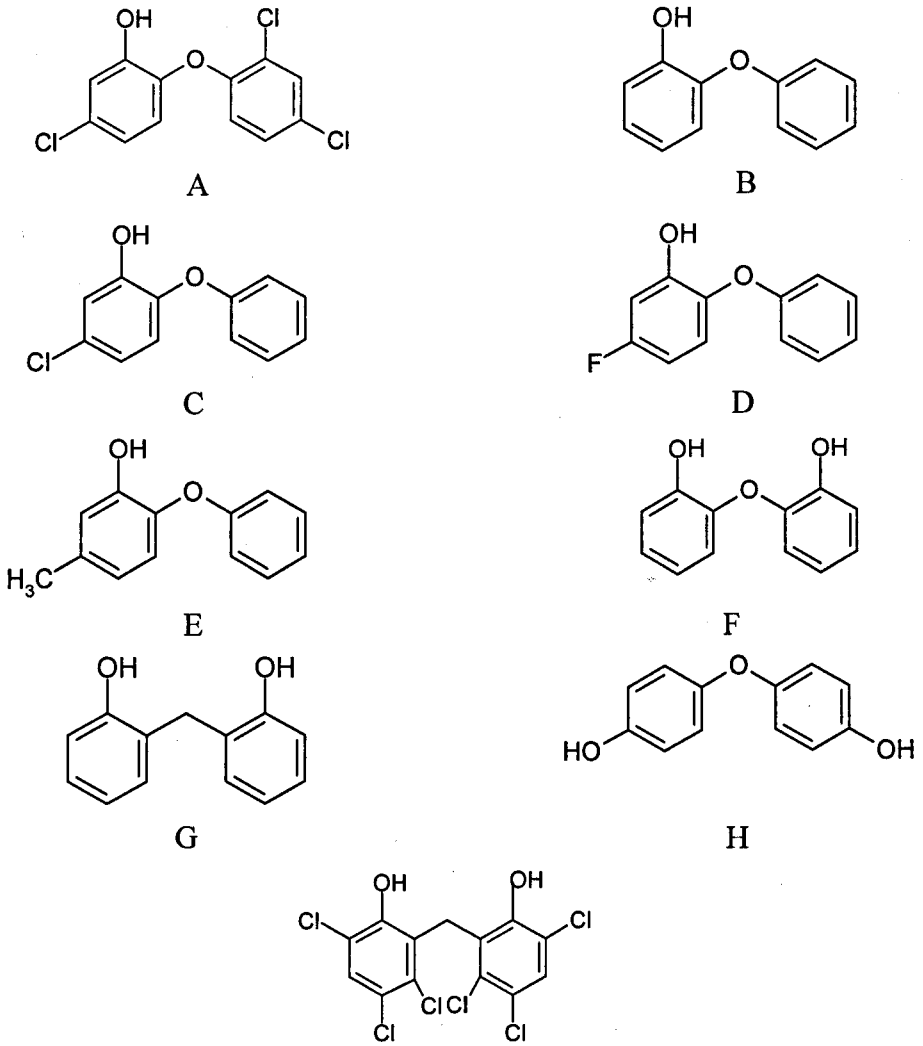


Figure 9.14. The structural formulae for several triclosan derivatives, produced using the program Isis draw.

Compound	MIC *E.coli* (μg/ml)	MIC *S.aureus* (μg/ml)
Triclosan (A)	0.3[a]	≤0.06[b]
B	3.7[a]	-
C	0.07[a]	-
D	0.6[a]	-
E	1.0[a]	-
F	2.0[a]	-
G	>40[a]	-
H	>100[a]	-
Hexachlorophene	>64[c]	1[c]

Table 9.1. The MIC values (μg/ml) for the triclosan derivatives given in Figure 9.14. The values are taken from **a** Sivaraman *et al.*, 2004, **b** Miller *et al.*, 2002 and **c** Heath *et al.*, 2000.

9.8 The possible development of resistance to new ENR inhibitors

9.8.1 Sequence conservation of residues involved in inhibitor binding

Analysis of the sequence conservation for those residues shown to be within 4Å of the compound 29 inhibitor in complex with EcENR showed that of the 11 residues identified (Gly^{93}, Phe^{94}, Ala^{95}, Leu^{100}, Tyr^{146}, Tyr^{156}, Met^{159}, Ala^{196}, Ile^{200}, Lys^{201} and Met^{206}) three residues are not conserved and may be hot spots for sequence variation (Met^{206}, Lys^{201} and Leu^{100}) ((**Figure 9.15a**)(**Figure 5.1**)). In contrast of those residues which are within 4Å of the inhibitor triclosan in the EcENR triclosan complex (Gly^{93}, Phe^{94}, Ala^{95}, Tyr^{146}, Tyr^{156}, Lys^{163}, Ala^{196}, Ala^{197}, Ile^{200} and Phe^{203}) (**Figure 9.15b**) all are either fully or strongly conserved. Seven residues form packing contacts with both triclosan and compound 29 (Gly^{93}, Phe^{94}, Ala^{95}, Tyr^{146}, Tyr^{156}, Ala^{196} and Ile^{200}).

9.8.2 Reported resistance of the ENR enzyme to triclosan

Although triclosan has been used in a variety of household products few cases of triclosan resistant bacterial strains have been reported. However triclosan resistance can occur through both the overexpression of the FabI gene and as a result of amino acid changes in ENR (Fan *et al.*, 2002). Several mutation within *S.aureus* ENR have been shown to decrease triclosan susceptibility and include G93V/S, which would introduce some steric hindrance within the binding site and A198G, L208F and F204C the latter of which reportedly reduces the van der Waals packing interaction to the inhibitor in the NAD^+/triclosan complex ((Brenwald & Fraise 2003)(Fan *et al.*, 2002)). Interestingly the F203C mutation does not change *S.aureus* ENR susceptibility to compound 29 or related compounds. Of the currently reported

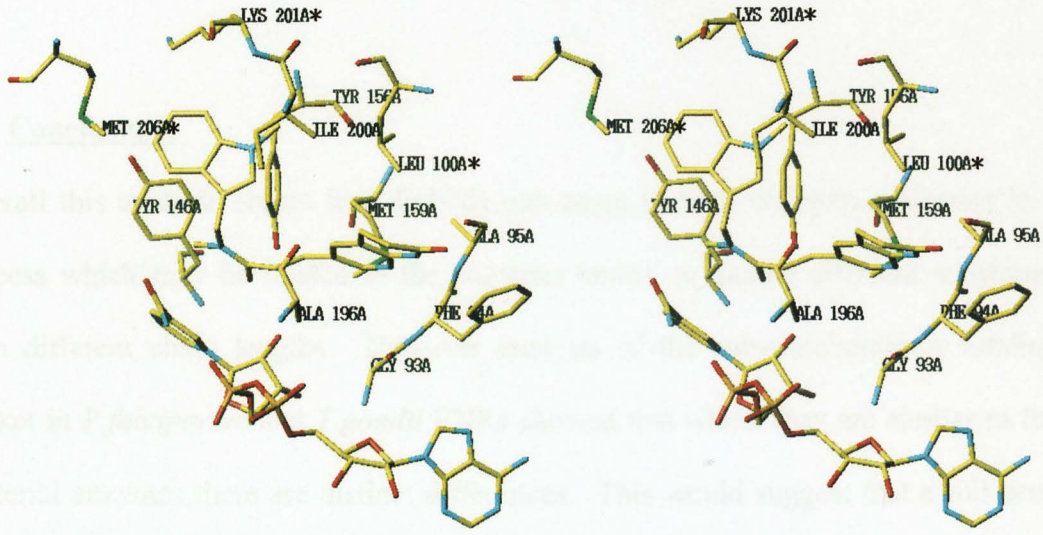


Figure 9.15 (a). A stereo view of the EcENR/NAD^+/compound 29 complex structure. Only those residues which are within 4.0Å of the bound inhibitor are shown and labelled. The figure is coloured with yellow for carbon, red for oxygen, blue for nitrogen, green for sulphur and orange for phosphorous. Those residues with low sequence conservation across the ENR family are highlighted by a star. The figure was produced in the graphics program TURBO/FRODO (Roussel *et al.*, 1990).

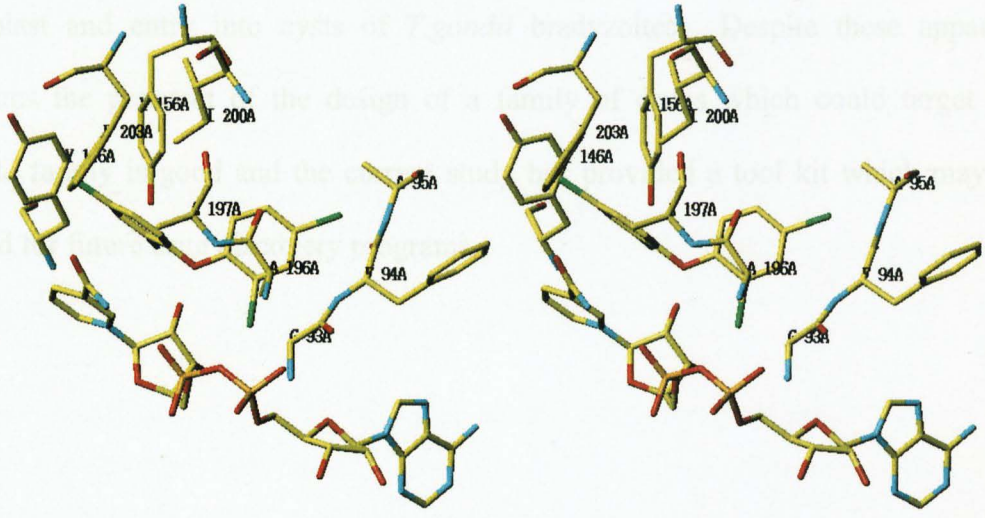


Figure 9.15 (b). A stereo view of the EcENR/NAD^+/triclosan complex structure. Only those residues which are within 4.0Å of the bound inhibitor are shown and labelled. The figure is coloured with yellow for carbon, red for oxygen, blue for nitrogen, green for sulphur and orange for phosphorous. The figure was produced in the graphics program TURBO/FRODO (Roussel *et al.*, 1990).

mutations within the ENR enzyme of clinical strains of *S.aureus* which confer triclosan resistance non totally abolish triclosan inhibition.

9.9 Conclusions

Overall this analysis shows how EcENR can adapt to bind different inhibitors in a process which may be related to the enzymes ability to handle different substrates with different chain lengths. However analysis of the substrate/inhibitor binding pocket in *P.falciparum* and *T.gondii* ENRs showed that whilst they are similar to the bacterial enzymes there are distinct differences. This would suggest that a full drug discovery program to develop an antiparasitic agent would need to utilise the parasite enzymes themselves in the screening and would require a full structural study of inhibitor binding to the Tg or Pf ENR enzyme. Moreover it is clear that such studies also have to address the problem of crossing the additional membrane of the apicoplast and entry into cysts of *T.gondii* bradyzoites. Despite these apparent problems the prospect of the design of a family of drugs which could target the parasite family is good and the current study has provided a tool kit which may be utilised for future drug discovery programs.

Anders, R. F. (1986). Multiple cross-reactivities amongst antigens of *Plasmodium falciparum* impair the development of protective immunity against malaria. *Parasite Immunol* **8**, 529-39.

Aravind, L., Iyer, L. M., Wellems, T. E. & Miller, L. H. (2003). *Plasmodium* biology: genomic gleanings. *Cell* **115**, 771-85.

Athappilly, F. K. & Hendrickson, W. A. (1995). Structure of the biotinyl domain of acetyl-coenzyme A carboxylase determined by MAD phasing. *Structure* **3**, 1407-1419.

Baldock, C., Rafferty, J. B., Sedelnikova, S. E., Baker, P. J., Stuitje, A. R., Slabas, A. R., Hawkes, T. R. & Rice, D. W. (1996). A mechanism of drug action revealed by structural studies of enoyl reductase. *Science* **274**, 2107-2110.

Baldock, C., Rafferty, J. B., Stuitje, A. R., Slabas, A. R. & Rice, D. W. (1998). The X-ray structure of *Escherichia coli* enoyl reductase with bound NAD^+ at 2.1 A resolution. *J Mol Biol* **284**, 1529-1546.

Barton, G. J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng* **6**, 37-40.

Banerjee, A., Dubnau, E., Quemard, A., Balasubramanian, V., Um, K. S., Wilson, T., Collins, D., de Lisle, G. & Jacobs, W. R., Jr. (1994). inhA, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*. *Science* **263**, 227-230.

Benach, J., Atrian, S., Gonzalez-Duarte, R. & Ladenstein, R. (1998). The refined crystal structure of *Drosophila lebanonensis* alcohol dehydrogenase at 1.9 Å resolution. *J Mol Biol* **282**, 383-399.

Bhargava, H. N. & Leonard, P. A. (1996). Triclosan: applications and safety. *Am J Infect Control* **24**, 209-218.

Blow, D. M (2002). Outline of crystallography for biologists. *Oxford University Press*.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S. & Schneider, M. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* **31**, 365-370.

Bohne, W., Heesemann, J. & Gross, U. (1994). Reduced replication of T*oxoplasma gondii* is necessary for induction of bradyzoite-specific antigens: a possible role for nitric oxide in triggering stage conversion. *Infect Immun* **62**, 1761-1767.

Boyer, K. & McLeod, R. (2002) Principles and practice of pediatric infectious diseases, eds. Long, S., Proeber, C. & Pickering, L. (Churchill Livingstone, New York), 2nd Ed., pp. 1303-1322.

Bragg, W.L. (1913). The structure of crystals as indicated by their diffraction of X-rays. *Proceedings of the Royal Society*, **A89**, 248-277.

Breman, J. G., Egan, A. & Keusch, G. T. (2001). The intolerable burden of malaria: a new look at the numbers. *Am J Trop Med Hyg* **64**, iv-vii.

Bricogne, G. (1997). Bayesian statistical viewpoint on structure determination: basic concepts and examples. *Methods in Enzymology,* **276,** 361-423.

Britton, D. (1972). Estimation of twinning parameters for twins with exactly superimposed reciprocal lattices. *Acta Cryst.* A**28**, 296-297.

Brunger, A. T. (1997). Free R value: cross-validation in crystallography. *Methods Enzymol.* **277**, 366-396.

Carruthers, V. B. (2002). Host cell invasion by the opportunistic pathogen *Toxoplasma gondii. Acta Trop* **81**, 111-122.

Clarke, J. L., Sodeinde, O. & Mason, P. J. (2003). A unique insertion in *Plasmodium berghei* glucose-6-phosphate dehydrogenase-6-phosphogluconolactonase: evolutionary and functional studies. *Mol Biochem Parasitol* **127**, 1-8.

Collaborative Computational Project No. 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. Sect. D,* **50**, 760-763.

Cowtan, K. (1994). DM an automated procedure for phase improvement by density modification. *In Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* (*Bailey, S. and Wilson, K., eds*) *SERC* Daresbury, **31**, 34-38.

Crowther, R.A., and Blow, D.M. (1967). A method of positioning a known molecule in an unknown crystal structure. *Acta. Crystallographica*, **23**, 544-548.

Daines, R. A., Pendrak, I., Sham, K., Van Aller, G. S., Konstantinidis, A. K., Lonsdale, J. T., Janson, C. A., Qiu, X., Brandt, M., Khandekar, S. S., Silverman, C. & Head, M. S. (2003). First X-ray cocrystal structure of a bacterial FabH condensing enzyme and a small molecule inhibitor achieved using rational design and homology modeling. *J Med Chem* **46**, 5-8.

Dessen, A., Quemard, A., Blanchard, J. S., Jacobs, W. R., Jr. & Sacchettini, J. C. (1995). Crystal structure and function of the isoniazid target of *Mycobacterium tuberculosis*. *Science* **267**, 1638-1641.

Delano, W. L. (2002) The PyMOL Molecular Graphics system. Delano scientific LLC, San Carlos, CA, USA. http://www.pymol.org

Djordjevic, S., Pace, C. P., Stankovich, M. T. & Kim, J. J. (1995). Three-dimensional structure of butyryl-CoA dehydrogenase from *Megasphaera elsdenii*. *Biochemistry* **34**, 2163-2171.

Dondorp, A. M., Kager, P. A., Vreeken, J. & White, N. J. (2000). Abnormal blood flow and red blood cell deformability in severe malaria. *Parasitol Today* **16**, 228-32.

van Dooren, G. G., Su, V., D'Ombrain, M. C. & McFadden, G. I. (2002). Processing of an apicoplast leader sequence in *Plasmodium falciparum* and the identification of a putative leader cleavage enzyme. *J Biol Chem* **277**, 23612-23619.

Drenth, J. (1994). Principles of protein X-ray crystallography. (Cantor, C.R., ed) Springer-Verlag, New York.

Dubey, J. P. (1998). Advances in the life cycle of *Toxoplasma gondii*. *Int J Parasitol* **28**, 1019-1024.

Dunn, C. R., Banfield, M. J., Barker, J. J., Higham, C. W., Moreton, K. M., Turgut-Balik, D., Brady, R. L. & Holbrook, J. J. (1996). The structure of lactate dehydrogenase from *Plasmodium falciparum* reveals a new target for anti-malarial design. *Nat Struct Biol* **3**, 912-915.

Eckstein-Ludwig, U., Webb, R. J., Van Goethem, I. D., East, J. M., Lee, A. G., Kimura, M., O'Neill, P. M., Bray, P. G., Ward, S. A. & Krishna, S. (2003). Artemisinins target the SERCA of *Plasmodium falciparum*. *Nature* **424**, 957-961.

Elborough, K. M., Winz, R., Deka, R. K., Markham, J. E., White, A. J., Rawsthorne, S. & Slabas, A. R. (1996). Biotin carboxyl carrier protein and carboxyltransferase subunits of the multi-subunit form of acetyl-CoA carboxylase from *Brassica napus*: cloning and analysis of expression during oilseed rape embryogenesis. *Biochem J* **315** (Pt 1), 103-112.

English, M., Sauerwein, R., Waruiru, C., Mosobo, M., Obiero, J., Lowe, B. & Marsh, K. (1997). Acidosis in severe childhood malaria. *Qjm* **90**, 263-270.

English[a], M., Waruiru, C. & Marsh, K. (1996). Transfusion for respiratory distress in life-threatening childhood malaria. *Am J Trop Med Hyg* **55**, 525-30.

English[b], M. C., Waruiru, C., Lightowler, C., Murphy, S. A., Kirigha, G. & Marsh, K. (1996). Hyponatraemia and dehydration in severe malaria. *Arch Dis Child* **74**, 201-205.

Evans, P. R. (1993). "Data reduction", proceedings of CCP4 study weekend, 1993, on data collection and processing. Pgs 114-122.

Ewald, P. P. (1979). A review of my papers on crystal optics. *Acta Cryst.* **A35**, 1-9

Fan, F., Yan, K., Wallis, N. G., Reed, S., Moore, T. D., Rittenhouse, S. F., DeWolf, W. E., Jr., Huang, J., McDevitt, D., Miller, W. H., Seefeld, M. A., Newlander, K. A., Jakas, D. R., Head, M. S. & Payne, D. J. (2002). Defining and combating the

mechanisms of triclosan resistance in clinical isolates of *Staphylococcus aureus*. *Antimicrob Agents Chemother* **46**, 3343-3347.

Fast, N. M., Kissinger, J. C., Roos, D. S. & Keeling, P. J. (2001). Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol Biol Evol* **18**, 418-426.

Fichera, M. E. & Roos, D. S. (1997). A plastid organelle as a drug target in apicomplexan parasites. *Nature* **390**, 407-409.

Filling, C., Berndt, K. D., Benach, J., Knapp, S., Prozorovski, T., Nordling, E., Ladenstein, R., Jornvall, H. & Oppermann, U. (2002). Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. *J Biol Chem* **277**, 25677-25684.

Filling, C., Nordling, E., Benach, J., Berndt, K. D., Ladenstein, R., Jornvall, H. & Oppermann, U. (2001). Structural role of conserved Asn179 in the short-chain dehydrogenase/reductase scaffold. *Biochem Biophys Res Commun* **289**, 712-717.

Fisher, M., Kroon, J. T., Martindale, W., Stuitje, A. R., Slabas, A. R. & Rafferty, J. B. (2000). The X-ray structure of *Brassica napus* beta-keto acyl carrier protein reductase and its implications for substrate binding and catalysis. *Structure Fold Des* **8**, 339-347.

Foth, B. J., Ralph, S. A., Tonkin, C. J., Struck, N. S., Fraunholz, M., Roos, D. S., Cowman, A. F. & McFadden, G. I. (2003). Dissecting apicoplast targeting in the malaria parasite *Plasmodium falciparum*. *Science* **299**, 705-708.

Fry, M. & Pudney, M. (1992). Site of action of the antimalarial hydroxynaphthoquinone, 2-[trans-4-(4'-chlorophenyl) cyclohexyl]-3-hydroxy-1,4-naphthoquinone (566C80). *Biochem Pharmacol* **43**, 1545-53.

Gewirth, D. (1999). The HKL manual edition 5, HKL Research. (www.hkl-xray.com/hkl/manual_online.pdf)

Ginsburg, H., Ward, S. A. & Bray, P. G. (1999). An integrated model of chloroquine action. *Parasitol Today* **15**, 357-360.

Gleeson, M. T. (2000). The plastid in Apicomplexa: what use is it? *Int J Parasitol* **30**, 1053-1070.

Grassberger, M. A., Turnowsky, F. & Hildebrandt, J. (1984). Preparation and antibacterial activities of new 1,2,3-diazaborine derivatives and analogues. *J Med Chem* **27**, 947-953.

He, X. & Reynolds, K. A. (2002). Purification, characterization, and identification of novel inhibitors of the beta-ketoacyl-acyl carrier protein synthase III (FabH) from *Staphylococcus aureus*. *Antimicrob Agents Chemother* **46**, 1310-1318.

Heath, R. J. & Rock, C. O. (1996). Roles of the FabA and FabZ beta-hydroxyacyl-acyl carrier protein dehydratases in *Escherichia coli* fatty acid biosynthesis. *J Biol Chem* **271**, 27795-27801.

Heath, R. J., Li, J., Roland, G. E. & Rock, C. O. (2000). Inhibition of the *Staphylococcus aureus* NADPH-dependent enoyl-acyl carrier protein reductase by triclosan and hexachlorophene. *J Biol Chem* **275**, 4654-4659.

Heerding, D. A., Chan, G., DeWolf, W. E., Fosberry, A. P., Janson, C. A., Jaworski, D. D., McManus, E., Miller, W. H., Moore, T. D., Payne, D. J., Qiu, X., Rittenhouse, S. F., Slater-Radosti, C., Smith, W., Takata, D. T., Vaidya, K. S., Yuan, C. C. & Huffman, W. F. (2001). 1,4-Disubstituted imidazoles are potential antibacterial agents functioning as inhibitors of enoyl acyl carrier protein reductase (FabI). *Bioorg Med Chem Lett* **11**, 2061-2065.

Higgins, D. & Thompson, J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting position-specific gap penalties and weighting matrix choice. *Nucleic Acids Res.* **22**. 4673-4680.

Hill, D. & Dubey, J. P. (2002). *Toxoplasma gondii*: transmission, diagnosis and prevention. *Clin Microbiol Infect* **8**, 634-640.

Hutchinson, E. G. & Thornton, J. M. (1996). PROMOTIF-a program to identify and analyze structural motifs in proteins. *Protein Sci* **5**, 212-220.

Jomaa, H., Wiesner, J., Sanderbrand, S., Altincicek, B., Weidemeyer, C., Hintz, M., Turbachova, I., Eberl, M., Zeidler, J., Lichtenthaler, H. K., Soldati, D. & Beck, E. (1999). Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* **285**, 1573-1576.

Kabsch, W. (1976) A solution for the best rotation to relate two sets of vectors. *Acta Cryst.* **A32** 922-923.(Lsqkab)

Kapoor, M., Dar, M. J., Surolia, A. & Surolia, N. (2001). Kinetic determinants of the interaction of enoyl-ACP reductase from *Plasmodium falciparum* with its substrates and inhibitors. *Biochem Biophys Res Commun* **289**, 832-837.

Kappes, B., Yang, J., Suetterlin, B. W., Rathgeb-Szabo, K., Lindt, M. J. & Franklin, R. M. (1995). A *Plasmodium falciparum* protein kinase with two unusually large kinase inserts. *Mol Biochem Parasitol* **72**, 163-178.

Keatinge-Clay, A. T., Shelat, A. A., Savage, D. F., Tsai, S. C., Miercke, L. J., O'Connell, J. D., 3rd, Khosla, C. & Stroud, R. M. (2003). Catalysis, specificity, and ACP docking site of *Streptomyces coelicolor* malonyl-CoA:ACP transacylase. *Structure (Camb)* **11**, 147-154.

Kohler, S., Delwiche, C. F., Denny, P. W., Tilney, L. G., Webster, P., Wilson, R. J., Palmer, J. D. & Roos, D. S. (1997). A plastid of probable green algal origin in Apicomplexan parasites. *Science* **275**, 1485-1489.

Kraulis, p. J. (1991). Molscript: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog* **24**, 946-950.

Lang, T. & Greenwood, B. (2003). The development of Lapdap, an affordable new treatment for malaria. *Lancet Infect Dis* **3**, 162-168.

Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.

Leesong, M., Henderson, B. S., Gillig, J. R., Schwab, J. M. & Smith, J. L. (1996). Structure of a dehydratase-isomerase from the bacterial pathway for biosynthesis of unsaturated fatty acids: two catalytic activities in one active site. *Structure* **4**, 253-264.

Leslie, A. G. (1999). Integration of macromolecular diffraction data. *Acta Cryst* ***D55***, 1690-1695.

Levy, C. W., Roujeinikova, A., Sedelnikova, S., Baker, P. J., Stuitje, A. R., Slabas, A. R., Rice, D. W. & Rafferty, J. B. (1999). Molecular basis of triclosan activity. *Nature* **398**, 383-384.

Levy, C. W., Baldock, C., Wallace, A. J., Sedelnikova, S., Viner, R. C., Clough, J. M., Stuitje, A. R., Slabas, A. R., Rice, D. W. & Rafferty, J. B. (2001). A study of the structure-activity relationship for diazaborine inhibition of *Escherichia coli* enoyl-ACP reductase. *J Mol Biol* **309**, 171-180.

Li, J. L. & Baker, D. A. (1998). A putative protein serine/threonine phosphatase from *Plasmodium falciparum* contains a large N-terminal extension and five unique inserts in the catalytic domain. *Mol Biochem Parasitol* **95**, 287-295.

Mac Sweeney, A. and D'Arcy, A. (2003). A simple and rapid method for mounting protein crystals at room temperature. *Applied Crystallography* **36**, 165-166

Magnuson, K., Jackowski, S., Rock, C. O. & Cronan, J. E., Jr. (1993). Regulation of fatty acid biosynthesis in *Escherichia coli. Microbiol Rev* **57**, 522-542.

Matthews, B. W. (1977). X-ray structure of proteins, edited by H Neurath & R.L.

McCarthy, A. D. & Hardie, D. G. (1984). Fatty acid synthase and example of protein evolution by gene fusion. *Trends Biochem.Sci.* **9**, 60-63

McFadden, G. I. & Roos, D. S. (1999). Apicomplexan plastids as drug targets. *Trends Microbiol* **7**, 328-333.

McLeod, R., Muench, S. P., Rafferty, J. B., Kyle, D. E., Mui, E. J., Kirisits, M. J., Mack, D. G., Roberts, C. W., Samuel, B. U., Lyons, R. E., Dorris, M., Milhous, W. K. & Rice, D. W. (2001). Triclosan inhibits the growth of *Plasmodium falciparum* and *Toxoplasma gondii* by inhibition of apicomplexan Fab I. *Int J Parasitol* **31**, 109-113.

McMurry, L. M., Oethinger, M. & Levy, S. B. (1998). Triclosan targets lipid synthesis. *Nature* **394**, 531-532.

Mead, P. S., Slutsker, L., Dietz, V., McCaig, L. F., Bresee, J. S., Shapiro, C., Griffin, P. M. & Tauxe, R. V. (1999). Food-related illness and death in the United States. *Emerg Infect Dis* **5**, 607-625.

de Medeiros, B. C., de Medeiros, C. R., Werner, B., Loddo, G., Pasquini, R. & Bleggi-Torres, L. F. (2001). Disseminated toxoplasmosis after bone marrow transplantation: report of 9 cases. *Transpl Infect Dis* **3**, 24-8.

Meshnick, S. R. (2002). Artemisinin: mechanisms of action, resistance and toxicity. *Int J Parasitol* **32**, 1655-1660.

Miller[a], L. H., Baruch, D. I., Marsh, K. & Doumbo, O. K. (2002). The pathogenic basis of malaria. *Nature* **415**, 673-9.

Miller[b], W. H., Seefeld, M. A., Newlander, K. A., Uzinskas, I. N., Burgess, W. J., Heerding, D. A., Yuan, C. C., Head, M. S., Payne, D. J., Rittenhouse, S. F., Moore, T. D., Pearson, S. C., Berry, V., DeWolf, W. E., Jr., Keller, P. M., Polizzi, B. J., Qiu, X., Janson, C. A. & Huffman, W. F. (2002). Discovery of aminopyridine-based inhibitors of bacterial enoyl-ACP reductase (FabI). *J Med Chem* **45**, 3246-3256.

Mitchell, D. J. (2000) Polyarginine enters cells more efficiently than other polycationic homopolymers. *J. Peptide Res.* **56**, 318-325

Molyneux, D. H., Floyd, K., Barnish, G. & Fevre, E. M. (1999). Transmission control and drug resistance in malaria: a crucial interaction. *Parasitol Today* **15**, 238-240.

Moon, J. H., Lee, H. H. & Suw, S. W. (2001). *H. pylori* ENR (unpublished work) PDB entry 1JW7 (www.rcsb.org/pdb/).

Mota, M. M., Pradel, G., Vanderberg, J. P., Hafalla, J. C., Frevert, U., Nussenzweig, R. S., Nussenzweig, V. & Rodriguez, A. (2001). Migration of *Plasmodium* sporozoites through cells before infection. *Science* **291**, 141-144.

Muench, S. P., Rafferty, J. B., McLeod, R., Rice, D. W. & Prigge, S. T. (2003). Expression, purification and crystallization of the *Plasmodium falciparum* enoyl reductase. *Acta Crystallogr D Biol Crystallogr* **59**, 1246-1248.

Murshudov, G. N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta. Crystallographica,* **D53**, 240-255.

Navaza, J. (1994). AMoRe-an automated package for molecular replacement. *Acta. Crystallographica,* **A50**, 157-163.

Newton, P. & White, N. (1999). Malaria: new developments in treatment and prevention. *Annu Rev Med* **50**, 179-192.

Nicholls, A, Sharp, K. & Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Protein. Structure. Function.& Genetics*. **11** 281.

Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* **10**, 1-6.

Olliaro, P. L., Haynes, R. K., Meunier, B. & Yuthavong, Y. (2001). Possible modes of action of the artemisinin-type compounds. *Trends Parasitol* **17**, 122-126.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology*, **276**, 307-326.

Parola, P. & Miller, R. S. (2002). Quinine in the modern treatment of *falciparum* malaria. *Lancet Infect Dis* **2**, 206-207.

Parsons, S. (2003). Introduction to twinning. *Acta Cryst.* D**59**, 1995-2003

Patterson, A. L. (1934). Ambiguities in the X-ray analysis of crystal structures. *Phys. Rev.* **65**, 195-201.

Payne, D. J., Miller, W. H., Berry, V., Brosky, J., Burgess, W. J., Chen, E., DeWolf Jr, W. E., Jr., Fosberry, A. P., Greenwood, R., Head, M. S., Heerding, D. A., Janson, C. A., Jaworski, D. D., Keller, P. M., Manley, P. J., Moore, T. D., Newlander, K. A., Pearson, S., Polizzi, B. J., Qiu, X., Rittenhouse, S. F., Slater-Radosti, C., Salyers, K. L., Seefeld, M. A., Smyth, M. G., Takata, D. T., Uzinskas, I. N., Vaidya, K., Wallis, N. G., Winram, S. B., Yuan, C. C. & Huffman, W. F. (2002). Discovery of a novel and potent class of FabI-directed antibacterial agents. *Antimicrob Agents Chemother* **46**, 3118-3124.

Perozzo, R., Kuo, M., Sidhu, A. S., Valiyaveettil, J. T., Bittman, R., Jacobs, W. R., Jr., Fidock, D. A. & Sacchettini, J. C. (2002). Structural elucidation of the specificity of the antibacterial agent triclosan for malarial enoyl acyl carrier protein reductase. *J Biol Chem* **277**, 13106-13114.

Perrakis, A., Sixma, T. K., Wilson, K. S., and Lamzin, V. S. (1997). wARP: improvement and extension of crystallographic phases by weighted averaging of multiple refined dummy atomic models. *Acta. Crystallographica*, **D53**, 448-455.

Pharmacopeia 2003 edition. www.usp.org

Pincock, S. (2003). Drug company to offer new malaria drug cheaply in Africa. *BMJ* **327**, 360.

Pizzi, E. & Frontali, C. (2001). Low-complexity regions in *Plasmodium falciparum* proteins. *Genome Res* **11**, 218-29.

Price[a], A. C., Choi, K. H., Heath, R. J., Li, Z., White, S. W. & Rock, C. O. (2001). Inhibition of beta-ketoacyl-acyl carrier protein synthases by thiolactomycin and cerulenin. Structure and mechanism. *J Biol Chem* **276**, 6551-6559.

Price[b], A. C., Zhang, Y. M., Rock, C. O. & White, S. W. (2001). Structure of beta-ketoacyl-[acyl carrier protein] reductase from *Escherichia coli*: negative cooperativity and its structural basis. *Biochemistry* **40**, 12772-12781.

Price, A. C., Zhang, Y. M., Rock, C. O. & White, S. W. (2004). Cofactor-induced conformational rearrangements establish a catalytically competent active site and a proton relay conduit in FabG. *Structure (Camb)* **12**, 417-428.

Qiu[a], X., Janson, C. A., Court, R. I., Smyth, M. G., Payne, D. J. & Abdel-Meguid, S. S. (1999). Molecular basis for triclosan activity involves a flipping loop in the active site. *Protein Sci* **8**, 2529-2532.

Qiu[b], X., Janson, C. A., Konstantinidis, A. K., Nwagwu, S., Silverman, C., Smith, W. W., Khandekar, S., Lonsdale, J. & Abdel-Meguid, S. S. (1999). Crystal structure of beta-ketoacyl-acyl carrier protein synthase III. A key condensing enzyme in bacterial fatty acid biosynthesis. *J Biol Chem* **274**, 36465-36471.

Qiu, X., Janson, C. A., Smith, W. W., Head, M., Lonsdale, J. & Konstantinidis, A. K. (2001). Refined structures of beta-ketoacyl-acyl carrier protein synthase III. *J Mol Biol* **307**, 341-356.

Rafferty, J. B., Simon, J. W., Baldock, C., Artymiuk, P. J., Baker, P. J., Stuitje, A. R., Slabas, A. R. & Rice, D. W. (1995). Common themes in redox chemistry emerge from the X-ray structure of oilseed rape (*Brassica napus*) enoyl acyl carrier protein reductase. *Structure* **3**, 927-938.

Ralph, S. A., D'Ombrain, M. C. & McFadden, G. I. (2001). The apicoplast as an antimalarial drug target. *Drug Resist Updat* **4**, 145-151.

Ramachandran GN, Ramakrishnan C, Sasisekharan V. (1963) Stereochemistry of polypeptide chain configurations. *J Mol Biol* **7**: 95-99.

Raviglione, M. C., Snider, D. E., Jr. & Kochi, A. (1995). Global epidemiology of tuberculosis. Morbidity and mortality of a worldwide epidemic. *Jama* **273**, 220-226.

Rawat, R., Whitty, A. & Tonge, P. J. (2003). The isoniazid-NAD adduct is a slow, tight-binding inhibitor of InhA, the *Mycobacterium tuberculosis* enoyl reductase: adduct affinity and drug resistance. *Proc Natl Acad Sci U S A* **100**, 13881-13886.

Razakantoanina, V., Nguyen Kim, P. P. & Jaureguiberry, G. (2000). Antimalarial activity of new gossypol derivatives. *Parasitol Res* **86**, 665-668.

Read, J. A., Wilkinson, K. W., Tranter, R., Sessions, R. B. & Brady, R. L. (1999). Chloroquine binds in the cofactor binding site of *Plasmodium falciparum* lactate dehydrogenase. *J Biol Chem* **274**, 10213-10218.

Ridley, R. G. (2002). Medical need, scientific opportunity and the drive for antimalarial drugs. *Nature* **415**, 686-693.

Ridley, R. G. (2003). Malaria: to kill a parasite. *Nature* **424**, 887-889.

Roberts, T., Murrell, K. D. & Marks, S. (1994). Economic losses caused by foodborne parasitic diseases. *Parasitology Today* **10**, 419-423.

Rock, C. O. & Cronan, J. E. (1996). *Escherichia coli* as a model for the regulation of dissociable (type II) fatty acid biosynthesis. *Biochim Biophys Acta* **1302**, 1-16.

Rossmann, M. G., Moras, D. & Olsen, K. W. (1974). Chemical and biological evolution of nucleotide-binding protein. *Nature* **250**, 194-199.

Roussel, A., Fontecilla-Camps, J.C., & Cambillau, C. (1990). TURBO-FRODO: a new program for protein crystallography and modeling. XV IUCr Congress Abstracts, Bordeux, France, pp. 66-67.

Roujeinikova[a], A., Levy, C. W., Rowsell, S., Sedelnikova, S., Baker, P. J., Minshull, C. A., Mistry, A., Colls, J. G., Camble, R., Stuitje, A. R., Slabas, A. R., Rafferty, J. B., Pauptit, R. A., Viner, R. & Rice, D. W. (1999). Crystallographic analysis of triclosan bound to enoyl reductase. *J Mol Biol* **294**, 527-535.

Roujeinikova[b], A., Sedelnikova, S., de Boer, G. J., Stuitje, A. R., Slabas, A. R., Rafferty, J. B. & Rice, D. W. (1999). Inhibitor binding studies on enoyl reductase reveal conformational changes related to substrate recognition. *J Biol Chem* **274**, 30811-30817.

Roujeinikova, A., Baldock, C., Simon, W. J., Gilroy, J., Baker, P. J., Stuitje, A. R., Rice, D. W., Slabas, A. R. & Rafferty, J. B. (2002). X-ray crystallographic studies on butyryl-ACP reveal flexibility of the structure around a putative acyl chain binding site. *Structure (Camb)* **10**, 825-835.

Rozwarski, D. A., Grant, G. A., Barton, D. H., Jacobs, W. R., Jr. & Sacchettini, J. C. (1998). Modification of the NADH of the isoniazid target (InhA) from *Mycobacterium tuberculosis*. *Science* **279**, 98-102.

Rozwarski, D. A., Vilcheze, C., Sugantino, M., Bittman, R. & Sacchettini, J. C. (1999). Crystal structure of the *Mycobacterium tuberculosis* enoyl-ACP reductase, InhA, in complex with NAD^+ and a C16 fatty acyl substrate. *J Biol Chem* **274**, 15582-15589.

Samuel, B. U., Hearn, B., Mack, D., Wender, P., Rothbard, J., Kirisits, M. J., Mui, E., Wernimont, S., Roberts, C. W., Muench, S. P., Rice, D. W., Prigge, S. T., Law, A. B. & McLeod, R. (2003). Delivery of antimicrobials into parasites. *Proc Natl Acad Sci U S A* **100**, 14281-14286.

Sanchez, C. P. & Lanzer, M. (2000). Changing ideas on chloroquine in *Plasmodium falciparum*. *Curr Opin Infect Dis* **13**, 653-658.

Sanchez, C. P., Stein, W. & Lanzer, M. (2003). Trans stimulation provides evidence for a drug efflux carrier as the mechanism of chloroquine resistance in *Plasmodium falciparum*. *Biochemistry* **42**, 9383-9394.

Schneider, T. R. (2000). Objective comparison of protein structures: error-scaled difference distance matrices. *Acta Cryst.* **D56**, 714-721.

Seefeld, M. A., Miller, W. H., Newlander, K. A., Burgess, W. J., DeWolf, W. E., Jr., Elkins, P. A., Head, M. S., Jakas, D. R., Janson, C. A., Keller, P. M., Manley, P. J., Moore, T. D., Payne, D. J., Pearson, S., Polizzi, B. J., Qiu, X., Rittenhouse, S. F., Uzinskas, I. N., Wallis, N. G. & Huffman, W. F. (2003). Indole naphthyridinones as inhibitors of bacterial enoyl-ACP reductases FabI and FabK. *J Med Chem* **46**, 1627-1635.

Serre, L., Verbree, E. C., Dauter, Z., Stuitje, A. R. & Derewenda, Z. S. (1995). The *Escherichia coli* malonyl-CoA:acyl carrier protein transacylase at 1.5-Å resolution. Crystal structure of a fatty acid synthase component. *J Biol Chem* **270**, 12961-12964.

Sibley, L. D. (2003). *Toxoplasma gondii*: perfecting an intracellular life style. *Traffic* **4**, 581-6.

Singer, G. A. & Hickey, D. A. (2000). Nucleotide bias causes a genomewide bias in the amino acid composition of proteins. *Mol Biol Evol* **17**, 1581-8.

Sivaraman, S., Sullivan, T. J., Johnson, F., Novichenok, P., Cui, G., Simmerling, C. & Tonge, P. J. (2004). Inhibition of the bacterial enoyl reductase FabI by triclosan: a structure-reactivity analysis of FabI inhibition by triclosan analogues. *J Med Chem* **47**, 509-518.

Smith, S. (1994). The animal fatty acid synthase: one gene, one polypeptide, seven enzymes. *Faseb J* **8**, 1248-1259.

Smith, S. (2003). Structural and functional organization of the animal fatty acid synthase. *Porg. Lipid res.* **42**, 289-317.

Srivastava, I. K., Morrisey, J. M., Darrouzet, E., Daldal, F. & Vaidya, A. B. (1999). Resistance mutations reveal the atovaquone-binding domain of cytochrome b in malaria parasites. *Mol Microbiol* **33**, 704-711.

Stewart, M. J., Parikh, S., Xiao, G., Tonge, P. J. & Kisker, C. (1999). Structural basis and mechanism of enoyl reductase inhibition by triclosan. *J Mol Biol* **290**, 859-865.

Surolia, N. & Surolia, A. (2001). Triclosan offers protection against blood stages of malaria by inhibiting enoyl-ACP reductase of *Plasmodium falciparum*. *Nat Med* **7**, 167-173.

Teng, T.Y. (1990). Mounting of crystals for macromolecular crystallography in a free-standing thin film. *Journal of Applied Crystallography,* **23**, 387-391.

Tenter, A. M., Heckeroth, A. R. & Weiss, L. M. (2000). *Toxoplasma gondii*: from animals to humans. *Int J Parasitol* **30**, 1217-1258.

Vagin, A. & Teplyakov, A. (2000). An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr* **56** Pt 12, 1622-1624.

Vance, D. E., Esders, T. W. & Bloch, K. (1973). On the role of a palmityl thioesterase in fatty acid elongation. *J Biol Chem* **248**, 2310-2316.

Wakil, S. J. (1989). Fatty acid synthase, A proficient multifunctional enzyme. *Biochemistry* **28**, 4523-4529.

Waldrop, G. L., Rayment, I. & Holden, H. M. (1994). Three-dimensional structure of the biotin carboxylase subunit of acetyl-CoA carboxylase. *Biochemistry* **33**, 10249-10256.

Waller, R. F., Keeling, P. J., Donald, R. G., Striepen, B., Handman, E., Lang-Unnasch, N., Cowman, A. F., Besra, G. S., Roos, D. S. & McFadden, G. I. (1998). Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. *Proc Natl Acad Sci U S A* **95**, 12352-12357.

Waller, R. F., Reed, M. B., Cowman, A. F. & McFadden, G. I. (2000). Protein trafficking to the plastid of *Plasmodium falciparum* is via the secretory pathway. *Embo J* **19**, 1794-1802.

Wang, B. C. (1985). Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol* **115**, 90-112.

Ward, W. H., Holdgate, G. A., Rowsell, S., McLean, E. G., Pauptit, R. A., Clayton, E., Nichols, W. W., Colls, J. G., Minshull, C. A., Jude, D. A., Mistry, A., Timms, D., Camble, R., Hales, N. J., Britton, C. J. & Taylor, I. W. (1999). Kinetic and structural characteristics of the inhibition of enoyl (acyl carrier protein) reductase by triclosan. *Biochemistry* **38**, 12514-12525.

Wender, P. A., Mitchell, D. J., Pattabiraman, K., Pelkey, E. T., Steinman, L. & Rothbard, J. B. (2000). The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: peptoid molecular transporters. *Proc Natl Acad Sci U S A* **97**, 13003-13008.

Wiesner, J., Borrmann, S. & Jomaa, H. (2003). Fosmidomycin for the treatment of malaria. *Parasitol Res* **90** Suppl **2**, S71-76.

Winstanley, P. (2001). Modern chemotherapeutic options for malaria. *Lancet Infect Dis* **1**, 242-250.

Winstanley, P. (2001). Modern chemotherapeutic options for malaria. *Lancet Infect Dis* **1**, 242-250.

Winstanley, P. (2003). The contribution of clinical pharmacology to antimalarial drug discovery and development. *Br J Clin Pharmacol* **55**, 464-468.

Wong, H. C., Liu, G., Zhang, Y. M., Rock, C. O. & Zheng, J. (2002). The solution structure of acyl carrier protein from *Mycobacterium tuberculosis*. *J Biol Chem* **277**, 15874-15880.

Wongsrichanalai, C., Pickard, A. L., Wernsdorfer, W. H. & Meshnick, S. R. (2002). Epidemiology of drug-resistant malaria. *Lancet Infect Dis* **2**, 209-218.

Worsham, L. M., Earls, L., Jolly, C., Langston, K. G., Trent, M. S. & Ernst-Fonberg, M. L. (2003). Amino acid residues of *Escherichia coli* acyl carrier protein involved in heterologous protein interactions. *Biochemistry* **42**, 167-76.

Xu, G. Y., Tam, A., Lin, L., Hixon, J., Fritz, C. C. & Powers, R. (2001). Solution structure of *B. subtilis* acyl carrier protein. *Structure (Camb)* **9**, 277-287.

Yao, X., Wei, D., Soden, C., Jr., Summers, M. F. & Beckett, D. (1997). Structure of the carboxy-terminal fragment of the apo-biotin carboxyl carrier subunit of *Escherichia coli* acetyl-CoA carboxylase. *Biochemistry* **36**, 15089-15100.

Yuvaniyama, J., Chitnumsub, P., Kamchonwongpaisan, S., Vanichtanankul, J., Sirawaraporn, W., Taylor, P., Walkinshaw, M. D. & Yuthavong, Y. (2003). Insights into antifolate resistance from malarial DHFR-TS structures. *Nat Struct Biol* **10**, 357-365.

Zhang, K.Y.J. and Main, P. (1990). The use of Sayre equation with solvent flattening and histogram matching for phase extension and refinement of protein structures. *Acta. Crystallographica,* **A46**,

Appendix I

Triclosan inhibits the growth of *Plasmodium falciparum* and *Toxoplasma gondii* by inhibiting Apicomplexan FabI

McLeod, R., Muench, S. P., Rafferty, J. B., Kyle, D. E., Mui, E. J., Kirisits, M. J., Mack, D. G., Roberts, C. W., Samuel, B. U., Lyons, R. E., Dorris, M., Milhous, W. K. & Rice, D. W.

Int J Parasitol **31**, 109-113.

ELSEVIER

International Journal for Parasitology 31 (2001) 109–113

INTERNATIONAL Journal for PARASITOLOGY

www.parasitology-online.com

Rapid communication

Triclosan inhibits the growth of *Plasmodium falciparum* and *Toxoplasma gondii* by inhibition of Apicomplexan Fab I

Rima McLeod[a,*], Stephen P. Muench[b], John B. Rafferty[b], Dennis E. Kyle[c], Ernest J. Mui[a], Michael J. Kirisits[a], Douglas G. Mack[a], Craig W. Roberts[d], Benjamin U. Samuel[e], Russell E. Lyons[d], Mark Dorris[d], Wilbur K. Milhous[c], David W. Rice[b,1]

[a]*Department of Ophthalmology and Visual Sciences, The University of Chicago, Chicago, IL 60637, USA*
[b]*Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, S10 2TN, UK*
[c]*Walter Reed Army Institute of Research, Silver Spring, MD 20307-5100, USA*
[d]*Department of Immunology, University of Strathclyde, Glasgow, Scotland, G4 0NR, UK*
[e]*Department of Pathology, Northwestern University, Chicago, IL 60611, USA*

Received 22 December 2000; received in revised form 11 January 2001; accepted 11 January 2001

Abstract

Fab I, enoyl acyl carrier protein reductase (ENR), is an enzyme used in fatty acid synthesis. It is a single chain polypeptide in plants, bacteria, and mycobacteria, but is part of a complex polypeptide in animals and fungi. Certain other enzymes in fatty acid synthesis in apicomplexan parasites appear to have multiple forms, homologous to either a plastid, plant-like single chain enzyme or more like the animal complex polypeptide chain. We identified a plant-like *Fab I* in *Plasmodium falciparum* and modelled the structure on the *Brassica napus* and *Escherichia coli* structures, alone and complexed to triclosan (5-chloro-2-[2,4 dichlorophenoxy] phenol]), which confirmed all the requisite features of an ENR and its interactions with triclosan. Like the remarkable effect of triclosan on a wide variety of bacteria, this compound markedly inhibits growth and survival of the apicomplexan parasites *P. falciparum* and *Toxoplasma gondii* at low (i.e. IC50 ≅ 150–2000 and 62 ng/ml, respectively) concentrations. Discovery and characterisation of an apicomplexan Fab I and discovery of triclosan as lead compound provide means to rationally design novel inhibitory compounds.

Keywords: FAB I; Triclosan; *Toxoplasma gondii*; *Plasmodium falciparum*; Lipid synthesis

New medicines are greatly needed for the treatment of apicomplexan infections. Recently a number of plant-like biochemical pathways associated with the vestigial plastid organelle of *Toxoplasma gondii* and *Plasmodium* species have been suggested as new targets for such medicines (Roberts et al., 1998; Waller et al., 1998; Zuther et al., 1999; Jomaa et al., 1999). A particularly attractive target in this respect is the fatty acid biosynthesis pathway as there are major differences between the structure of the plastid-associated enzymes found in plants and the cytosolic enzymes found in mammals (Roberts et al., 1998; Zuther et al., 1999; Waller et al., 2000). Most significantly, enzymes of mammalian lipid synthesis form domains on a multi-functional protein, whereas those enzymes in plants, and certain bacteria are found on discrete mono-functional polypeptides. These differences have already been exploited by a number of compounds which selectively inhibit bacterial or plant enzymes, but not mammalian enzymes (Roberts et al., 1998; Waller et al., 1998; Zuther et al., 1999; Payne et al., 2000). Notably, both *T. gondii* and *Plasmodium falciparum* have been shown to possess mono-functional, plant- or bacterial-like fatty acid biosynthesis enzymes which are targeted to the plastid organelle via a bipartite, N-terminal transit sequence (Waller et al., 1998; Zuther et al., 1999; Roos et al., 1999; DeRocher et al., 2000). Compounds such as aryloxyphenoxypropionates (Zuther et al., 1999) and, cyclohexanedione (Zuther et al., 1999) herbicides and thiolactomycin (Waller et al., 2000) which inhibit acetyl-CoA carboxylase (ACC) and β-ketoacyl-ACP synthase (Fab H), respectively, have been demonstrated to restrict the growth of *T. gondii* in vitro. Herein, we describe the first apicom-

* Corresponding author. Also affiliated with: Departments of Medicine, Pathology, Committees on Genetics and Immunology and The College. Tel.: +1-773-834-4152; fax: +1-773-834-3577.

E-mail addresses: rmcleod@midway.uchicago.edu (R. McLeod), d.rice@sheffield.ac.uk (D.W. Rice).

[1] Co-Corresponding author. Tel.: +44-114-222-4242; fax: 44-114-222-2850.

0020-7519/01/$20.00
PII: S0020-7519(01)00111-4

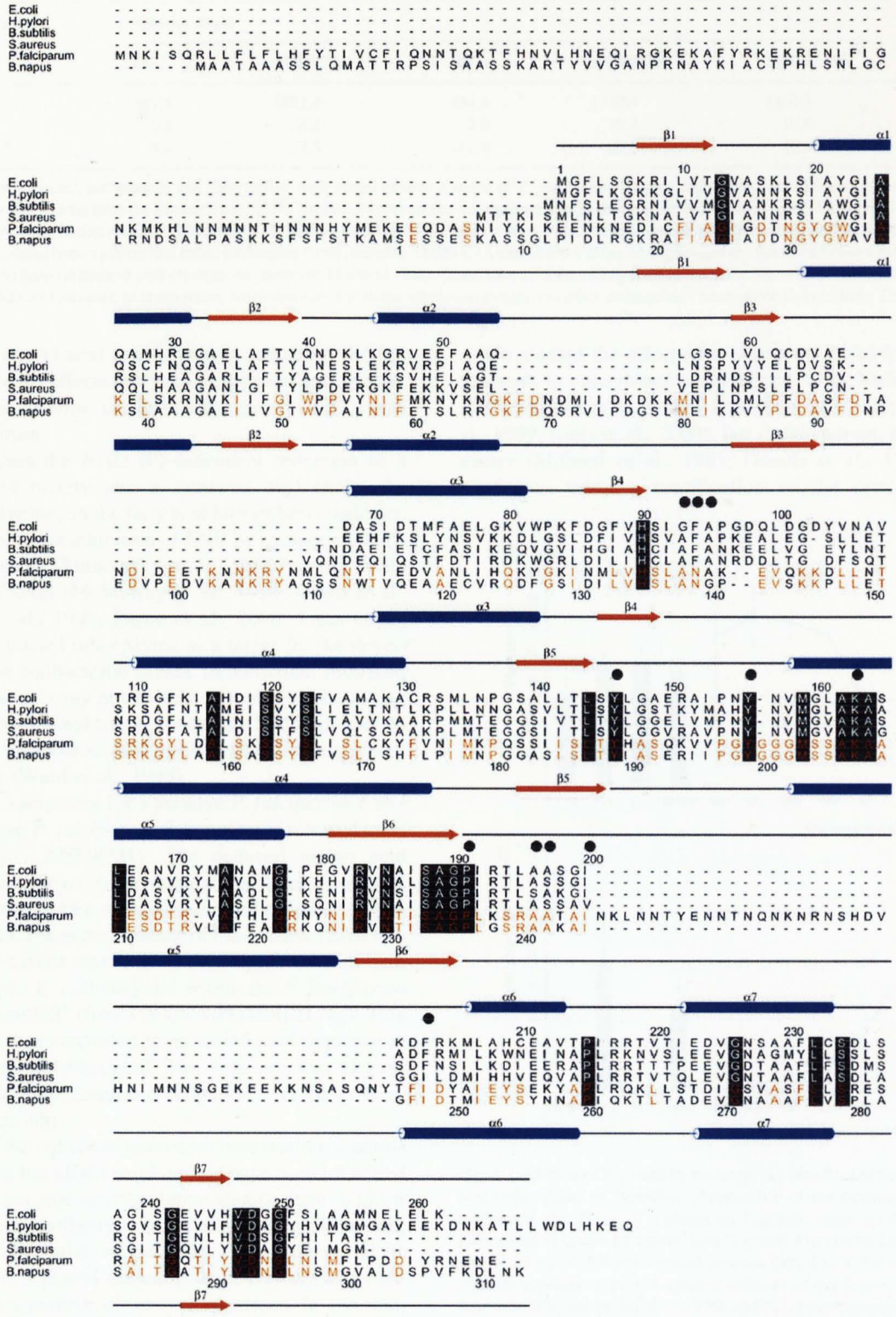


Fig. 1. A multiple structure-based sequence alignment of the enoyl reductases from *E. coli*, *H. pylori*, *B. subtilis*, *S. aureus*, *P. falciparum* and *B. napus*. The secondary structures and sequence numbers of the *E. coli* and *B. napus* enzymes are shown above and below the alignment, respectively. The residues which are completely conserved are in reverse type and those involved in triclosan binding are indicated with a black, filled circle above. Residues in red are conserved between *B. napus* and *P. falciparum* ENRs.

Table 1
IC50[a] of Triclosan, chloroquine, and mefloquine when cultured with *P falciparum* (ng/ml)

Antimicrobial Agents	Parasite strain				
	D6	TM90C2A	W2	TM90C2B	TM91C235
Triclosan	387.1	1891.4	154.4	1330.4	1800.5
Mefloquine	5.3	24.5	2.0	19.3	19.6
Chloroquine	3.8	57.3	162.4	82.7	46.1

[a] The activity of triclosan, mefloquine, and chloroquine were tested against a series of *P. falciparum* isolates and clones with differing susceptibilities to antimalarial drugs. D6, a clone from the African Sierra I/UNC isolate, is chloroquine and pyrimethamine susceptible; W2 is a clone of the Indochina I isolate and is chloroquine and pyrimethamine resistant. TM90C2A, TM90C2B, and TM91C235 are isolates from Thailand and all are chloroquine and mefloquine resistant. TM91C235 was isolated from a patient that failed mefloquine twice, whereas TM90-C2A and TM90-CB are admission and recrudescent isolates, respectively, of the first patient who failed treatment with atovaquone (alone) in Thailand. Subsequent susceptibility testing demonstrated that the recrudescent isolate (2B) was approximately 2000-fold resistant to atovaquone, when compared with the admission isolate and other atovaquone - susceptible isolates from Thailand.

plexan Fab I (enoyl acyl carrier protein reductase (ENR)) and investigate the effects of triclosan, a potent and specific inhibitor of this enzyme, on the in vitro growth of *T. gondii* and *P. falciparum*.

ENR catalyses the NAD (P)-dependent reduction of a trans-2,3 enoyl moiety into a saturated acyl chain, the second reductive step in the fatty acid biosynthesis pathway. Recent studies on the inhibition of ENR by compounds such as the diazaborines (Turnowsky et al., 1989; Baldock et al., 1996) and triclosan (McMurray et al., 1998; Heath et al., 1998; Levy et al., 1999; Payne et al., 2000; Jones et al., 2000) have validated this enzyme as a target for the development of new antibacterial agents. In particular, triclosan, which is found in many house-hold formulations including soaps, deodorants, hand lotion, toothpaste and impregnated into plastic as an anti-bacterial agent is an extremely potent ENR inhibitor (Ward et al., 1999).

We located a sequence for a putative *P. falciparum Fab I* on the aggregate *P. falciparum* chromosomes referred to as 'blob' (Asn no. AF338731). The deduced amino acid sequence and a multisequence alignment with representative enoyl acyl carrier protein reductases are shown in Fig. 1 (Asn no. AF338 731). The gene sequence of *Plasmodium ENR* was obtained with a Blast search using the sequences from both the *B. napus* and *E. coli* enzymes within the *P. falciparum* database 'PlasmoDB' (found at www.PlasmoDB.org). This sequence was then converted to an amino acid sequence at www.expasy.ch/tools/dna.html. The sequence was aligned using the 'Multiple Sequence Alignment' at http://searchlaun.bcm.tmc.edu.

Analysis of the pattern of sequence conservation confirms that this protein has all the residues that have been identified as crucial for enzyme activity. Interestingly, there is much greater sequence similarity with the plant enzyme than with the ENRs of bacterial origin. The *P. falciparum* ENR appears to have a plastid targeting sequence (Waller et al., 2000) and has a number of internal insertions. In addition, the *P. falciparum* protein has an extremely polar additional internal insertion for which no counterpart exists in any of the previously described ENRs. The function of the latter insertion remains to be determined.

We studied the effect of triclosan on *P. falciparum* and *T. gondii* in vitro (Mack and McLeod, 1984; Milhous et al., 1985; Oduola et al., 1988; Roberts et al., 1998; Zuther et al., 1999; Jones et al., 2000). For *P. falciparum*, the in vitro assays (Milhous et al., 1985; Oduola et al., 1988) were conducted using a modification of the semiautomated

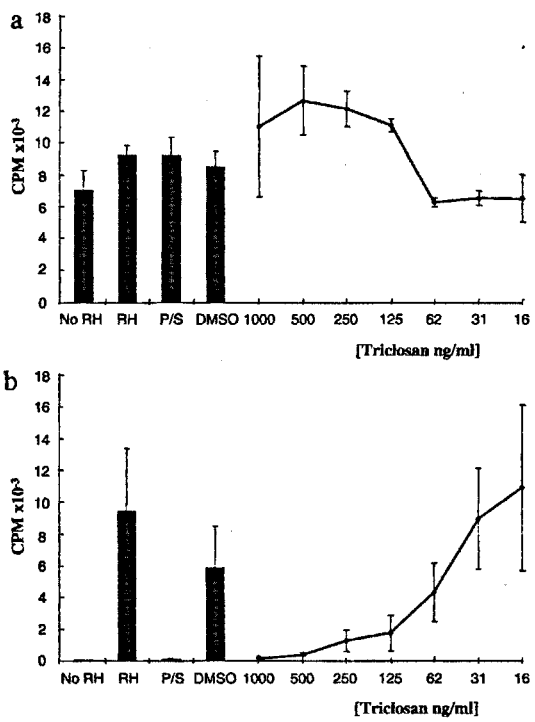


Fig. 2. Inhibition of *T. gondii* by triclosan. (a). No effect of triclosan on the host cells uptake of thymidine. Appearance of the monolayer also was unchanged. (b). Effect of triclosan on *T. gondii* uracil uptake. Triclosan reduces uracil uptake by intracellular *T. gondii* 4 days after infection. IC50 was ≅ 62 ng/ml. Effect increased between days 1 to 4. For example, in a separate experiment, for 125 ng/ml of triclosan on day 1, percentage inhibition was 20% and on day 4 was 72% and for pyrimethamine/sulfadiazine percentages of inhibition at these times were 63 and 100%, respectively. Abbreviations: RH, RH strain of *T. gondii* within fibroblasts; No RH, Control with fibroblasts alone; DMSO, fibroblasts with highest concentration of DMSO; P/S, fibroblasts, *T. gondii*, pyrimethamine and sulfadizine used as a positive control for the assay; CPM, counts per minute.

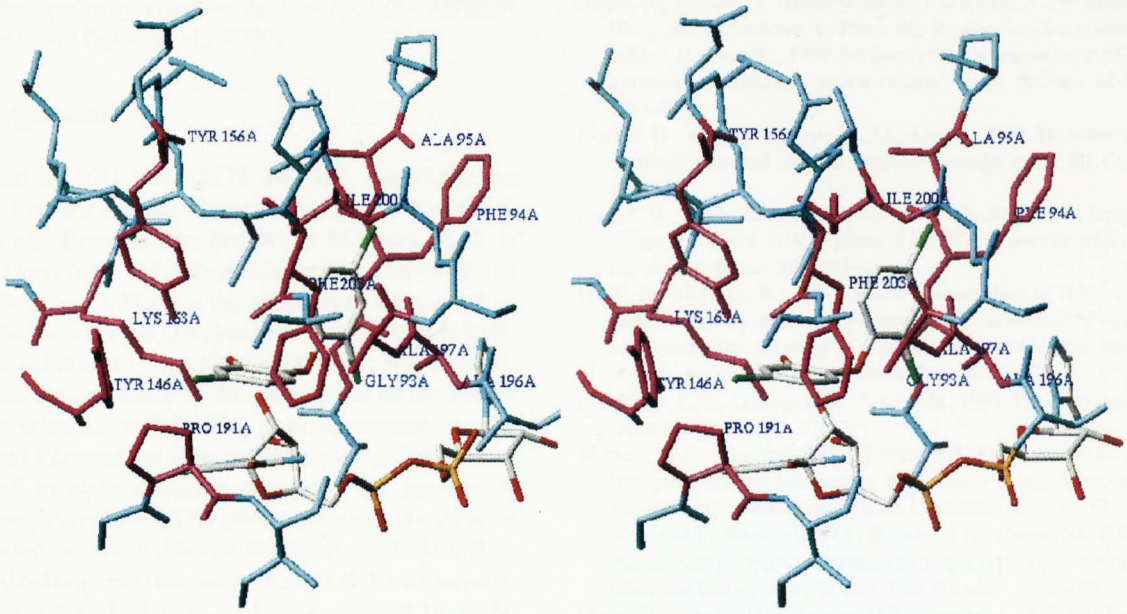


Fig. 3. A stereo view of the three dimensional arrangement of the atoms that form the binding pocket for triclosan, in *E. coli* enoyl reductase, with the 11 residues that have any atom within 4 Å of the inhibitor, labelled and coloured red. This is important in assigning the relative contributions made to the interaction with triclosan by the critical amino acids that are also present in the *P. falciparum* enzyme.

microdilution technique for assessing antifolate antagonists. Instead of dialysed human plasma, 10% Albumax I (Gibco BRL), a serum-free substitute, was used to supplement the RPMI 1640 medium. All test compounds were dissolved in DMSO and diluted 400-fold into complete medium before serial dilution over 11 concentrations. Incubation was at 37°C in 5% O_2, 5% CO_2 and 90% N_2 for 48 h. [^{3}H]-Hypoxanthine incorporation was measured as described previously (Milhous et al., 1985; Oduola et al., 1988). *P. falciparum* strain W2 is susceptible to mefloquine, but resistant to pyrimethamine, sulphadoxine and quinine and less susceptible to chloroquine than *P. falciparum* strain D6. Strain D6 is susceptible to pyrimethamine and sulphadoxine, but similar to *P. falciparum* strains TM90C2A and TM90C2B, and strain TM91C235 is less susceptible to mefloquine. The effect of triclosan on *P. falciparum* in vitro was studied with pyrimethamine sensitive and resistant organisms and those with varying sensitivities to chloroquine and mefloquine, simultaneously with studies of effect of chloroquine or mefloquine on these parasites (Table 1). Triclosan was effective against pyrimethamine resistant *P. falciparum* (W2) at low concentrations (IC50s of 150 ng/ml [triclosan] and 160 ng/ml [Chloroquine], respectively) (Table 1). Interestingly, the pattern of relative susceptibility of triclosan and mefloquine were identical. This similarity suggests that triclosan and mefloquine may share a common mechanism of influx or efflux, because such differences in transporters are believed to be the basis of the differences in susceptibility of malaria parasites to mefloquine although other mechanisms also are possible.

For *T. gondii*, growth inhibition was assessed over a 4-day period as described previously (Mack and McLeod, 1984; Roberts et al., 1998; Zuther et al., 1999) using human foreskin fibroblasts (HFF) infected with 10^5 tachyzoites of the RH strain of *T. gondii*. Triclosan also was effective against *T. gondii*, in nanomolar amounts (Fig. 2). IC50 was 62 ng/ml. There was no toxicity to host cells at these concentrations.

Analysis of the binding site for triclosan in *B. napus* and *E. coli* ENR shows that 11 residues have contacts less than 4 Å with one or more atoms of the triclosan (Fig. 3). Inspection of the sequence for *P. falciparum* ENR reveals that it shares sequence identity at each of these positions with either the sequence of the *B. napus* or *E. coli* enzymes providing a clear explanation for the inhibitory properties of this agent against *P. falciparum*.

The discovery and characterisation of an apicomplexan Fab I and discovery of triclosan as a lead compound provide means to rationally design novel inhibitory compounds with considerable promise. These findings may provide novel ways to counteract the increasing resistance of *Plasmodium* species to the current armoury of antimalarial agents and provide a new approach to the great need for additional, less toxic antimicrobial agents effective against *T. gondii*. We (Zuther et al., 1999) and others (Waller et al., 1998) also have identified other novel inhibitors of sequential enzymatic steps in the apicomplexan lipid synthesis pathway, that we predict will be synergistic with triclosan and other inhibitors of Fab I (Baldock et al., 1996). This also raises the exciting possibility of a rational basis for discovery of synergistic inhi-

bitors of this pathway effective against multiple different microorganisms (Payne et al., 2000).

Acknowledgements

Supported by NIH R01 44572, BBSRC, the Wellcome Trust, the Howard Hughes Foundation and the Koshland, Blackmon and Brennan families. Rima McLeod, M.D. is Jules and Doris Stein RPB Professor at The University of Chicago. Douglas G. Mack is the recipient of a Research to Prevent Blindness Career Development Award. John B Rafferty is a Royal Society Olga Kennard Fellow. We thank H. McGuinness for assistance in the preparation of the manuscript. The scientists and funding agencies comprising the international *Plasmodium* Genome Sequencing Consortium have agreed to make sequence data from the genome of *Plasmodium falciparum* (3D7) public before publication of the completed sequence. Sharen Bowman, Neil Hall, Bart Barrell, and colleagues at the Sanger Centre (UK) are sequencing chromosomes 1, 3–9 and 13, with support from the Wellcome Trust. A US consortium composed of Malcolm Gardner, Leda Cummings, Claire Fraser and colleagues at The Institute for Genome Research, along with Dan Carucci, Steven Hoffman and colleagues at the Navy Military Research Center, are sequencing chromosomes 2, 10, 11 and 14 (supported by NIAID/NIH, the Burroughs Wellcome Fund, and the U.S. Department of Defense). Richard Hyman, Eula Fung, Ron Davis and colleagues at the Stanford Genome Technology Center (USA) are sequencing chromosome 12, supported by the Burroughs Wellcome Fund. The *Plasmodium* Genome Database http://PlasmoDB.org (Nucleic Acids Res. 29, in press; 2001) is a collaborative effort supported by the Burroughs Wellcome Fund.

References

Baldock, C., Rafferty, J.B., Sdelnikova, S.E., Baker, P.J., Stitje, A.R., Slabas, A.R., Hawkes, T.R., Rice, D.W., 1996. A mechanism of drug action revealed by structural studies of enoyl reductase. Science 274, 2107–10.

DeRocher, A., Hage, C.B., Froehlich, J.E., Feagin, J.E., Parsons, M., 2000. Analysis of targeting sequences demonstrates that trafficking to the toxoplasma gondii plastid branches off the secretory system. J. Cell Sci. 113(Pt. 22), 3969–77.

Heath, R.J., Yu, Y.T., Shapiro, M.A., Olson, E., Rock, C.O., 1998. Broad-spectrum antimicrobial biocides target the Fab I component of fatty acid synthesis. J. Biol. Chem. 273, 30316–20.

Jomaa, H., Wiesner, J., Sanderbrand, S., Altincicek, B., Weidemeyer, C., Hintz, M., Turbachova, I., Eberl, M., Zeidler, J., Lichtenhaler, H.K., Soldati, D., Beck, E., 1999. Inhibtors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. Science 285 (5433), 1573–6.

Jones, R.D., Janpani, H., Newman, J.L., Lee, A., 2000. Triclosan: a review of effectiveness and safety in healthcare settings. Am. J. Inf. Control 28, 184–96.

Levy, C.W., Roujeinikova, A., Sedelnikova, S.E., Baker, P.J., Stuitje, A.R., Slabas, A.R., Rice, D.W., Rafferty, J.B., 1999. Molecular basis of triclosan activity. Nature 398, 383–4.

Mack, D., McLeod, R., 1984. A new micromethod to study effects of antimicrobial agents on *Toxoplasma gondii*: comparison of sulfadoxine and sulfadiazine and study of clindamycin, metronidazole, and cyclosporin A. Antimicrob. Agents Chemother. 26, 26–30.

McMurray, L.M., Oethinger, M., Levy, S.B., 1998. Triclosan targets lipid synthesis. Nature 394, 531–2.

Milhous, W.K., Weatherly, N.F., Bowdre, J.H., Desjardins, R.E., 1985. In vitro activities and mechanisms of resistance to anti-folates and antimalarial drugs. Antimicrob. Agents Chemother. 27, 525–30.

Oduola, A.M.J., Weatherly, N.J., Bowdre, J.H., Desjardins, R.E., 1988. *Plasmodium falciparum* – cloning by single erythrocyte micromanipulation and heterogeneity in vitro. Exp. Parasitol. 66, 86–95.

Payne, D.J., Wallis, N.G., Gentry, D.R., Rosenberg, M., 2000. The impact of genomics on novel antibacterial targets. Curr. Opin. Drug Disc. Dev. 3 (2), 177–90.

Roberts, F., Roberts, C.W., Johnson, J., Kyle, D.E., Krell, T., Coggins, J.R., Coombs, G.H., Milhous, W.K., Tzipori, S., Ferguson, D.J.P., Chakrabarti, D., McLeod, R., 1998. Evidence for the shikimate pathway in apicomplexan parasites. Nature 393, 801–5.

Roos, D.S., Crawford, M.J., Donald, R.G., Kissinger, J.C., Klimczak, L.J., Striepen, B., 1999. Origin, targeting, and function of the apicomplexan plastid. Curr. Opin. Microbiol. 2(4), 426–32.

Turnowsky, F., Fuchs, K., Jeschek, C., Hogenauer, G., 1989. env M genes of Salmonella typhimurium and *Escherichia coli*. J. Bacteriol. 171, 6555–65.

Waller, R.F., Reed, M.D., Cowman, A.F., McFadden, G.I., 2000. Protein trafficking to the plastid of Plasmodium falciparum is via the secretory pathway. EMBO J. 19 (8), 1794–802.

Waller, R.F., Kelling, P.J., Donald, R.G., Striepen, B., Handman, E., Lan-Unnasch, N., Cowman, A.F., Besra, G.S., Roos, D.S., McFadden, G.I., 1998. Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. Proc. Natl. Acad. Sci. USA 95 (21), 12352–7.

Ward, W.H.J., Holdgate, G.A., Rowesell, S., McLean, E.G., Pauptit, R.A., Clayton, E., Nichols, W.W., Colls, J.G., Minshull, C.A., Jude, D.A., Mistry, A., Timms, D., Camble, R., Hales, N.J., Britton, C.J., Taylor, I.W.F., 1999. Kinetic and structural characteristics of the inhibition of enoyl (acyl carrier protein) reductase by triclosan. Biochemistry 38, 12514–25.

Zuther, E., Johnson, J.J., Haselkorn, R., McLeod, R., Gornicki, P., 1999. Growth of *Toxoplasma gondii* is inhibited by aryloxyphenoxypropionate herbicides targeting acetyl-CoA carboxylase. Proc. Natl. Acad. Sci. USA 96, 13387–92.

Appendix II

Expression, purification and crystallization of the *Plasmodium falciparum* enoyl reductase

Muench, S. P., Rafferty, J. B., McLeod, R., Rice, D. W. & Prigge, S. T.

Acta Cryst. (2003) **59**, 1246-1248.

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Expression, purification and crystallization of the *Plasmodium falciparum* enoyl reductase

Stephen P. Muench,[a] John B. Rafferty,[a] Rima Mcleod,[b] David W. Rice[a] and Sean T. Prigge[c,d]*

[a]Krebs Institute for Biomolecular Research, Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield S10 2TN, England, [b]Department of Ophthalmology and Visual Sciences, The University of Chicago, Chicago, IL 60637, USA, [c]Department of Molecular Microbiology and Immunology, Malaria Research Institute, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA, and [d]Division of Experimental Therapeutics, Walter Reed Army Institute of Research, Silver Spring, MD 20910, USA

Correspondence e-mail: sprigge@jhsph.edu

Received 9 December 2002
Accepted 17 April 2003

New hope has been gained in the control of the malaria parasite *Plasmodium falciparum* (pf) with the discovery that the parasite contains a prokaryotic type II fatty-acid synthase (FAS). Since enzymes of this type are absent in humans, they are potential targets for the development of new drugs. The enoyl reductase enzyme (ENR) belonging to this pathway is of particular interest because it has been shown to be inhibited by submicromolar concentrations of the antimicrobial agent triclosan. Here, the development of an efficient overexpression system for pfENR as a fusion protein with maltose-binding protein, its simple one-step purification and cleavage from its fusion protein and crystallization under new conditions with bound NAD^+ cofactor and triclosan are reported. The crystals belong to the space group $P2_1$, with approximate unit-cell parameters $a = 88.2$, $b = 82.4$, $c = 94.8$ Å, $\beta = 90.77°$, and contain a tetramer in the asymmetric unit. Cryocooled crystals (100 K) diffracted to beyond 2.2 Å resolution at the Daresbury Synchrotron Radiation Source.

1. Introduction

The malaria parasite, *P. falciparum*, is responsible for over 500 million cases of malaria per year, killing more than 2 million people annually (Breman, 2001). Recently, a new area for drug development has come to light with the discovery of a prokaryotic type II fatty-acid synthase (FAS) in this parasite (Waller *et al.*, 1998; Waters *et al.*, 2002). Type II FAS systems, in which the enzymes of this pathway are found on separate polypeptides, are common in bacteria and plants (Magnuson *et al.*, 1993). In contrast, in the type I FAS systems found in metazoans the domains that carry out the equivalent catalytic steps are quite different in amino-acid sequence and reside on one or two much larger multifunctional proteins (Smith, 1994). Thus, the enzymes of type II FAS systems are prime targets for the discovery of drugs with potentially low toxicity in man. Previous studies have shown that enoyl reductase (ENR) is a key enzyme in type II systems and is the target of several classes of antimicrobial compounds, including the diazaborines and triclosan (Baldock *et al.*, 1996; Levy *et al.*, 1999, 2001). More recently, it has been shown that triclosan also inhibits pfENR with a K_i of 0.4 nM (Kapoor *et al.*, 2001) and inhibits the growth of *P. falciparum* in culture with an IC_{50} of 0.7 µM (Surolia & Surolia, 2001).

The sequence of the pfENR shows significant sequence similarity to enoyl reductases from other species with, for example, 47% identity to the enzyme from *Brassica napus*. However, the pfENR sequence differs from all other ENR sequences determined to date in that it contains a 43 amino-acid insert in a loop which in the ENRs from other species is known to undergo substantial conformation change on inhibitor (and presumably substrate) binding (Levy *et al.*, 1999, 2001; Qiu *et al.*, 1999). The structure of this loop is therefore of particular interest in the exploitation of pfENR for drug discovery owing to its possible interactions with substrate, cofactor or inhibitors. Recently, the structure of pfENR has been solved to 2.4 Å (Perozzo *et al.*, 2002), but unfortunately no density could be seen for the 43 amino-acid insert owing to disorder in this part of the structure, thus leaving an important piece of the structure undetermined. In order to address this problem, we have developed a new expression system for this enzyme. We report the successful expression and purification of pfENR to give crystals of a new form which diffract to high resolution.

2. Materials and methods

2.1. Expression and purification of pfENR using the pMALc2x vector

The coding sequence of *P. falciparum* ENR was amplified from gDNA of the 3D7 strain of *P. falciparum* using PfuTurbo polymerase (Stratagene). The primers (forward) 5′-GGTGGTGAATTCTCAAACATAAACAAAATTAAAGAAG-3′ and (reverse) 5′-GGTGGT-

GTCGACTTATTCATTTTCATTGCGATATATATC-3′ were used to amplify nucleotides encoding amino acids 85–432 and to introduce a proximal *Eco*RI and distal *Sal*I site (bold) in the PCR product. Nucleotides encoding the amino-terminal 84 residues of pfENR were excluded because this region is principally composed of a signal peptide and an organellar transit peptide. The resulting amplicon was digested with *Eco*RI and *Sal*I and ligated into the pMALc2x vector (New England Biolabs). The resulting pSTP6 was transformed into BL21 Star(DE3) cells (Invitrogen). These cells were cotransformed with the pRIL plasmid isolated from BL21-CodonPlus(DE3) cells (Stratagene) and used for the expression of MBP-ENR fusion protein. Cells were grown in LB medium at 310 K to an optical density at 600 nm of 0.8 and then induced by addition of IPTG to a final concentration of 0.4 m*M*. The culture was maintained in shaker flasks at 293 K for 12 h and then harvested by centrifugation.

Cells were resuspended in lysis buffer [20 m*M* sodium/potassium phosphate pH 7.5, 1 mg ml^{-1} lysozyme (Sigma), 2.5 µg ml^{-1} DNAse I (Sigma), 200 m*M* NaCl] and sonicated. Cell lysate was clarified by centrifugation and applied to a 10 ml amylose column (New England Biolabs) for affinity purification (Fig. 1*a*, lane 2). Purified MBP-ENR fusion protein was digested with factor Xa (New England Biolabs) at a ratio of 1 mg factor Xa per 500 mg of fusion protein in the presence of 1 m*M* calcium chloride at 277 K (Fig. 1*a*, lane 3). The reaction mixture was desalted with a HiPrep 26/10 desalting column (Pharmacia) and applied to an SP Sepharose cation-exchange column (Pharmacia). Column fractions containing pure pfENR protein were pooled for further analysis.

2.2. Expression and purification of pfENR using the pMALcHT vector

A second construct of pfENR also containing residues 85–432 was designed for *in vivo* cleavage by the TEV (tobacco etch virus) protease. The amplicon described above was ligated into a modified version of the pMALc2x vector (pMALcHT) in which the linker region was altered to contain nucleotides encoding a TEV (tobacco etch virus) protease cleavage site followed by a six-histidine tag (Fig. 1*b*). The resulting ligation product, pSTP7, was transformed into BL21 Star(DE3) cells (Invitrogen). These cells were cotransformed with the pRIL plasmid from BL21-CodonPlus(DE3) cells (Stratagene) and plasmid (pKM586) encoding the TEV protease (Kapust & Waugh, 2000). Cells were grown, harvested and lysed as above. Cell lysate was clarified by centrifugation and applied to a 5 ml HiTrap Chelating HP column (Pharmacia). Column fractions containing cut pfENR were desalted with a HiPrep 26/10 desalting column (Pharmacia) and loaded onto a 5 ml HiTrap SP Fast Flow column (Pharmacia). Pure pfENR (Fig. 1*a*, lane 4) was then concentrated to 12 mg ml^{-1} for crystallization trials.

2.3. Crystallization of pfENR and data collection.

Crystals of pfENR were grown using the hanging-drop vapour-diffusion technique by mixing 2.5 µl of the protein solution (12 mg ml^{-1} pfENR in 20 m*M* sodium/potassium phosphate pH 8.0, 150 m*M* NaCl, 5 µ*M* NAD^+ and 6 µ*M* triclosan) with 2.5 µl of the reservoir solution at 290 K. Initial screening of crystallization conditions was conducted using Crystal Screen 1, Crystal Screen 2 and PEG/Ion Screen (Hampton Research), of which PEG/Ion screen solution 11 [20%(*w*/*v*) PEG 3350 and 200 m*M* KI] produced the best-quality crystals. This aforementioned screen was optimized, changing both the pH and precipitant concentrations, to achieve an optimal reservoir solution composed of 19.5%(*w*/*v*) PEG 3350 and 230 m*M* KI. This condition produced good-quality crystals which took four to five weeks to reach dimensions of 0.15 × 0.10 × 0.10 mm. X-ray analysis at the Daresbury Synchrotron Radiation Source (SRS) of crystals frozen at 100 K using 20% glycerol as a cryoprotectant showed that they diffracted to beyond 2.2 Å. Rotation images were collected with 1° oscillation width and 1 min exposure times on an ADSC Quantum 4 detector at station 14.1. These crystallization conditions were very different to those previously reported by Perozzo *et al.* (2002), which were based on ammonium sulfate. They also give crystals whose diffraction properties appear to be superior, giving a lower R_{merge} (0.095 compared with 0.123) at a higher resolution of 2.2 *versus* 2.4 Å.

3. Results and discussion

We were able to efficiently overexpress pfENR as a fusion protein with the maltose-binding protein (MBP) using the pMALc2x expression vector. The resulting MBP-ENR fusion protein was readily purified through affinity chromatography (Fig. 1*a*, lane 2). However, endoproteolytic digestion of pure

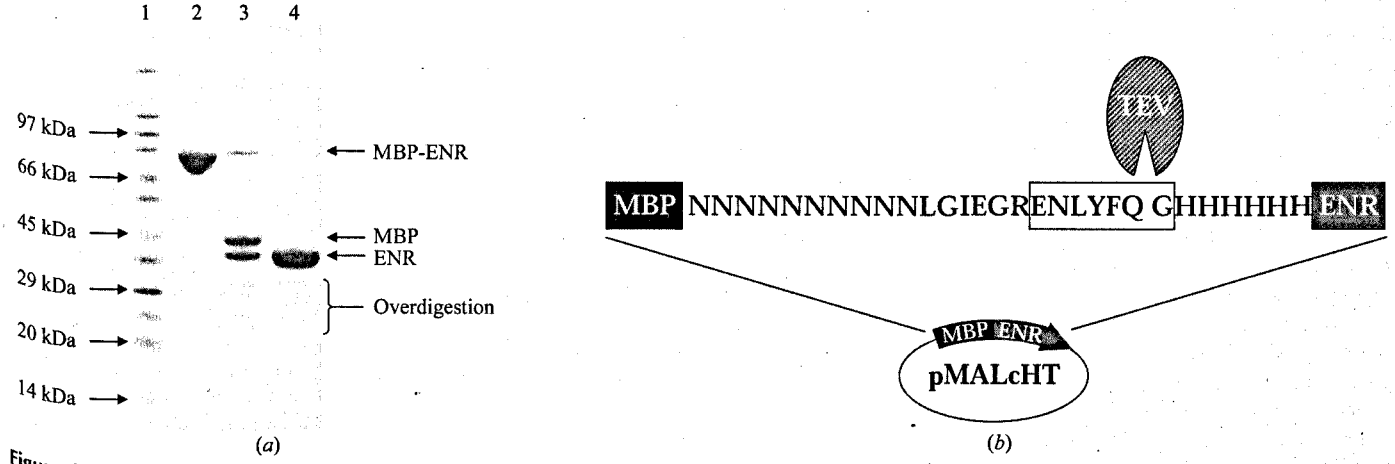


Figure 1
(*a*) SDS–PAGE of pfENR purification. SigmaMarker wide-range molecular-weight markers are shown in lane 1. Pure MBP-ENR fusion protein produced by the pMALc2x vector (lane 2) and factor Xa-digested fusion protein (lane 3) are shown. Lane 4 shows pfENR produced by *in vivo* cleavage using the pMALcHT vector. (*b*) Schematic representation of the pMALcHT vector. Amino acids in the linker region between the maltose-binding protein (MBP) and pfENR (ENR) are shown using single-letter abbreviations. The seven-residue TEV protease recognition site is boxed and a gap indicates the protease cleavage site.

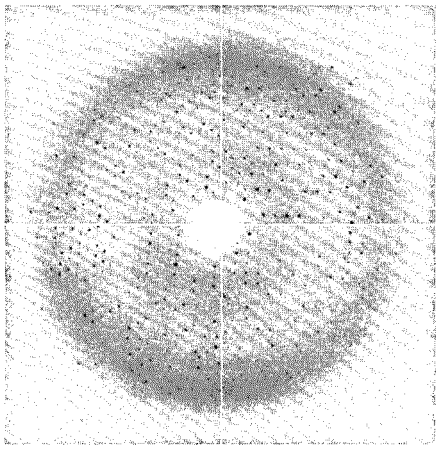


Figure 2
A representative 1° oscillation image of data collected from a crystal of pfENR complexed with NAD^+ and triclosan on a Quantum Q4 CCD detector on station 14.1 at the SRS Daresbury Laboratory. The resolution at the edge of the image corresponds to a resolution of 2.2 Å.

MBP-ENR with factor Xa protease yielded overdigestion products under conditions that did not completely digest all of the MBP-ENR fusion protein (Fig. 1*a*, lane 3). Better results were obtained using *in vivo* cleavage of the fusion protein by the tobacco etch virus (TEV) protease. The linker region of the pMALc2x vector was modified to contain a TEV protease cut site followed by a six-histidine tag, generating the pMALcHT expression vector (Fig. 1*b*). This vector was cotransformed into *Escherichia coli* along with a plasmid encoding the TEV protease (pKM586) and a plasmid encoding three rare tRNAs (pRIL). A long low-temperature induction period during the expression of the MBP-ENR fusion protein also served to expose the fusion protein to digestion by constitutively expressed TEV protease. The resulting pfENR digestion product could then be purified *via* a histidine tag exposed by cleavage. Since pfENR is a multimer (see below), uncut MBP-ENR will tend to copurify with cut pfENR. We were able to remove trace amounts of uncut MBP-ENR by using a subtractive amylose column.

The data collected at the Daresbury Synchrotron Radiation Source (SRS) were processed and scaled using the *DENZO*/*SCALEPACK* package (Otwinowski & Minor, 1997). Analysis of the diffraction data using the autoindexing routine in the program *DENZO* showed that the crystals belong to the primitive monoclinic system,

Table 1
Data collection and processing for pfENR crystals.

Values in parentheses refer to data in the highest resolution shell.

Space group	$P2_1$
Wavelength used (Å)	0.9600
Resolution range (Å)	0–2.18 (2.26–2.18)
Unique reflections	77321 (7521)
Multiplicity	3.1 (3.0)
Completeness (%)	97.2 (94.7)
$I/\sigma(I) > 3$ (%)	72.6 (49.6)
R_{merge}† (%)	0.095 (0.395)

† $R_{merge} = \sum_{hkl} |I_i - I_m| / \sum_{hkl} I_i$, where I_i and I_m are the observed intensity and mean intensity of related reflections, respectively.

with unit-cell parameters $a = 88.18$, $b = 82.37$, $c = 94.82$ Å, $\alpha = \gamma = 90$, $\beta = 90.77°$. Reflections were observed along the $0k0$ axis (b^*) only where they satisfied the condition $k = 2n$, indicating that the crystals belonged to space group $P2_1$. Significant reflections were observed to the edge of the image-plate detector (Fig. 2) and a good-quality data set was collected to 2.2 Å. Data-collection and processing statistics can be found in Table 1. Gel-filtration studies indicate that pfENR is a tetramer in solution like the enzymes from *E. coli* and *B. napus* (Baldock *et al.*, 1998; Rafferty *et al.*, 1995); consideration of the possible values of V_M suggest that the asymmetric unit contains a complete tetramer with a V_M of 2.2 $Å^3$ Da^{-1}, which is within the range observed for protein crystals (Matthews, 1974).

A preliminary attempt to solve the structure using molecular replacement with the *AMoRe* program (Navaza, 1994) as implemented within the *CCP*4 package (Collaborative Computational Project, Number 4, 1994) has been carried out using the coordinates of *B. napus* ENR (PDB code 1d70) as a search model. A promising solution was obtained and ultimately it is hoped that refinement of the structure to high resolution may provide insights into the details of inhibitor binding and allow us to see any important interactions with the additional loop present in the pfENR.

This work was supported by the Wellcome Trust and the BBRSC. The Krebs Institute is a designated BBRSC Biomolecular Science Centre and a member of the North of England Structural Biology Centre. This work was also supported by the USAMRMC (US Army Medical Research and Materiel Command). Preliminary sequence data were obtained from the Plasmodium Genome Database, http://PlasmoDB.org, which is a collaborative effort supported by the Burroughs Wellcome Fund. JBR is the Royal Society Olga Kennard Fellow and RML is the Jules and Dorris Stein RPB Professor at the University of Chicago, USA.

References

Baldock, C., Rafferty, J. B., Sedelnikova, S. E., Baker, P. J., Stuitje, A. R., Slabas, A. R., Hawkes, T. R. & Rice, D. W. (1996). *Science*, **274**, 2107–2110.
Baldock, C., Rafferty, J. B., Stuitje, A. R., Slabas, A. R. & Rice, D. W. (1998). *J. Mol. Biol.* **284**, 1529–1546.
Breman, J. G. (2001). *Am. J. Trop. Med. Hyg.* **64**, 1–11.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Kapoor, M., Dar, M. J., Surolia, A. & Surolia, N. (2001). *Biochem. Biophys. Res. Commun.* **289**, 832–837.
Kapust, R. B. & Waugh, D. S. (2000). *Protein Expr. Purif.* **19**, 312–318.
Levy, C. W., Baldock, C., Wallace, A. J., Sedelnikova, S., Viner, R. C., Clough, J. M., Stuitje, A. R., Slabas, A. R., Rice, D. W. & Rafferty, J. B. (2001). *J. Mol. Biol.* **309**, 171–180.
Levy, C. W., Roujeinikova, A., Sedelnikova, S., Baker, P. J., Stuitje, A. R., Slabas, A. R., Rice, D. W. & Rafferty, J. B. (1999). *Nature (London)*, **398**, 383–384.
Magnuson, K., Jackowski, S., Rock, C. O. & Cronan, J. E. Jr (1993). *Microbiol. Rev.* **57**, 522–542.
Matthews, B. W. (1974). *J. Mol. Biol.* **82**, 513–526.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perozzo, R., Kuo, M., Sidhu, A. S., Valiyaveetil, J. T., Bittman, R., Jacobs, W. R. Jr, Fidock, D. A. & Sacchettini, J. C. (2002). *J. Biol. Chem.* **277**, 13106–13114.
Qiu, X., Janson, C. A., Court, R. I., Smyth, M. G., Payne, D. J. & Abdel-Meguid, S. S. (1999). *Protein Sci.* **8**, 2529–2532.
Rafferty, J. B., Simon, J. W., Baldock, C., Artymiuk, P. J., Baker, P. J., Stuitje, A. R., Slabas, A. R. & Rice, D. W. (1995). *Structure*, **3**, 927–938.
Smith, S. (1994). *FASEB J.* **8**, 1248–59.
Surolia, N. & Surolia, A. (2001). *Nature Med.* **7**, 167–173.
Waller, R. F., Keeling, P. J., Donald, R. G., Striepen, B., Handman, E., Lang-Unnasch, N., Cowman, A. F., Besra, G. S., Roos, D. S. & McFadden, G. I. (1998). *Proc. Natl Acad. Sci. USA*, **95**, 12352–12357.
Waters, N. C., Kopydlowski, K. M., Guszczynski, T., Wei, L., Sellers, P., Ferlan, J. T., Lee, P. J., Li, Z., Woodard, C. L., Shallom, S., Gardner, M. J. & Prigge, S. T. (2002). *Mol. Biochem. Parasitol.* **123**, 85–94.

Appendix III

Delivery of antimicrobials into parasites

Samuel, B. U., Hearn, B., Mack, D., Wender, P., Rothbard, J., Kirisits, M. J., Mui, E., Wernimont, S., Roberts, C. W., Muench, S. P., Rice, D. W., Prigge, S. T., Law, A. B. & McLeod, R.

Proc Natl Acad Sci U S A. (2003) **100**, 14281-14286.

Delivery of antimicrobials into parasites

B. U. Samuel*†, B. Hearn†‡, D. Mack*†, P. Wender‡§¶, J. Rothbard‡§, M. J. Kirisits*, E. Mui*, S. Wernimont*, C. W. Roberts‖, S. P. Muench**, D. W. Rice**, S. T. Prigge††, A. B. Law††, and R. McLeod*¶

*Department of Visual Sciences, University of Chicago, 5841 South Maryland, AMB S-208, Chicago, IL 60637; ‡Department of Chemistry, Stanford University, Stanford, CA 94305; §Cell Gate Inc., 552 Del Rey Avenue, Sunnyvale, CA 94086; ‖Department of Immunology, University of Strathclyde, G4 ONR Glasgow, Scotland; **Department of Molecular Biology, Sheffield University, Firth Court, Western Bank, Sheffield S10 2TN, United Kingdom; and ††Malaria Research Institute, The Johns Hopkins University School of Public Health, Baltimore, MD 21205

Contributed by P. Wender, September 25, 2003

To eliminate apicomplexan parasites, inhibitory compounds must cross host cell, parasitophorous vacuole, and parasite membranes and cyst walls, making delivery challenging. Here, we show that short oligomers of arginine enter *Toxoplasma gondii* tachyzoites and encysted bradyzoites. Triclosan, which inhibits enoyl-ACP reductase (ENR), conjugated to arginine oligomers enters extracellular tachyzoites, host cells, tachyzoites inside parasitophorous vacuoles within host cells, extracellular bradyzoites, and bradyzoites within cysts. We identify, clone, and sequence *T. gondii enr* and produce and characterize enzymatically active, recombinant ENR. This enzyme has the requisite amino acids to bind triclosan. Triclosan released after conjugation to octaarginine via a readily hydrolyzable ester linkage inhibits ENR activity, tachyzoites *in vitro*, and tachyzoites in mice. Delivery of an inhibitor to a microorganism via conjugation to octaarginine provides an approach to transporting antimicrobials and other small molecules to sequestered parasites, a model system to characterize transport across multiple membrane barriers and structures, a widely applicable paradigm for treatment of active and encysted apicomplexan and other infections, and a generic proof of principle for a mechanism of medicine delivery.

The vast majority of interesting small molecules will never be used for therapeutics because they cannot traverse biologic barriers to reach their targets. A special challenge for the development of antimicrobials effective against apicomplexan parasites is delivery. Effective agents have to gain entry to host cells, cross the parasitophorous vacuole, and enter parasites and their specialized organelles. This is potentially more difficult with *Toxoplasma gondii* bradyzoites, which reside in cysts composed partly of host and partly of parasite constituents. The studies described herein were performed to determine whether short oligomers of arginine (Fig. 1*A*) could provide a solution to these challenges. Arginine oligomers linked to immunomodulatory and antitumor compounds can actively transport such compounds across a variety of epithelial barriers, including skin, lung, and eye (1–4). Oligoarginine-medicine conjugates have been administered safely to experimental animals and humans (ref. 1 and unpublished data).

Better approaches for treating apicomplexan parasite infections (e.g., toxoplasmosis and malaria) are needed. For example, *T. gondii* bradyzoites infect 30–50% of people throughout the world, but no currently used medicines eradicate this slowly growing, encysted, latent parasite that causes chronic, lifelong infections in brain and eye with recrudescence causing severe illness (5). Available antimicrobial agents to eliminate rapidly growing *T. gondii* tachyzoites, which cause tissue destruction during active infection, are limited by toxicity and allergy (5).

Enoyl-ACP reductase (ENR) is a key enzyme active in fatty acid synthesis (6). Bacterial and apicomplexan fatty acid synthesis occurs by the type II pathway (6–10) (Fig. 5, which is published as supporting information on the PNAS web site) and depends on monofunctional polypeptides, including ENR. In contrast, mammalian fatty acid synthesis occurs by the type I pathway in which the key enzymes are present on a polyfunctional single polypeptide (6–10). In apicomplexans, type II fatty acid synthesis enzymes are in a plastid organelle (6, 8).

Some monofunctional type II enzymes had been identified in *Plasmodia* and *T. gondii* (6–10), but *T. gondii* ENR (TgENR) had not been identified. The antimicrobial agent 5-chloro-2-[2,4-dichlorophenoxy]phenol (triclosan) had been found to inhibit bacterial ENRs (11). We (6) and Surolia and Surolia (9) had found that μM concentrations of triclosan dissolved in DMSO inhibit replication of apicomplexan parasites, presumably by binding to, and inhibiting, apicomplexan ENR. Triclosan was active only when solubilized in DMSO (6, 9). Structures of triclosan, oligomers, and conjugates of oligoarginines are shown in Fig. 1 (see also Fig. 5).

Materials and Methods

Short Arginine Oligomers, Conjugates, and Controls. Peptides were synthesized by using solid-phase techniques and commercially available fluorenylmethoxycarbonyl amino acids, resins, and reagents with an Applied Biosystems 433 peptide synthesizer. Fastmoc cycles were used with *O*-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HATU). Peptides and conjugates were cleaved from resin by using 95% trifluoroacetic acid and 5% triisopropylsilane for 24 h. Peptides were subsequently filtered from resin, precipitated with diethyl ether, purified by using HPLC reverse-phase columns, and characterized by using ^{1}H NMR and electrospray MS.

For synthesis of triclosan-Orn(FITC)-Arg_n-$CONH_2$, **4** and **5**, triclosan was reacted with α-bromoacetic acid to provide the desired α-substituted acetic acid product. Attachment of the resin-bound carrier was accomplished by reaction of this acetic acid product with the appropriate oligomer of L-arginine(2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl, Pbf) containing an N-terminal L-ornithine(methyltrityl). After cleavage of the side-chain methyltrityl protecting group with 1% trifluoroacetic acid in dichloromethane, the conjugates were labeled with FITC in the presence of diisopropylethylamine (DIPEA). The conjugates were then deprotected and cleaved from the resin by using the previously described protocol. For the synthesis of triclosan-Orn(tetramethylrhodamine isothiocyanate)-Arg_8-$CONH_2$, **ent-6**, tetramethylrhodamine isothiocyanate was reacted in the solution phase with triclosan-D-Orn-D-Arg_8-$CONH_2$ in the presence of DIPEA. Triclosan was reacted with glutaric anhydride (GA) in the presence of DIPEA to provide the desired glutaric acid product, i.e., triclosan-GA-Arg_n-$CONH_2$, **7** and **8**. Attachment of the carrier was accomplished by the solution-phase reaction with the appropriate oligomer of D- or L-arginine. For the synthesis of triclosan-acetic

Abbreviations: ENR, enoyl-ACP reductase; TgENR, *Toxoplasma gondii* ENR; Pbf, 2,2,4,6,7-pentamethyldihydrobenzofuran-5-sulfonyl.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. AY372520).

†B.U.S., B.H., and D.M. contributed equally to this work.

¶To whom correspondence should be addressed. E-mail: wenderp@leland.stanford.edu or rmcleod@midway.uchicago.edu.

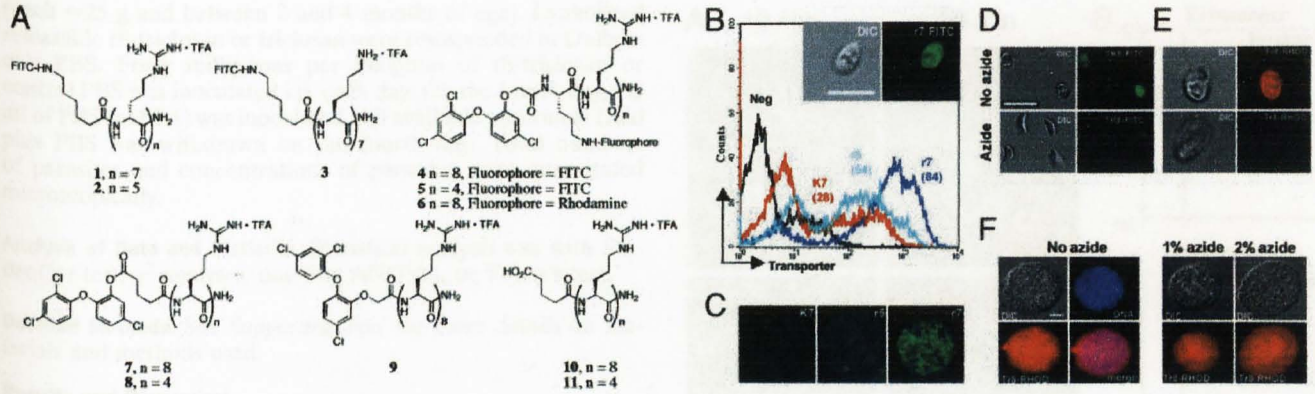


Fig. 1. Schematic representations of oligomers and conjugates, uptake of short oligomers of arginine and lysine by *T. gondii* tachyzoites, encysted bradyzoites, and effect of azide on uptake of arginine triclosan conjugates by parasites. (*A*) Structures of short oligomers of arginine and conjugates used in these studies. Releasable conjugates use readily hydrolyzable ester linkages for the attachment of triclosan to the oligoarginine carrier (structures **7** and **8**). Nonreleasable triclosan conjugates use nonhydrolyzable ether bonds for the attachment of triclosan (structures **4**, **5**, **6**, and **9**). Control compounds **10** and **11** are the hydrolysis products of releasable conjugates **7** and **8**, respectively. Ent designates the dextro isomer. Synthetic details are in *Materials and Methods*. (*B*) Comparative flow cytometry to measure labeling of *T. gondii* tachyzoites incubated with FITC-conjugated polyamino acid compounds. Numbers in parentheses indicate percentage of positive cells. (*Inset*) Fluorescence and differential interference contrast (DIC) microscopy assays of uptake of r7 FITC, compound **1**, by *T. gondii* tachyzoites. (*C*) Fluorescence images showing uptake of FITC-conjugated polyamino acid compounds by *T. gondii* tissue cysts. (Scale bar = 10 μm.) (*D–F*) Fluorescein and DIC microscopy images showing inhibition of uptake of arginine triclosan conjugates by azide in *T. gondii*. (*D*) Tachyzoites. (*E*) Excysted bradyzoites. (*F*) Uptake of r8 triclosan (compound **7**) and less inhibition in isolated brain cysts. The arginine carrier was conjugated to fluorophore and triclosan in a nonreleasable manner (**4** and **ent-6**). (Scale bar = 10 μm.)

acid-Arg_8-$CONH_2$, **9**, the previously described α-substituted acetic acid product was attached to the carrier by reaction with the resin-bound octa-L-arginine(Pbf). The conjugate was subsequently cleaved from the resin, and the Pbf sulfonamides were removed. For the synthesis of GA-Arg_n-$CONH_2$, **10** and **11**, the resin-bound oligomer of L-arginine(Pbf) was reacted with GA in the presence of DIPEA, followed by cleavage of the resin and Pbf deprotection.

The half-life of conjugate **7** decomposition in PBS was determined by using HPLC to monitor the change in the ratio of conjugate-to-internal standards (benzoic acid) over time.

***T. gondii* Parasites.** Tachyzoites of the RH strain were maintained and used in experiments as described (12, 13). GFP-, plastid-labeled RH strain parasites were kindly provided by B. Streipen and D. Roos (University of Pennsylvania, Philadelphia). Production and isolation of *T. gondii* cysts were as described (14).

Studies on the Uptake of Short Arginine and Lysine Oligomers into *T. gondii*. Live cells incubated with conjugates were examined by using a Zeiss Axiovert inverted fluorescence microscope or Zeiss Axioplan equipped with cooled charge-coupled device camera. For flow cytometry, parasites and cells were incubated in PBS/2% FCS containing 12.5 μM of each oligomer at 18°C for 10 min, washed in cold PBS three times, resuspended in PBS, and then analyzed by flow cytometry with a FACScan and CELL QUEST software. Before analysis, intracellular parasites were exposed to 12.5 μM conjugate for 10 or 30 min.

Treatment with Azide. Parasites were preincubated for 30 min with 0.5% azide (tachyzoites, bradyzoites) or 0.5–2% azide (cysts), all in 2% FCS/PBS buffer, before the addition of 12.5 μM conjugate.

Cloning and Sequencing of TgENR. A cDNA library was screened to identify and characterize the TgENR gene as described (15, 16) (see *Supporting Text*, which is published as supporting information on the PNAS web site).

Expression and Purification of Recombinant TgENR. A construct of TgENR containing residues 103–417 (and lacking the putative signal and transit peptides) was designed for *in vivo* cleavage by tobacco etch virus (TEV) protease. Amplified DNA encoding residues 103–417 of TgENR was ligated into a modified version of the pMALc2x vector (pMALcHT) in which the linker region was altered to contain nucleotides encoding a TEV protease cleavage site followed by a six-histidine tag (17). The resulting ligation product, pSTP8, was transformed into BL21 Star (DE3) cells (Invitrogen). These cells were cotransformed with the pRIL plasmid from BL21-CodonPlus (DE3) cells (Stratagene) and a plasmid (pKM586) encoding the TEV protease (18). Cells were grown, harvested, and lysed as described (17). TgENR was purified by using metal chelate and anion exchange chromatography (see Fig. 3*B*).

Effect of r8-Triclosan on Recombinant TgENR. The activity of TgENR was assayed by using crotonyl-CoA as a substrate and monitoring the consumption of the NADH cofactor. Inhibition by r8-triclosan over a 24-h period was measured by incubating TgENR at 37°C with different concentrations of r8-triclosan. Aliquots were removed at 0, 12, and 24 h and assayed for TgENR activity. Briefly, 10 μl of 1 mM NADH and 10 μl of 1 mM crotonyl-CoA were added to 80 μl of incubated TgENR/r8-triclosan, and the reaction was followed for 120 s in a Beckman DU-640 spectrophotometer. Final concentration of TgENR was 11.6 nM, and the final concentrations of r8-triclosan were 2 μM, 0.4 μM, 80 nM, 16 nM, 3.2 nM, 640 pM, and 128 pM. TgENR was inhibited by r8-triclosan with an IC_{50} value >1 μM at 0 h, an IC_{50} value = 40 nM ± 10 nM at 12 h, and an IC_{50} value <16 nM at 24 h (see Fig. 3*C*). The IC_{50} value at 24 h is an upper limit because inhibition cannot be properly measured near the concentration of TgENR used in these assays (11.6 nM).

Assays to Assess Inhibition of *T. gondii* Tachyzoite Growth and Effect of Antimicrobial Agents on Host Cells *in Vitro*. These studies were performed as described (12, 13).

Effect of r8-Triclosan on Tachyzoites *in Vivo*. One thousand RH strain parasites were inoculated i.p. into five female SW mice

(each ≈25 g and between 2 and 4 months of age). Lypholized releasable r8-triclosan or triclosan were resuspended in Dulbecco's PBS. Forty milligrams per kilogram of r8-triclosan or control PBS was inoculated i.p. each day. On the fourth day, 1.5 ml of PBS (pH 7.4) was inoculated. All available peritoneal fluid plus PBS was withdrawn on the fourth day. Total numbers of parasites and concentrations of parasites were quantitated microscopically.

Analysis of Data and Statistics. Statistical analysis was with Student's t test, χ^2 analyses, one-way ANOVA, or Tukey's test.

Detailed Methods. See *Supporting Text* for more details on materials and methods used.

Results and Discussion

Initially, uptake of fluoresceinated hepta-D-arginine (r7, compound **1**, Fig. 1*A*) and control compounds penta-D-arginine (r5, compound **2**, Fig. 1*A*) and hepta-L-lysine (K7, compound **3**, Fig. 1*A*) into extracellular and intracellular tachyzoites, and extracellular and encysted bradyzoites, was studied by using deconvolution fluorescence microscopy and flow cytometry (Fig. 1 *B–F*). R designates levo isomer, r designates dextro isomer, and the number after r indicates the number of arginines in the compound. Short arginine oligomers were found to cross all membranes and to enter each life cycle stage and compartment (Fig. 1 *B–F*). Deconvolution fluorescence microscopy revealed that r7 facilitated uptake of fluoresceinated compounds into tachyzoites (Fig. 1*B Inset*). Flow cytometric analysis (Fig. 1*B*) demonstrated that uptake of r7 was greater than that of r5 or K7 (seven arginines ≫ five arginines or seven lysines). This finding was consistent with earlier studies demonstrating that longer oligomers of arginines (seven or eight) entered certain eukaryotic cells better than shorter oligomers (one to four). Deconvolution microscopy demonstrated that short arginine oligomers could cross cyst walls and enter encysted bradyzoites (Fig. 1*C* and Movie 1, which is published as supporting information on the PNAS web site). Uptake of short arginine oligomers into tachyzoites and bradyzoites was very rapid (seconds), and in bradyzoites, the fluoresceinated compound moved from the cytoplasm to the nucleus over the first day (data not shown).

Floresceinated, nonreleasable triclosan R8, conjugate **4**, then was studied by microscopic analyses. The conjugated triclosan was taken up by extracellular tachyzoites and bradyzoites within cysts ≈10 s after addition (Fig. 1 *D–F*). Treatment with azide blocked uptake of short arginine oligomers linked to triclosan by tachyzoites (Fig. 1*D*) and bradyzoites (Fig. 1*E*) but did not completely inhibit uptake into encysted bradyzoites under the conditions studied (Fig. 1*F*). Thus, uptake into tachyzoites and bradyzoites is a facilitated process. Uptake of fluoresceinated triclosan R8 (conjugate **4**) into tachyzoites was substantially greater (one and a half logs) than triclosan R4 (conjugate **5**), after incubation for 10 s, 1 min, and 10 min. Uptake of triclosan R8 increased over time (data not shown). As uptake into isolated bradyzoites was blocked by azide, the inability of azide to completely abrogate uptake by encysted bradyzoites is surprising and suggested that more than a single mechanism of uptake might be involved in uptake across the cyst wall.

Fluorescein-tagged conjugate **4** and its enantiomer both entered host fibroblasts. *T. gondii* is an obligate intracellular parasite and resides inside a specialized vacuolar compartment known as the parasitophorous vacuole delimited by the parasitophorous vacuole membrane. To determine whether the carrier conjugated to triclosan can enter this specialized compartment, tachyzoites within a vacuole in human fibroblasts were studied. Intracellular tachyzoites took up fluoresceinated r8-triclosan (conjugate **ent-4**) within minutes (Fig. 2*A*). This finding was also confirmed by flow cytometric analysis (Fig. 2*B*).

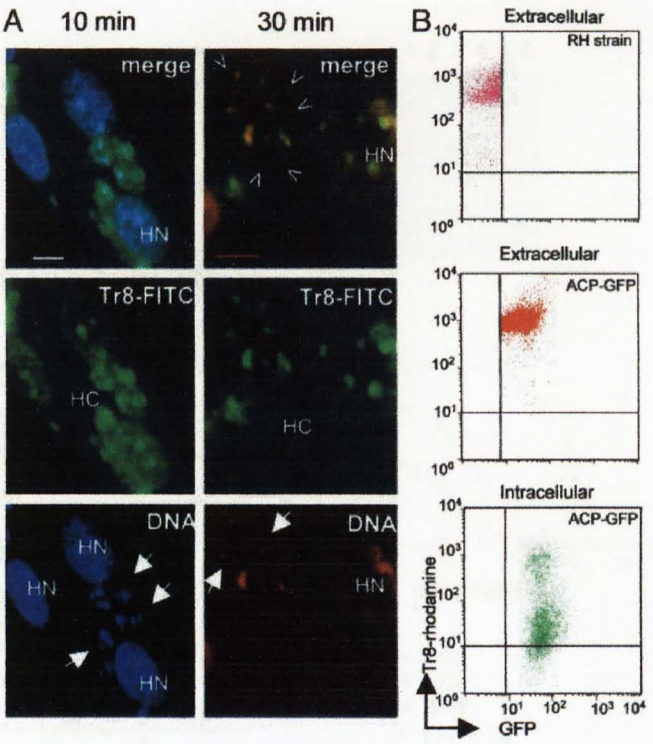


Fig. 2. Intracellular and extracellular tachyzoites take up triclosan conjugated to r8. (*A*) Fluorescence images of live intracellular tachyzoites after 10 or 30 min of exposure to r8-triclosan conjugate (nonreleasable conjugate **4**). Parasite nuclei (arrows) and host cell nuclei (HN) were stained with Hoechst. FITC signal is seen in host cell cytoplasm (HC) and nuclei (HN), and in parasites residing in parasitophorous vacuole (v). (Scale bar = 10 μm.) (*B*) Uptake of triclosan r8 conjugate (nonreleasable conjugate **4**) by extracellular and intracellular tachyzoites. Flow cytometric analysis showing representative dot blots of RH strain (pink) and transgenic parasites expressing the GFP (ACP-GFP, extracellular, red; intracellular, green) exposed to r8-triclosan conjugated to rhodamine (**ent-6**) for 10 min. Intracellular parasites were exposed to the carrier and then released from the host cells before analyses.

Previously, the only apicomplexan ENR identified was that of *Plasmodium falciparum*. We (17) and others (19) have cloned, sequenced, and solved the structure of *P. falciparum* ENR. It was of interest to compare *P. falciparum* ENR and TgENR for the present study, particularly with reference to their triclosan binding sites. Portions of TgENR were identified in a database, full-length cDNA clones were isolated and sequenced, deduced amino acid sequences were determined, and enzymatically active recombinant protein was produced (Fig. 3). Analysis of the deduced amino acid sequence revealed that key amino acid residues are conserved between *T. gondii* (Fig. 3*A*, GenBank accession no. AY372520), *P. falciparum*, *Brassica napus*, and bacterial ENRs and that 11 amino acid residues involved in triclosan binding are also conserved (Fig. 3*A*). This result suggests that triclosan is likely to have a similar mode of inhibition in *T. gondii* as in other species. In keeping with a plastid localization of this enzyme in other apicomplexans, a bi-partite transit sequence consisting of a cleavable von Heijne secretory signal (the first 27 aa) followed by a chloroplast-like targeting sequence (amino acid numbers 28–66) were identified. Alignment with the genomic DNA sequence in the database indicated that there are four introns. In analyzing the amino acid sequence of *P. falciparum* ENR, several amino acid insertions were found. *B. napus* ENR had inserts corresponding to two of those from *P. falciparum* ENR. The sequence of TgENR has multiple amino acid sequences that coincide with those inserts

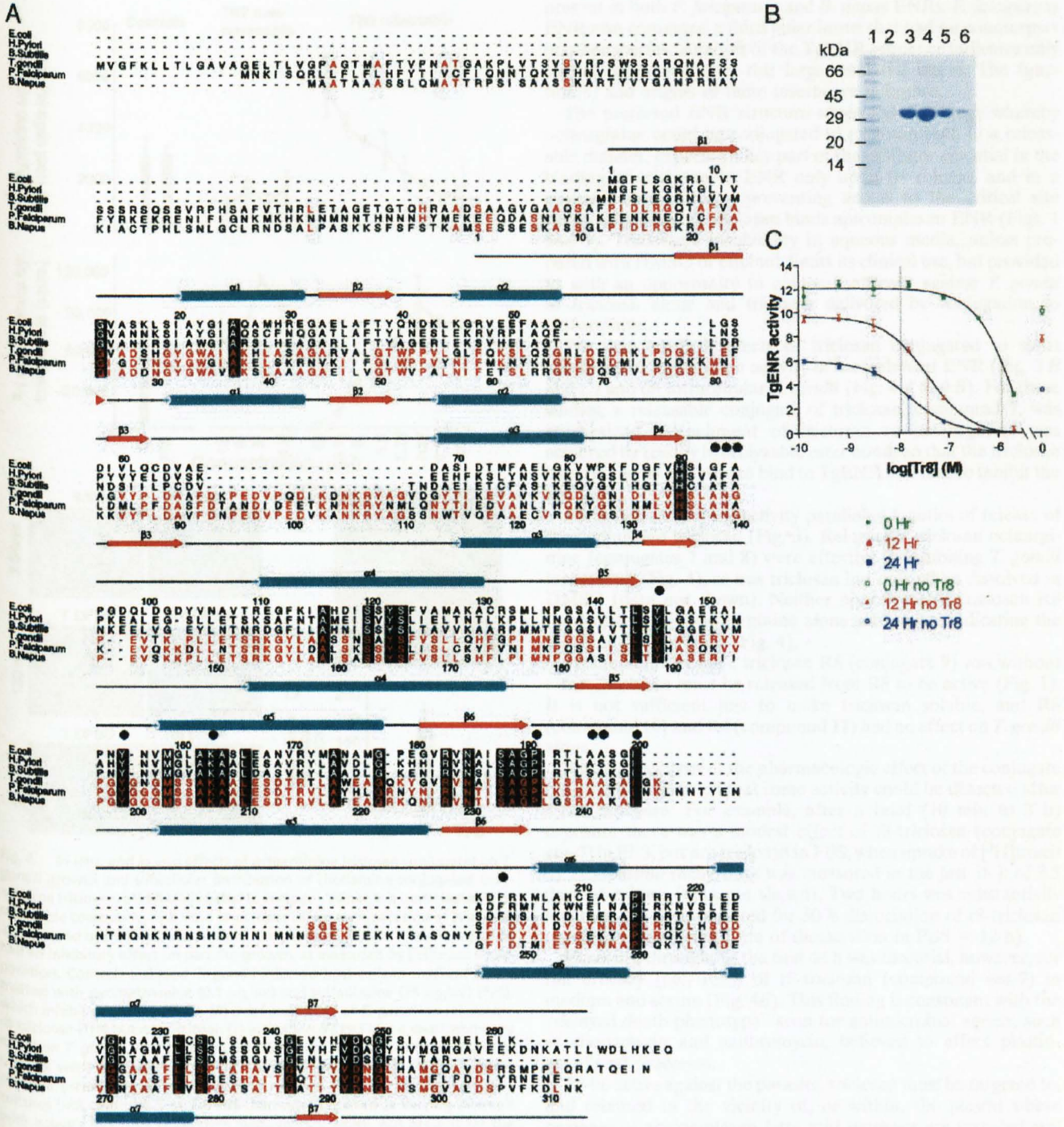


Fig. 3. TgENR: Deduced amino acid sequence and structure, purification of recombinant protein, and kinetics of enzyme activity in the presence of r8 releasable conjugate. (*A*) Multiple structure-based sequence alignment showing deduced amino acid sequence of TgENR and enoyl reductases of *Escherichia coli*, *Helicobacter pylori*, *Bacillus subtilis*, *P. falciparum*, and *B. napus*. This sequence was confirmed by isolation and sequencing of five separate clones from an RH strain *T. gondii* library kindly provided by J. Ajioka (Cambridge University, Cambridge, U.K.). The secondary structures and sequence numbers of the *E. coli* and *B. napus* enzymes are shown above and below the alignment, respectively. The residues that are completely conserved are in reverse type, and those involved in triclosan binding are indicated by ●. The residues in red are conserved between *B. napus* ENR, *P. falciparum* ENR, and TgENR. Red arrows indicate β-sheets, and blue-green cylinders indicate α-helices. (*B*) SDS/PAGE of TgENR purification. SigmaMarker (Sigma) wide-range molecular weight markers are shown in lane 1. Lanes 2–6 show fractions eluted from the HiTrap Q Fast Flow column containing pure TgENR. (*C*) Time course of TgENR inhibition by releasable r8-triclosan conjugate (Tr8), compound **7**. Inhibition curves were measured at 0 h (■), 12 h (▲), and 24 h (●) of incubation at 37°C. The three open symbols to the right show TgENR activity at these time points with no added inhibitor. Enzyme activity is expressed as turnovers per TgENR molecule per s, and inhibitor concentration is expressed as the log of the Tr8 concentration in molar units. IC_{50} values calculated by nonlinear regression analysis are displayed graphically as dashed vertical lines for the three inhibition curves, showing an increase of inhibitory activity with time. A black vertical dashed line marks the concentration of TgENR used in these assays and thus represents a value below which inhibition cannot be properly measured in these assays. Error bars show the variation between triplicate measurements at each point. When not visible, the variation is smaller than the symbol.

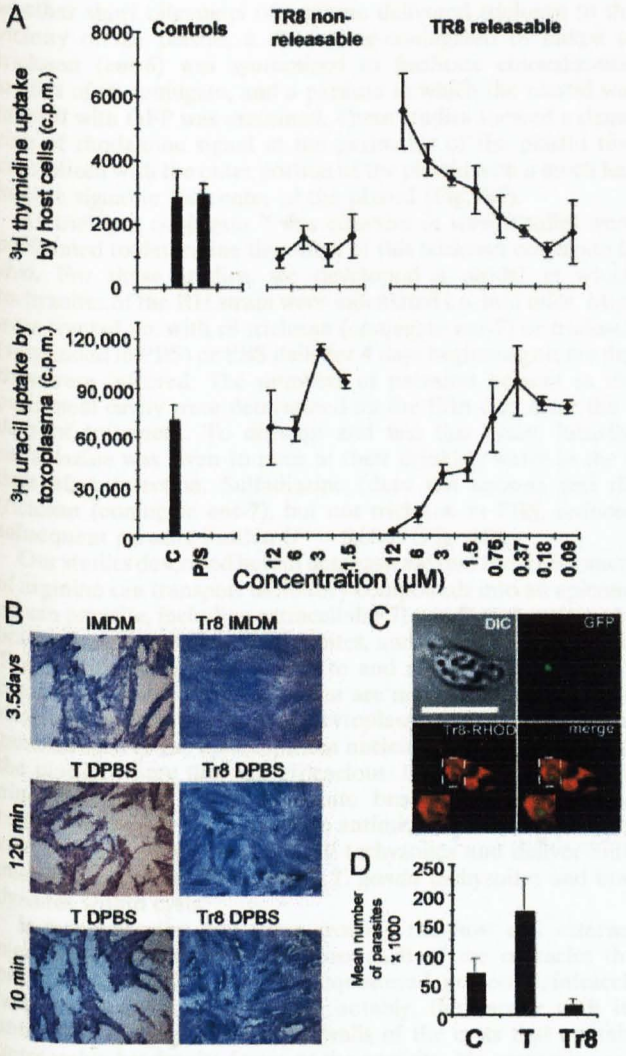


Fig. 4. *In vitro* and *in vivo* effects of octaarginine triclosan conjugates on *T. gondii* growth and subcellular distribution of rhodamine-conjugated octaarginine triclosan. (*A*) R8 conjugated to triclosan, either as a nonreleasable or releasable compound, had little or no toxic effect on replication of host cells as measured by [^{3}H]thymidine uptake. Only releasable R8 triclosan conjugate had an inhibitory effect on parasite growth, as measured by [^{3}H]uracil incorporation. Controls included *T. gondii*-infected host cells in medium (C) or treated with pyrimethamine (0.1 μg/ml) and sulfadiazine (25 μg/ml) (P/S), which inhibit parasite growth. (*B*) Brief exposures of *T. gondii* to releasable r8-triclosan (Tr8) but not triclosan (T) in Dulbecco's PBS had a small inhibitory effect on *T. gondii*. Two hours after infection of host cells with tachyzoites, cultures were incubated with T or Tr8 (12.5 μM) either in PBS or growth medium (Iscove's modified Dulbecco's medium–FCS). For the PBS-exposed cultures (not controls), T or Tr8 PBS then were replaced at varying intervals with Iscove's modified Dulbecco's medium–FCS, which was present for the remainder of the 3.5-day culture period. Parasite growth also was determined by [^{3}H]uracil uptake at the end of 3.5 days of culture in growth medium (data not shown). Brief exposures to Tr8 partially inhibited parasite growth and the 3.5-day exposure (or a 48-h exposure, data not shown) inhibited parasite growth almost completely. Unconjugated triclosan had no effect. The photomicrographs (Giemsa stain) on the right show the "protective" effect of Tr8, but not T, in PBS on the infected host cell monolayer. The large empty areas represent host cell destruction by intracellular parasite growth, which is inhibited by r8-conjugated triclosan and not by the triclosan in PBS. (*C*) Subcellular distribution of r8-triclosan conjugate in tachyzoites was determined by deconvolution microscopy using nonreleasable r8-triclosan conjugated to rhodamine, and transgenic parasites expressing the GFP targeted to the parasite plastid organelle. Fluorescence and differential interference contrast (DIC) images showing rhodamine signal surrounds the GFP signal, partly overlapping it. (*Inset*) Higher magnification of the boxed area showing rhodamine signal partly overlapping plastid. (Scale bar = 10 μm.) (*D*) *In vivo* administration of releasable r8-triclosan, **ent-7**, but not triclosan in PBS, reduces numbers of i.p. tachyzoites 5 days after challenge ($P = 0.006$).

present in both *P. falciparum* and *B. napus* ENRs. *P. falciparum* ENR also contained a third polar insert that had no counterpart in other species. Analysis of the TgENR sequence indicates only partial conservation of this larger malarial insert. The function(s) and origins of these inserts are unknown.

The predicted ENR structure suggested a strategy whereby octaarginine could be conjugated to triclosan both in a releasable manner, exposing a key part of the inhibitor essential in the binding of triclosan to ENR only upon its release, and in a nonreleasable manner, preventing access to the critical site where a hydroxyl in triclosan binds apicomplexan ENR (Figs. 1 and 5). Triclosan's insolubility in aqueous media, unless prepared with DMSO or ethanol, limits its clinical use, but provided us with an opportunity to compare efficacy against *T. gondii* of triclosan alone and triclosan delivered by conjugation to octaarginine.

We then studied effects of triclosan conjugated to short oligomers of arginine on activity of recombinant ENR (Fig. 3 *B* and *C*) and on intracellular *T. gondii* (Fig. 4 *A* and *B*). For these studies, a releasable conjugate of triclosan, compound **7**, was synthesized. Attachment of triclosan to octaarginine was achieved by readily hydrolyzable ester bonds so that the triclosan cargo could be released to bind to TgENR and thus to inhibit the parasite.

Inhibition of enzyme activity paralleled kinetics of release of the conjugated triclosan (Fig. 3). Releasable triclosan octaarginine (conjugates **7** and **8**) were effective in inhibiting *T. gondii* tachyzoites (Fig. 4), as was triclosan but only when dissolved in DMSO (data not shown). Neither nonreleasable triclosan R8 (conjugate **9**) nor octaarginine alone was active, indicating the specificity of triclosan (Fig. 4).

As the nonreleasable triclosan R8 (conjugate **9**) was without effect, triclosan must be released from R8 to be active (Fig. 1). It is not sufficient just to make triclosan soluble, and R8 (compound **10**) and R4 (compound **11**) had no effect on *T. gondii* alone.

A kinetic analysis of the pharmacologic effect of the conjugate ($T_{1/2}$ = 12 h) revealed that some activity could be detected after short exposure: For example, after a brief (10 min to 2 h) exposure there was a modest effect of r8-triclosan (conjugate **ent-7**) in PBS, but not triclosan in PBS, when uptake of [^{3}H]uracil by *Toxoplasma* tachyzoites was measured in the last 18 h of 3.5 days of culture (data not shown). Two hours was substantially less than the time required for 50% dissociation of r8-triclosan (conjugate **ent-7**) (half-life of dissociation in PBS = 12 h).

Exposure throughout the first 48 h was essential, however, for full efficacy (i.e., IC_{90}) of r8-triclosan (compound **ent-7**) in medium and serum (Fig. 4*B*). This finding is consonant with the "delayed death phenotype" seen for antimicrobial agents, such as clindamycin and azithromycin, believed to effect plastid-associated processes.

To be active against the parasite, triclosan must be targeted to, and released in the vicinity of, or within, the plastid where enzymes of apicomplexan fatty acid synthesis are targeted and active (6–10). Because short oligomers of arginine alone and linked to triclosan rapidly entered the parasite cytoplasm and nucleus, we wanted to determine whether short oligomers of arginine linked to triclosan also were in close proximity to or entered the plastid, as the delayed death phenotype of triclosan suggested that they were likely to act on a plastid associated process. Also, analysis of TgENR suggested that the plastid was a likely site of action of this inhibitor of ENR. To determine

whether short oligomers of arginine delivered triclosan to the vicinity of the plastid, a rhodamine-conjugated r8 linked to triclosan (**ent-6**) was synthesized to facilitate colocalization studies of r8 conjugate, and a parasite in which the plastid was labeled with GFP was examined. These studies showed a dense ring of rhodamine signal at the perimeter of the plastid that colocalized with the outer portion of the plastid with a much less intense signal in the center of the plastid (Fig. 4*C*).

As triclosan conjugate **7** was effective *in vitro*, studies were performed to determine the effect of this triclosan conjugate *in vivo*. For these studies, we developed a model in which tachyzoites of the RH strain were inoculated i.p. into mice. Mice were treated i.p. with r8-triclosan (conjugate **ent-7**) or triclosan (suspended in PBS) or PBS daily for 4 days beginning on the day they were infected. The numbers of parasites present in the peritoneal cavity were determined on the fifth day, after the 4 days of treatment. To develop and test this assay, initially, sulfadiazine was given to mice in their drinking water in the 4 days after infection. Sulfadiazine (data not shown) and r8-triclosan (conjugate **ent-7**), but not triclosan in PBS, reduced subsequent parasite burden ($P < 0.006$) (Fig. 4*D*).

Our studies described herein demonstrate that short oligomers of arginine can transport inhibitory compounds into an apicomplexan parasite, including extracellular *T. gondii* tachyzoites and bradyzoites, intracellular tachyzoites, and encysted bradyzoites. The peptide carrier can deliver to and release lead inhibitory compounds inside tachyzoites that are multiplying within parasitophorous vacuoles, to their cytoplasm and subcellular organelles such as the apicomplexan nucleus and the perimeter of the plastid, where they are efficacious. Furthermore, octaarginine can deliver compounds into bradyzoites in cysts and tachyzoites *in vivo*. Oligoarginine antimicrobial conjugates provide a way to eliminate *T. gondii* tachyzoites and deliver antimicrobial agents to intracellular *T. gondii* tachyzoites and bradyzoites within cysts.

It is noteworthy that after crossing the host cell external membrane, octaarginine can cross each of the obstacles the parasite has specified to create sequestered, protected, intracellular or intracyst niches. Most notably, the carrier with its antimicrobial cargo crosses the walls of the cysts that contain untreatable bradyzoite forms of the parasite. No currently used antimicrobial compound previously has been shown to cross the cyst wall.

We selected, designed, and developed a method of delivery as a generic proof of principle, showing the ability of a carrier to deliver small molecules to microbes that are sequestered by multiple membranes and structures. This proof of principle is based on discoveries reported herein of a promising enzyme target for antimicrobials (TgENR), production and purification of recombinant TgENR, demonstration and characterization of TgENR activity, and kinetics of its inhibition by triclosan that parallels kinetics of the release of triclosan from the octaarginine carrier. Deductions concerning the structure of this promising enzyme target are then used to design an approach for testing efficacy and toxicity of releasable octaarginine inhibitor conjugates (Fig. 4). Identification and characterization of this enzyme in *T. gondii* provides useful information for further development of unique classes of antimicrobial agents effective against apicomplexans. It was not obvious or certain *a priori* that *T. gondii* would have this enzyme as certain organisms, e.g., the *pneumococcus*, that use type II fatty acid synthesis do not have an ENR. As a proof of concept, our observations show the potential for delivery, with efficacy and without toxicity, of an inhibitor of ENR by a releasable carrier *in vitro* and *in vivo* (Figs. 3 and 4).

We demonstrate a mechanism of medicine delivery to an important pathogen target enzyme. Octaarginine delivers a "prodrug" processed by the organism to an enzyme that is essential for pathogen survival. A number of inhibitors of ENR described so far are inherently poorly soluble in part because they reflect the lipophilic nature of the enzyme substrate. One clear advantage of the delivery system is that the specific solubilizing agent attaches a strongly hydrophilic functionality to the drug, which gives rise to freely soluble compounds.

As short oligomers of arginine are being optimized for use in delivery across skin, eye, and blood-brain barrier, results described herein provide potential for significantly improving treatment of apicomplexan, and perhaps other, infections. Future characterization of the mechanisms whereby the parasite cyst wall and multiple host and parasite membranes (of prokaryotic and eukaryotic lineage) are traversed by short arginine oligomers to deliver its antimicrobial agent cargo will be of considerable interest.

We thank B. Streipen and D. Roos for providing ACP-GFP-transfected tachyzoites, the RITA Corporation for donating triclosan, S. Bond in the Digital Light Microscopy Facility at the University of Chicago, and M. Sing, E. Castro, and L. Kallal for their assistance with preparation of this manuscript. This work was supported by the National Institutes of Health National Institute of Allergy and Infectious Diseases TMP program (Grant R01 AI43228), The Research to Prevent Blindness Foundation, National Institutes of Health Grants CA 31841 and 31845, gifts from the Keiweit, Blackmon, Brennan, Koshland, and Langel families, the Biotechnology and Biological Sciences Research Council, and The Wellcome Trust. R.M. is the Jules and Doris Stein RPB Professor at the University of Chicago. B.H. was the recipient of a National Institutes of Health postdoctoral fellowship.

1. Rothbard, J. B., Garlington, S., Lin, Q., Kirschberg, T., Kreider, E., McGrane, P. L., Wender, P. A. & Khavari, P. A. (2000) *Nat. Med.* **6**, 1253–1257.
2. Wender, P. A., Mitchell, D. J., Pattabiraman, K., Pelkey, E. T., Steinman, L. & Rothbard, J. B. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 13003–13008.
3. Mitchell, D. J., Steinman, L., Kim, D. T., Fathman, C. G. & Rothbard, J. B. (2000) *J. Peptide Res.* **56**, 318–325.
4. Rothbard, J. B., Kreider, E., Van Deusan, C. L., Wright, L., Wylie, B. L. & Wender, P. A. (2003) *Handbook of Cell-Penetrating Peptides* (CRC, Boca Raton, FL).
5. Boyer, K. & McLeod, R. (2002) in *Principles and Practice of Pediatric Infectious Diseases*, eds. Long, S., Proeber, C. & Pickering, L. (Churchill Livingstone, New York), 2nd Ed., pp. 1303–1322.
6. McLeod, R., Muench, S., Rafferty, J., Kyle, D., Mui, E., Kirisits, M., Mack, D., Roberts, C., Samuel, B., Lyons, R., *et al.* (2001) *Int. J. Parasitol.* **31**, 109–113.
7. Zuther, E., Johnson, J. J., Haselkorn, R., McLeod, R. & Gornicki, P. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 13387–13392.
8. Waller, R. F., Keeling, P. J., Donald, R. J., Streipen, B., Handman, E., Lang Unasch, N., Cowman, A. F., Bersa, G. S., Roos, D. S. & McFadden, G. I. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 12352–12357.
9. Surolia, N. & Surolia, A. (2001) *Nat. Med.* **7**, 167–173.
10. Roberts, C., McLeod, R., Rice, D., Ginger, M., Chance, M. & Goad, L. (2003) *Mol. Biochem. Parisitol.* **126**, 129–142.
11. Levy, C. W., Roujeinikova, A., Sedelnikova, S. E., Baker, P. J., Stuitje, A. R., Slabas, A. R., Rice, D. W. & Rafferty, J. B. (1999) *Nature* **398**, 383–384.
12. Roberts, F., Roberts, C. W., Johnson, J., Kyle, D. E., Krell, T., Coggins, J. R., Coombs, G. H., Milhous, W. K., Tzipori, S., McLeod, R., *et al.* (1998) *Nature* **393**, 801–805.
13. Mack, D. & McLeod, R. (1984) *Antimicrob. Agents Chemother.* **26**, 26–30.
14. Denton, H., Roberts, C. W., Alexander, J., Thong, K. W. & Coombs, G. H. (1996) *FEMS Microbiol. Lett.* **137**, 103–108.
15. Nielsen, H., Englebrecht, J., Brunak, S. & von Heijne, G. (1997) *Protein Eng.* **10**, 1–6.
16. Emanuelsson, O., Nielsen, H. & von Heijne, G. (1999) *Protein Sci.* **8**, 978–984.
17. Muench, S. P., Rafferty, J. B., McLeod, R., Rice, D. W. & Prigge, S. T. (2003) *Acta Crystallogr. D* **59**, 1246–1248.
18. Kapust, R. B. & Waugh, D. S. (2000) *Protein Expression Purif.* **19**, 312–318.
19. Perozzo, R., Kuo, M., Sidhu A. S., Valyaveettil, J. T., Bittman, R., Jacobs, W. R., Fidock, D. A. & Sacchettini, J. C. (2002) *J. Biol. Chem.* **277**, 13106–13114.

Appendix IV

Crystallization and preliminary X-ray crystallographic studies on the class II cholesterol oxidase from *Burkholderia cepacia* containing bound flavin.

Aunpad, R., Muench, S. P., Baker, P. J., Sedelnikova, S., Panbangred, W., Doukyu, N., Aono, R. & Rice, D. W.

Acta Cryst. (2002) 58, 2182-3.

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Crystallization and preliminary X-ray crystallographic studies on the class II cholesterol oxidase from *Burkholderia cepacia* containing bound flavin

Ratchaneewan Aunpad,[a]† Stephen P. Muench,[a] Patrick J. Baker,[a] Svetlana Sedelnikova,[a] Watanalai Panbangred,[b] Noriyuki Doukyu,[c] Rikizo Aono[c] and David W. Rice[a]*

[a]Krebs Institute for Biomolecular Research, Department of Molecular Biology and Biotechnology, The University of Sheffield, Sheffield S10 2TN, England, [b]Department of Biotechnology, Faculty of Science, Mahidol University, Bangkok, Thailand, and [c]Department of Biological Information, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Yokohama, Japan

† Permanent address: Department of Biotechnology, Faculty of Science, Mahidol University, Bangkok, Thailand.

Correspondence e-mail: d.rice@sheffield.ac.uk

Received 16 April 2002
Accepted 23 September 2002

Burkholderia cepacia cholesterol oxidase (ChoS) is a 58.7 kDa molecular-weight flavoenzyme which has been categorized as a 3β-hydroxysteroid oxidase converting the 3β-hydroxyl group of a range of hydroxysteroids to the corresponding ketone. Analysis of enzymes with this activity has shown that two classes of cholesterol oxidase can be defined. Enzymes belonging to class I contain non-covalently bound FAD, whereas the class II enzymes contain FAD covalently bound to an active-site histidine. Despite catalysing the same chemical reaction, the class I and class II enzymes show no sequence similarity and have a different molecular architecture. Crystals of a recombinant class II enzyme from *B. cepacia* have been grown by the hanging-drop vapour-diffusion method using polyethylene glycol as a precipitating agent. The crystals belong to space group $P3_121$, with unit-cell parameters $a = b = 119.6$, $c = 101.1$ Å, and have one subunit in the asymmetric unit. These crystals diffract to at least 2.0 Å resolution at the Daresbury SRS and are suitable for a full structure determination. Ultimately, analysis of the structure of *B. cepacia* ChoS may allow the characteristics and structural features which contribute to its suitability as a diagnostic reagent for the detection of cholesterol and unresolved mechanistic features of the class II enzymes to be understood.

1. Introduction

Cholesterol oxidase (3β-hydroxysteroid oxidase; EC 1.1.3.6), a key enzyme in cholesterol metabolism, is an FAD-dependent enzyme which catalyses the oxidation of 3β-hydroxyl groups of steroids to the corresponding ketone. This enzyme has an important application in clinical assay kits for the determination of serum levels of cholesterol in a coupled system with cholesterol esterase and peroxidase (Allain *et al.*, 1974). Biochemical studies of cholesterol oxidase from *Brevibacterium sterolicum* have suggested that two classes of enzyme can be distinguished (Croteau & Vrielink, 1996). The class I enzymes have been shown to contain non-covalently bound FAD. In contrast, the class II enzymes contain FAD covalently linked to an active-site histidine *via* the C8α atom of the flavin isoalloxazine ring. Despite catalysing the same chemical reaction, these two classes of enzyme show no significant sequence similarity.

Cholesterol oxidase (ChoS) from *Burkholderia cepacia* oxidizes various 3β-hydroxysteroids with the activity for cholesterol being greatest (Doukyu & Aono, 1998). This enzyme is a particularly attractive target for the development of a cholesterol assay because the Michaelis constant of the *Burkholderia* sp. ChoS enzyme is lower than that of other commercially available oxidases. Biochemical studies have shown that a further favourable feature of the *Burkholderia* sp. ChoS is the finding that in the organic solvents used in the clinical assay the enzyme is both stable and shows an oxidation rate enhanced by more than threefold (Doukyu & Aono, 1998). Analysis of the sequence of the cholesterol oxidase from *B. cepacia* showed no similarity to the class I enzymes. Rather, this enzyme shows 46% sequence identity to the class II cholesterol oxidase from *Br. sterolicum*, indicating that the *B. cepacia* enzyme is a class II cholesterol oxidase containing covalently bound FAD.

The structures of representative class I enzymes containing non-covalently bound FAD have been determined for the enzymes from both *Br. sterolicum* (Vrielink *et al.*, 1991) and *Streptomyces* sp. (Yue *et al.*, 1999). More recently, the structure of the *Br. sterolicum* class II cholesterol oxidase containing covalently bound FAD from has been solved to 1.7 Å resolution and has shown that the fold of this enzyme is unrelated to that of the class I enzyme (Coulombe *et al.*, 2001). Analysis of the class II *Br. sterolicum* ChoS has shown that the active site consists of a cavity sealed off from the exterior of the protein. Moreover, a

novel feature suggested for this enzyme from the structure is the observation that a water-filled hydrophobic channel extending to the flavin moiety may act as the entry point for oxygen to the catalytic site during the oxidative half-reaction. Control of access by oxygen to this cavity is suggested to involve concerted movements of key arginine, glutamate and isoleucine residues which were observed to occupy multiple conformations in the structure (Coulombe *et al.*, 2001). In order to contribute to a full understanding of the mechanism of enzymes belonging to this class and, in particular, to the nature of the conformational changes which control gating to molecular oxygen, we have initiated crystallographic studies on *Burkholderia* sp. ChoS. Furthermore, the structure might also provide important insights into the molecular features of *B. cepacia* cholesterol oxidase that make it stable in the presence of organic solvents and therefore suitable as a diagnostic reagent for cholesterol. In this paper, we report the crystallization and preliminary X-ray analysis of ChoS.

2. Materials and methods

The overexpression and initial purification of the *B. cepacia* ChoS were initially carried out as previously described (Doukyu & Aono, 1998). However, to further purify the protein before crystallization trials, an additional step of preparative gel-filtration chromatography was carried out using a Superdex 200 HiLoad 16/60 column (Pharmacia) equilibrated with 10 m*M* PBS buffer pH 8.0. After this step, the enzyme was concentrated to 10 mg ml^{-1} (estimated by the Bradford assay; Bradford, 1976) in 10 m*M* Tris–HCl pH 8.5. The protein sample was then centrifuged at 25 000*g* for 10 min to clarify the solution before initiating any crystal trials.

Preliminary crystallization experiments were carried out at 290 K with the Hampton Crystal Screens I and II and the PEG ion screen using the hanging-drop vapour-diffusion technique, mixing 2 µl of protein solution and an equivalent volume of reservoir solution. Promising trials were refined to give optimum conditions, the best of which were based on a reservoir solution consisting 28% PEG 400, 0.1 *M* Na HEPES pH 7.5 and 0.2 *M* calcium chloride. Under these conditions, crystals with the morphology of square plates of dimensions up to 0.2 × 0.25 × 0.2 mm were obtained after two weeks. Preliminary X-ray analysis of the crystals, flash-frozen using the reservoir buffer as a cryoprotectant, showed that they diffracted well to high resolution. Subsequently, a complete data set was collected to 2.0 Å resolution at 100 K, using 1° rotation frames and X-rays with a wavelength of 0.978 Å on an ADSC Quantum 4 detector at station PX 14.2 at the Daresbury SRS. The data were processed using the *DENZO/SCALEPACK* package (Otwinowski & Minor, 1997).

Table 1
Data-collection statistics for the 2.0 Å ChoS data set.

Values in parentheses indicate values for the highest resolution shell.

Resolution (Å)	15–2.0 (2.15–2.0)
No. of unique reflections	45559
Completeness (%)	88.9 (91)
R_{merge} (%)	0.059 (0.348)
$I/\sigma(I) > 3$ (%)	82.6 (58.7)
Multiplicity (%)	4.48 (4.30)

3. Results and discussion

Autoindexing and scaling of the data using *DENZO/SCALEPACK* and analysis of the pattern of systematic absences is consistent with a primitive trigonal crystal system, identifying the space group as $P3_121$ (or its enantiomorph $P3_221$), with unit-cell parameters $a = b = 119.6$, $c = 101.1$ Å. Significant reflections were observed to the edge of the image-plate detector and a high-quality data set was collected to this resolution (Table 1). Considerations of possible values of V_M suggest that the asymmetric unit could either contain a monomer with a V_M of 3.5 Å^3 Da^{-1} or a dimer with a V_M of 1.75 Å^3 Da^{-1}, both of which lie within the range observed by Matthews (1977), with the former being more probable. Gel-filtration studies on ChoS strongly suggest that this enzyme is a monomer in PBS buffer solution, consistent with this conclusion.

In order to provide a completely independent solution to the structure, we initially carried out a heavy-atom screen which yielded a number of very promising derivatives based on either mercury, gold or gadolinium. Using these derivatives a preliminary map at 2.4 Å resolution was calculated and showed a clear molecular boundary. Whilst this work was in progress, the coordinates for the class II ChoS from *Br. sterilocum* were deposited in the Protein Data Bank (Berman *et al.*, 2000) (PDB code 1i19). A preliminary attempt to solve the structure by molecular replacement gave a solution in space group $P3_121$ corresponding to a monomer in the asymmetric unit, which made acceptable packing interactions in the cell. Given the quality of the derivatives and in order to examine the conformation of key residues at the active site in a manner in which any bias is minimized, we have chosen to proceed with the structure determination using an isomorphous replacement approach. Ultimately, it is hoped that a complete solution of this structure will lead to a better understanding of the enzymes belonging to the cholesterol oxidase family and so help in the design of more efficient clinical tools for the assay of cholesterol.

This work was supported by the New Energy and Industrial Development Organization, the Wellcome Trust, the BBRSC and the British Council. RA is supported by the National Centre for Genetic Engineering and Biotechnology (BIOTEC), National Science and Technology Development Agency (NSTDA), Thailand. The Krebs Institute is a designated BBRSC Biomolecular Science Centre and a member of the North of England Structural Biology Centre.

References

Allain, C. C., Poon, L. S., Chan, C. G. S., Richimond, W. & Fu, P. C. (1974). *Clin. Chem.* **20**, 470–475.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.
Coulombe, R., Yue, K, Q., Ghisla, S. & Vrielink, A. (2001). *J. Biol. Chem.* **276**, 30435–30441.
Croteau, N. & Vrielink, A. (1996). *J. Struct. Biol.* **116**, 317–319.
Doukyu, N. & Aono, R. (1998). *Appl. Environ. Microbiol.* **64**, 1929–1932.
Matthews, B. W. (1977). *X-ray Structure of Proteins*, 3rd ed., edited by H. Neurath & R. L. Hill, Vol. 3, pp. 468–477. New York: Academic Press.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Vrielink, A., Lloyd, L. F. & Blow, D. M. (1991). *J. Mol. Biol.* **219**, 533–554.
Yue, Q. K., Kass, I. J., Sampson, N. S. & Vrielink, A. (1999). *Biochemistry*, **38**, 4277–4286.