Structural and Functional Studies of Peptidyl-prolyl cis-trans isomerase A and 1-deoxy-D-xylulose-5-phosphate reductoisomerase from Mycobacterium tuberculosis

LENA M HENRIKSSON
Abstract

Mycobacterium tuberculosis, the causative pathogen of tuberculosis, currently infects one-third of the world’s population, resulting in two million deaths annually. This clearly shows that tuberculosis is one of the most serious diseases of our times. The often unpleasant side effects from the current drugs, combined with the difficulty of ensuring patient compliance, and the emergence of drug-resistant and multidrug-resistant strains, makes the need for new and better drugs urgent.

In this thesis, all the steps, from cloning, purification, crystallization, to activity determination, and structure determination are presented for two different M. tuberculosis enzymes. The structures, which were modeled from X-ray crystallographic data, provide the framework for structure-based drug design. Here, new potential inhibitors can be tailor-made based on the specific interactions in the enzyme’s active site.

The bacteria have two different peptidyl-prolyl cis-trans isomerases that catalyze the isomerization of peptide bonds preceding proline residues, a process of high importance for correct folding. Here we present the structure of peptidyl-prolyl cis-trans isomerase A, an enzyme present inside the bacteria, and distinguish it from the B form of the enzyme, which is membrane bound, placing its active site outside the bacteria.

The enzyme 1-deoxy-D-xylulose-5-phosphate reductoisomerase catalyzes the second step within the non-mevalonate pathway, which leads to the production of isopentenyl diphosphate. This compound is the precursor of various isoprenoids, vital to all living organisms. In humans, isopentenyl diphosphate is produced via a different pathway, indicating that all the enzymes within the non-mevalonate pathway may be suitable drug targets in M. tuberculosis. Several structures of both wild type and mutant 1-deoxy-D-xylulose-5-phosphate reductoisomerase in complex with different substrates, and also with the known inhibitor fosmidomycin, provide valuable information not only to the field of drug design, but also, in this case, into the catalysis.

Keywords: Mycobacterium tuberculosis, Rv0009, peptidyl-prolyl cis-trans isomerase, Rv2870c, 1-deoxy-D-xylulose-5-phosphate reductoisomerase, non-mevalonate pathway, DOXP/MEP pathway, X-ray crystallography

Lena M Henriksson, Department of Cell and Molecular Biology, Box 596, Uppsala University, SE-75124 Uppsala, Sweden

© Lena M Henriksson 2007

ISSN 1651-6214
urn:nbn:se:uu:diva-8253 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-8253)
List of publications

This thesis is based on the following publications, which in the text are referred to by their roman numerals:


The papers are reprinted with permission from the copyright holders.

Cover illustration kindly provided by Daniel Ericsson.
Contents

Introduction.................................................................................................................................9

Aim and outline of thesis ...........................................................................................................11

General Methods.........................................................................................................................13
  Isolating a gene.......................................................................................................................13
  Vectors used in cloning............................................................................................................13
  Expression ...............................................................................................................................14
  Purification...............................................................................................................................15
  Crystallization ........................................................................................................................15
  X-ray crystallography............................................................................................................18
  Molecular replacement............................................................................................................24
  Model building and refinement .............................................................................................27
  Validation.................................................................................................................................29

The tubercle bacillus..................................................................................................................31
  Robert Koch and the discovery of \textit{M. tuberculosis} ......................................................31
  Bacillus characteristics and route of infection .................................................................31
  The \textit{M. tuberculosis} genome .....................................................................................33
  Diagnostics, vaccines, and current drugs ................................................................................34
  Essentiality and target selection .............................................................................................35

Peptidyl-prolyl \textit{cis-trans} isomerase A (Paper I) .................................................................37
  \textit{M. tuberculosis} PpiA and PpiB ....................................................................................37
  Cloning, expression, and purification.....................................................................................38
  Activity determination ............................................................................................................39
  The \textit{Mt}PpiA crystals .........................................................................................................41
  From data collection to final model ....................................................................................42
  Overall structure and active site ............................................................................................43
  Concluding remarks .............................................................................................................44

1-deoxy-D-xylulose-5-phosphate reductoisomerase (Paper II and III) ..................................45
  The mevalonate and non-mevalonate pathways ....................................................................45
  Constructs, constructs, and more constructs .........................................................................49
  Cloning, expression and purification of \textit{Mt}DXR and \textit{Mt}DXR\textsuperscript{NQ} ...........................................51
  Activity, and inhibition by fosmidomycin ...............................................................................52
  Activity assay .......................................................................................................................52
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CoA</td>
<td>Coenzyme A</td>
</tr>
<tr>
<td>DXP</td>
<td>1-deoxy-D-xylulose 5-phosphate</td>
</tr>
<tr>
<td>DXR</td>
<td>1-deoxy-D-xylulose-5-phosphate reductoisomerase</td>
</tr>
<tr>
<td>DXS</td>
<td>1-deoxy-D-xylulose-5-phosphate synthase</td>
</tr>
<tr>
<td>EcDXR</td>
<td>DXR from <em>Escherichia coli</em></td>
</tr>
<tr>
<td>F</td>
<td>Fosmidomycin</td>
</tr>
<tr>
<td>IFN-γ</td>
<td>Interferon γ</td>
</tr>
<tr>
<td>IPP</td>
<td>Isopentenyl diphosphate</td>
</tr>
<tr>
<td>IspC</td>
<td>1-deoxy-D-xylulose-5-phosphate reductoisomerase</td>
</tr>
<tr>
<td>IspD</td>
<td>4-diphosphocytidyl-2-C-methyl-D-erythritol synthase</td>
</tr>
<tr>
<td>IspE</td>
<td>4-diphosphocytidyl-2-C-methyl-D-erythritol kinase</td>
</tr>
<tr>
<td>IspF</td>
<td>2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase</td>
</tr>
<tr>
<td>IspG</td>
<td>4-hydroxy-3-methylbut-2-en-1-yl-diphosphate synthase</td>
</tr>
<tr>
<td>IspH1/H2</td>
<td>4-hydroxy-3-methylbut-2-enyl-diphosphate reductase 1/2</td>
</tr>
<tr>
<td>M</td>
<td>Mn²⁺</td>
</tr>
<tr>
<td>MEP</td>
<td>2-C-methyl-D-erythritol 4-phosphate</td>
</tr>
<tr>
<td>MDR-TB</td>
<td>Multidrug-resistant tuberculosis</td>
</tr>
<tr>
<td>MtDXR</td>
<td>DXR from <em>Mycobacterium tuberculosis</em></td>
</tr>
<tr>
<td>MtPpiA</td>
<td>PpiA from <em>Mycobacterium tuberculosis</em></td>
</tr>
<tr>
<td>N, NADPH</td>
<td>Nicotinamide adenine dinucleotide phosphate, reduced form</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene glycol</td>
</tr>
<tr>
<td>Ppi</td>
<td>Peptidyl-prolyl cis-trans isomerase</td>
</tr>
<tr>
<td>r.m.s.</td>
<td>Root mean square</td>
</tr>
<tr>
<td>S</td>
<td>Sulfate</td>
</tr>
<tr>
<td>TB</td>
<td>Tuberculosis</td>
</tr>
<tr>
<td>TNF-α</td>
<td>Tumor necrosis factor α</td>
</tr>
<tr>
<td>TraSH</td>
<td>Transposon site hybridization</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
<tr>
<td>XDR-TB</td>
<td>Extensively drug-resistant tuberculosis</td>
</tr>
</tbody>
</table>
Introduction

On the wall in my grandparent’s home hangs a painting showing a man and his son buying their Christmas tree while snowflakes swirl around them. The painter was my grandfather’s younger brother who died of tuberculosis (TB) in 1947 at the age of 26. Hearing this story, told by my mother, is probably the first contact I had with the term tuberculosis. Even though this disease was a reality for my grandparents I must confess that growing up, I believed that tuberculosis was more or less extinct. And even though, in my opinion, this perception is still common, that situation lies far from the truth. According to the World Health Organization (WHO, http://www.who.org), 2 billion people, one-third of the world’s population, are currently infected with *Mycobacterium tuberculosis*, the causative pathogen of tuberculosis. Of these, 5-10% will become infectious or ill during some time of their life. For a person who is also HIV-positive this risk is much higher. The lethal combination of HIV and a tuberculosis infection is seen, perhaps most clearly, in Africa, where HIV is the main cause for the increase of new TB cases. But, even though the highest incidence, and prevalence, of tuberculosis is found in Africa and South-East Asia, we could even in Sweden during 2005 read about a small outbreak in one of our primary schools (Berggren-Palme et al., 2005). The emergence of multidrug-resistant TB (MDR-TB), caused by TB bacilli resistant to the most common first-line TB drugs isoniazid and rifampin, and also recently of extensively drug-resistant TB (XDR-TB), caused by TB bacilli resistant to second-line drugs, as well, adds to the picture that tuberculosis is a disease that we cannot ignore. In 1993 the World Health Organization declared tuberculosis a global emergency, and included in their Global Plan to STOP TB is the goal of producing the first new TB drug in 40 years, by the year 2010.

Proteins are one of the groups of macromolecules found in living organisms, and that participate in all processes within the cells. They were first described and named in 1838 by Jöns Jakob Berzelius. The name protein comes from the Greek word “prota”, meaning “primary importance”, which is a very descriptive name for this diverse and highly essential group of molecules. They have mechanical functions, as they build up muscles and other tissues, they participate in cell signaling, and in immune responses. A large group of the proteins, called enzymes, is responsible for catalyzing all biochemical reactions. Each protein is build up from chains of amino acids that are attached to each other by peptide bonds. The protein’s corresponding
gene, present in the genetic material, defines the order, in which the amino acids are placed within the protein. In order for the protein to form a stable structure it arranges itself in \(\alpha\)-helices and \(\beta\)-sheets, where the amino acids can hydrogen bond to each other in an ordered way. The three-dimensional structure is important for the protein’s function, and provides valuable information when it comes to the development of new drugs. By knowing the appearance of the protein, and knowing how it moves and the nature of the contact surfaces between the protein and its substrate, molecules can be designed to bind in the substrate’s place, thereby hopefully inhibiting the protein’s function.

The most common method used for determining a protein structure is X-ray crystallography. Here, a crystal, containing a three-dimensional repetitive pattern of protein molecules, is exposed to X-rays. Due to the ordered arrangement of the molecules within the crystal, the rays diffracted by the atoms will produce a diffraction pattern, which can be recorded on a detector. Different computer programs can in turn process these data and the result will eventually be visualized in the form of an electron density map. By analyzing the appearance of the density, the amino acids of the protein can be placed within it, until a complete three-dimensional model is formed. The first protein structure ever to be solved was that of sperm whale myoglobin (Kendrew et al., 1958), presented in 1958 by John Cowdery Kendrew, an achievement that four years later was recognized by the Nobel Prize in Chemistry.
Aim and outline of thesis

The RAPID (Rational Approaches to Pathogen Inhibitor Discovery) program at Uppsala University began officially in January 2003. Here, medicinal and computational chemistry, as well as structural biology, is combined with the aim of drug discovery, targeting foremost the diseases tuberculosis and malaria. Within the structural biology group, a medium-throughput system has been developed for the cloning, expression, purification, crystallization, and finally structure determination of enzymes. The targeted enzymes are foremost selected for analysis based on their predicted essentiality for the pathogen. Biological issues such as extracellular location or a potential transport pathway through which a drug could make its way into the pathogen are also evaluated. Among the selected enzymes are those participating in metabolism, replication, survival, and for *M. tuberculosis*, also in cell-wall biosynthesis. The catalytic activities of the recombinant enzymes are evaluated, and for those enzymes that are annotated in public databases as unknown, a function is assigned if possible.

The medicinal chemistry group search for, and design, compounds that could act as new potential inhibitors. The compounds are evaluated using such chemical criteria as their predicted solubility, drug-like characteristics and whether they provide good scaffolds for generating a series of similar compounds. Within the computational chemistry group, the selected compounds are docked into the enzyme’s three-dimensional structure, by using molecular dynamics simulations. The mode of binding, as well as the energy associated with this binding, is used as a measure of the compound’s specificity. When a potential inhibitor is found, the work is directed back to the structural biology group. The inhibitory effect is then evaluated, and co-crystallization of the enzyme and compound is performed in order to obtain a new three-dimensional structure, in which the specific interactions can be seen. At different stages throughout the experimental work, the RAPID group has also had a most fruitful collaboration with AstraZeneca India Pvt. Ltd.

The aim of this thesis is to present two different projects, covering the *M. tuberculosis* enzymes, peptidyl-prolyl *cis-trans* isomerase A (*MtPpiA*) and 1-deoxy-D-xylulose-5-phosphate reductoisomerase (*MtDXR*). The results, and questions raised during cloning, expression, purification, crystallization, and structure determination are covered. Catalytic and inhibitory measurements are also presented. The three-dimensional enzyme structures are de-
scribed in detail, and in the case of a complex structure, the interactions with cofactors, ligands, or inhibitors are evaluated. Comparisons to other related enzyme sequences and structures are also presented, and in the case of MtDXR, a discussion is also presented concerning mechanisms and future perspectives for the project. Some of the sections included in the description of the different projects are added mainly in order to enable the reader to fully follow the different steps performed. Other sections are extended compared to what has previously been presented in the published papers. This has been done to highlight a few interesting questions and conclusions encountered during the projects.

Prior to the presentation and discussion of the *M. tuberculosis* enzymes, I have prepared two other sections. In the first, the general methods used in these studies, and the questions and problems sometimes associated with them, are presented. In the second, a brief description of the discovery and attributes of the pathogen *Mycobacterium tuberculosis* are given, as well as a discussion about essentiality and target selection.
General Methods

The road from target selection to a refined protein structure is one filled with hurdles, which can sometimes be difficult to pass. This methods section is not in any way a complete description of all the techniques that can be used, or the problems one can face. It merely gives a short presentation of the theory associated with X-ray crystallography, and a description and discussion of the main methods used during the projects presented in this thesis.

Isolating a gene

Isolation and amplification of a gene is preformed by PCR (Polymerase Chain Reaction). The gene can either be isolated from total DNA or from a cosmid, an artificially constructed cloning vector, containing the gene of interest. The reaction is initiated when specially designed primers, small DNA sequences, attach to opposite ends of the two DNA strands. Then, an enzyme, a polymerase, continues to build strands, complementary to the first strands, by attaching nucleotides present in the reaction mixture. A high GC content in a primer decreases its specificity, and can therefore cause problems in the PCR reaction. In order to produce a soluble protein, purify it, and later to be able to crystallize it, the exact sequence being cloned is of utter importance. Therefore, several constructs should preferably be made to increase the chances of reaching the goal, which here is that of a protein structure. An ideal construct should produce a globular, compact protein, lacking flexible, disordered ends. The construct design will be discussed more in detail in the expression and crystallization sections.

The two genes of interest in this thesis, designated as Rv0009 and Rv2870c (http://genolist.pasteur.fr/Tuberculist/), were both originally isolated from total DNA from the H37Rv strain of *M. tuberculosis* (Cole et al., 1998). All constructs had an N-terminal affinity tag, containing 6 hisidine residues, in order to, further on, ease the purification.

Vectors used in cloning

The genes discussed here, were cloned into one of the vectors, pCR®/T7/CT-TOPO® or pET101/D-TOPO® (Invitrogen). These have several characteris-
tics in common. They both have a TOPO® cloning site, allowing rapid insertion of the PCR product, as well as a T7 promotor, which provides the binding site for T7 RNA polymerase responsible for transcription (Studier et al., 1990). They also have a ribosome-binding site located close to the cloning site, optimizing the translation of the gene. Further, an ampicillin resistance gene allows for selection of the vectors in *E. coli*. The pET101/D-TOPO® vector differs from pCR®T7/CT-TOPO® by also having a repressor-binding site in its DNA, located close to the ribosome-binding site. Since it also has a gene coding for the repressor in question, the repressor will under normal conditions bind to the repressor-binding site, thereby preventing the ribosome from binding to its site. The repressor will only change its conformation and detach from its site in the presence of isopropyl-β-D-thiogalactopyranoside. This means that, as long as the expression is not physically induced, there will be almost no basal expression of the gene. Therefore, pET101/D-TOPO® offers a tighter regulation.

**Expression**

In order to produce a protein, the vector containing the corresponding gene must be inserted into an expression strain.

The two different strains used in the projects are *E. coli* BL21-AI™ and BL21-STAR™(DE3) (Invitrogen). BL21-STAR™(DE3) contains a copy of the T7 RNA polymerase gene under the control of the UV5 promotor. This promotor is induced by isopropyl-β-D-thiogalactopyranoside, but even though the induction is dose-dependent, the range for varying the expression is very narrow. Even without induction there is a basal level of protein expression with the UV5 promotor, a property which is a problem for expression of a toxic protein, but which can be exploited for a non-toxic protein. In addition, this cell line contains a mutation in the gene encoding RNaseE, which thereby stabilizes the mRNA. The BL21-AI™ cells have a tighter regulation of the polymerase expression, making them suitable for also expressing proteins toxic to *E. coli*. Here, the araBAD promotor is used instead of the UV5 promotor, and expression is induced by adding L-arabinose, with which the level of expression is easy to vary. Since the solubility, and folding, of a protein can be affected by the level of expression it is important to find the optimal combination of vector and expression strain.

Over-expression of a heterologous protein frequently results in situations where a fraction, or sometimes all of the material, is found in an insoluble form. One reason for this can be that the protein is not correctly folded. There are several proteins, including from our experience proteases and ribonucleotide reductase, that are known to have N-terminal sequences that are crucial for the folding, and that are later cleaved off from the pro-protein for activation. Therefore, leaving an extra sequence, N-terminal to the full-
length gene, can be one method of enabling correct folding. Another reason for insolubility could be the presence of hydrophobic patches on the protein surface. In order to increase the charge, and thereby the solubility, elongation of the C-terminus by a poly-lysine sequence can be used, as for Hepatitis C protease (Yan et al., 1998). Also, membrane-spanning proteins are known to be difficult to express in soluble form (Wagner et al., 2006). Here, cleaving off the hydrophobic trans-membrane section can have a positive affect, provided that the remaining protein is able to attain a stable fold. In some cases a protein might also need a “companion”. The companion can be in the form of a cofactor, or another protein that it usually co-exists with in complex form.

**Purification**

All the recombinant proteins used in this work contained an N-terminal His$_6$-tag, which was exploited in the first purification step. Here, the protein is purified by immobilized-metal affinity chromatography (IMAC) when applied to a Ni-NTA-column (Qiagen). Four of the six coordination sites around a nickel ion are bound to the resin NTA, nitrilotriacetic acid, leaving two sites available for interactions with the histidines in the protein tag. After washing away proteins that are not specifically bound, the protein is eluted by adding imidazole, which competes with the histidines for the interaction with the metal. Final purification was performed by size exclusion chromatography on a hiLoad$^{\text{TM}}$ 16/60 Superdex$^{\text{TM}}$ 75, or 200, column (Amersham Biosciences). Here, the matrix consists of dextran, covalently bound to highly cross-linked agarose beads. Since a smaller protein can take more advantage of the space inside the matrix, the path traveled will be longer than for a bigger protein. Consequently, a larger protein will elute from the column before the smaller protein, thereby resulting in a size separation.

**Crystallization**

For a protein to crystallize it must reach the thermodynamically unstable state of supersaturation. This is easily demonstrated in a vapor diffusion experiment, one of the most common techniques used to crystallize a protein. A schematic experiment, with a hanging drop, is demonstrated in Figure 1. Here, a drop, containing equal volumes of protein and precipitant solution, is dispensed on the inside of a plastic cap. The cap is sealed tight over a container with precipitant solution at the bottom. Due to the fact that the concentration of the precipitant solution is lower in the protein drop than in the bottom solution, water will start to diffuse from the drop to the bottom. In the process of reaching equilibrium, the protein will eventually be-
come supersaturated, as the drop gets smaller. At supersaturation several things can happen. The two most common scenarios are that the drop either remains in a supersaturated state, or that the protein forms a precipitate of insoluble material. In some rare cases though, a nucleation event will occur. Protein molecules will then have started to make contacts with each other, spontaneously forming a three-dimensional repetitive pattern, a crystal. Even though the system will reach equilibrium, the crystals can continue to grow as long as the protein concentration is high enough.

Figure 1. In a vapor diffusion experiment, using a hanging drop, the protein in the mixed protein/precipitant drop, will go from unsaturated (1), to supersaturated (2), as the system moves towards equilibrium. At supersaturation, nucleation can occur, leading to the formation of crystals. These can then grow (3) as long as there is an excess of protein molecules in the drop.

The process of vapor diffusion can also be illustrated in a simplified phase diagram, shown in Figure 2, experiment 1. Here, the protein concentration is plotted against the precipitant concentration and we can follow the drop from the unsaturated state to the nucleation zone, within the supersaturated area. When nucleation has occurred the precipitant concentration is basically at a steady state. The formed nuclei will start to grow within the growth zone, and continue to do so until the solubility curve is reached. In many cases, as already mentioned, the system never reaches the nucleation zone, or goes straight through it without forming nuclei, leaving the experiment unsuccessful in terms of crystals.

During the projects presented in this thesis, vapor diffusion, using both sitting and hanging drops, has been used. In addition, both vapor diffusion experiments and a batch experiment in combination with seeding, have been used with the aim of crystal optimization. When seeding, a needle, cat whisker or hair from a horsetail is used for picking up crystalline material from an already existing crystal. The tip is then dragged through a new crystallization drop, which has been left to equilibrate for a period of time rang-
ing from a few hours up to a day or two. By performing this so called “streak seeding”, nucleation sites are added to the drop. In Figure 2, experiment 2, one can see that by doing so the drop does not have to reach the nucleation zone but crystal growth can start immediately. In a batch experiment, number 3 in Figure 2, the protein drop and bottom solution are set up in such a way that they are already in equilibrium, and the supersaturated protein drop is seeded immediately, initiating crystal growth.

Figure 2. The different states, in which a crystallization drop can exist, are here illustrated in a phase diagram, where the protein concentration is plotted against the precipitant concentration. Examples are given for three different crystallization setups, those of a vapor diffusion experiment, without (1) and with (2) seeding, and that of a batch experiment (3). The star indicates a spontaneous nucleation, while the encircled stars show nucleation by seeding.

Not only the concentration of precipitant is important, but also the choice of precipitant used in the experiment. Common precipitants are salts, like ammonium sulfate, organic solvents, including ethylene glycol and polyethylene glycol (PEG) 200, or different polymers, represented mostly by larger PEGs. The precipitants enable crystallization in different ways. Salts work by influencing the “water of solvation”, which means that the salt, and protein molecules, compete for the water molecules in the drop. By attracting water, the salt forces the protein molecules closer to each other and thereby the chances of the molecules to start forming a crystal will increase. PEGs
instead work through “size of exclusion”. Since PEGs have large radii, and are quite mobile, their movement, and the place they will occupy, will force the protein molecules to gather in certain areas. During a lecture I attended in 2005, Alexander McPherson compared it to a “boogie-woogie” couple dancing in the middle of a group of waltzing couples, forcing the quieter dancers to the edge of the dance floor (personal communication).

Other parameters affecting the crystallization experiment are temperature, pH and protein concentration. Also, the state of the protein sample is important. One strives for a stable, pure and homogeneous sample. This can to a certain extent be reached through the design of the construct. Time spent on comparisons to homologous sequences and structures, and making secondary structure predictions, is not wasted. By identifying N-terminal or C-terminal stretches not predicted to have a defined structure, flexible parts, with the potential of disturbing crystal contacts, can be largely avoided. Certain specific residues can also cause problems. Lysines predicted to be positioned on the surface could be mutated into alanine, since the more flexible lysines rarely participate in crystal packing (Goldschmidt et al., 2007). Further, cysteins not predicted to form disulfide bonds in the native structure, and not thought to be involved in some other aspect of structure or functions, might be removed in order to avoid incorrect folding of the protein. However, no matter how much effort is put into preparation, certain problems are not noticed until the protein is being expressed or purified, sometimes forcing the crystallographer to consider designing a new construct.

X-ray crystallography

The three-dimensional repetitive pattern of molecules forming a crystal can be looked upon as being built up by building blocks, here called unit cells, see Figure 3. The lengths of the unit cell edges are commonly referred to as a, b, and c, and the angles between them are called $\alpha$, $\beta$, and $\gamma$. Since the unit cells have to build up the entire crystal, leaving no gaps, the unit cell is defined as the smallest part of the crystal from which the entire crystal can be formed, by using pure translations, each a multiple of a full unit cell length, along the corresponding edges. A unit cell can contain one or several molecules. In most cases, these molecules are related to each other by symmetry, and the unit cell can be further divided into asymmetric units, which can be looked upon as the smallest unique portion of the crystal. The combinations of symmetry operations, with which identical molecules from different asymmetric units are related to each other, are referred to as the crystal’s space group. Sometimes the asymmetric unit can itself contain multiple molecules. Here, the molecules are not in identical environments, although they are generally related to each other by some sort of symmetry operation; this situation is referred to as non-crystallographic symmetry.
Figure 3. The crystal can be seen as being built up from small building blocks called unit cells. Due to symmetry between molecules within one unit cell, the cell can usually be further divided into asymmetric units, which are then the smallest unique elements of the crystal. The unit cells are not always perfectly aligned, something that will affect the amount of detail seen in the protein structure.

By placing a point at every corner of each unit cell, the crystal can also be looked upon as a point lattice. The edges a, b, and c here lie along the x, y, and z-axis, respectively. By placing planes, at specific distances from each other, containing all points in the lattice, a family of planes is formed, see Figure 4. The crystal contains many families of planes, since they can be placed with different angles to each other, or have different spacing in between them. Each family of planes is defined by the number of segments that the, a, b, and c edges are divided into. These numbers are called $h$, $k$, and $l$ for a, b, and c, respectively, and are referred to as the Miller indices. A specific plane does not have to contain a lattice point, but as long as all points are placed in some plane, it accounts for a valid family of planes.

Figure 4. The family of planes with Miller indices, $(h,k,l)$, of $(2,2,1)$ are here drawn within a unit cell.
Figure 5. A family of planes is hit by an incident wave of X-rays at an angle $\theta$. In order for the diffracted waves, reflected from different planes within the family to cause constructive interference, they must scatter in phase. From the figure one can see, that by geometry, the distances $x$ and $2x$ must be equal to $d_{hkl} \sin \theta$ and $2d_{hkl} \sin \theta$, respectively. Consequently, for the waves to scatter in phase, the double distance of $x$ and $2x$, must be equal to an integer, $n$, times the wavelength, $\lambda$. This condition for diffraction from a family of planes is described in Bragg’s law, stating that 

$$n\lambda = 2d_{hkl} \sin \theta.$$ 

When accelerated electrons hit metal, energy will be released in the form of X-ray photons. The X-rays can be looked upon as particles, but also as waves having a specific amplitude and phase. During data collection, the crystal is exposed to X-rays of a specific wavelength, ranging from 0.5-1.6 Å for a typical experiment. When the X-rays hit the crystal, most of them will go straight through, but some will hit the atoms in the crystal, causing them to oscillate. The X-rays will be reemitted with the same amplitude and phase as before, but in a different angle. If one considers the reflections of the incident X-rays from a family of planes, as shown in Figure 5, one can see that unless the path difference between the planes is an integer times the wavelength for the incident wave, there will be destructive interference. The conditions for diffraction, caused by perfect constructive interference, are stated in Bragg’s law

$$n\lambda = 2d_{hkl} \sin \theta.$$
where \( d_{hkl} \) is the distance between the planes in a family, \( \theta \) is the angle between the incident wave and the planes, \( \lambda \) the wavelength of the incident wave, and \( n \) an integer. From this, one can also see that the smallest distance, from a theoretical standpoint, between planes within a family is limited by the wavelength. Even though the law describes the diffraction from specific planes this is just a way in which to illustrate the diffraction. Therefore it should be stated that all the atoms in the crystal contribute to the scattering, not only those positioned in a plane.

In the experimental setup, X-rays diffracted by the crystal will hit a detector, which can measure the intensity. The diffraction pattern so generated is related to the Fourier transform of the crystal. Each diffraction spot in the pattern is related to a specific family of planes. The spot’s position on the detector, is related to the angle \( \theta \), and hence, to the distance \( d_{hkl} \) between the family of planes. In a similar way, the distances and angles between different spots are related to the size and shape of the unit cell. The lattice so generated in diffraction space is called the reciprocal lattice. This is an imaginary lattice that, as explained, is reciprocal to the crystal point lattice, which is defined in real space. In a construction called the Ewald’s sphere, see Figure 6, the direction of diffraction can easily be demonstrated. Here one can see that when a reciprocal lattice vector, corresponding to a certain family of planes, passes a sphere of radius \( 1/\lambda \) the diffracted beam will be at an angle of \( 2\theta \) to the incident beam, where \( \theta \) is the angle with which the beam hits the family of planes. The \( d_{hkl} \) spacing for the diffraction spots with the largest \( \theta \) values is termed the resolution of the diffraction study. Other things being equal, the resolution defines the ultimate reliability of a diffraction experiment.

In order to explain the necessary calculations going from the individually diffracted waves to a protein structure, the diffracted wave for each family of planes, having both an amplitude and a phase, are described by its structure factor equation, \( F(h,k,l) \). For a good description, see (McPherson, 2003).

\[
F(h,k,l) = V \sum_{j=1}^{N} f_j \left[ \exp \left( -\frac{B_j \sin^2 \theta}{\lambda^2} \right) \cos 2\pi(hx + ky + lz) + i \sin 2\pi(hx + ky + lz) \right]
\]

The structure factor is a summation over all atoms (\( N \)) in the unit cell, and therefore it takes into account the scattering power, here called the scattering factor, \( f_j \), for each atom, \( j \). The scattering factor is proportional to the number of electrons in the atom, and it is constant at \( \theta = 0 \), but then falls off. \( V \) is the volume of the unit cell, \( \lambda \) the wavelength, and \( \theta \), again, the angle at which the incident beam hits the family of planes. \( B_j \), commonly called the temperature factor or \( B \) factor, is related to the r.m.s. displacement of each atom. \( B_j = 8\pi^2 \left\{ \Delta_j^2 \right\} \), where \( \Delta \) is the mean-square
Figure 6. The crystal, indicating the origin of real space, O, is placed in the middle of a sphere of radius $1/\lambda$, called the Ewald’s sphere. The incident beam is referred to as $s_0$, and the diffracted beam as $s$. The origin of the reciprocal space is at $O^*$.

When the end of a reciprocal lattice vector, with length $1/d_{hkl}$, passes through the sphere, here at point P, the waves diffracted from a family of planes with interspacing $d_{hkl}$, will scatter in phase, and produce a diffraction spot on the detector.

Displacement of each atom from its resting position. R.m.s displacements of 1.0 Å and 0.5 Å result in temperature factors of 79 and 20 Å², respectively. The coordinates in units of the fractional unit cell are denoted $(x,y,z)$, and $(h,k,l)$ are the indices of each reflection in reciprocal space.

$F(h,k,l)$ is a complex variable, having both a real and imaginary part, and can therefore be expressed in the form $F(h,k,l) = A(h,k,l) + iB(h,k,l)$. As such, the structure factor has an amplitude which is $\sqrt{A^2 + B^2}$, and a phase of $\arctan(B/A)$. The structure factor can therefore also be written as

$$F(h,k,l) = |F(h,k,l)|e^{i\alpha(h,k,l)}$$

where $\alpha(h,k,l)$ represents the phase.

Since diffraction can occur from both sides of a family of planes, $F(h,k,l)$ and $F(-h,-k,-l)$ will, during normal diffraction, show equal intensity. They will, however, have phases of opposite sign, which will lead to a center of symmetry in the diffraction pattern that is independent of the
symmetry within the crystal. The relation between \( F(h,k,l) \) and \( F(-h,-k,-l) \) is called Friedel’s law. Since the collection of structure factors represents the Fourier transform of the crystal, a back transform has to be used in order to be able to visualize the contents of the unit cell. This second Fourier transform, shown below, is called the electron density equation, \( \rho(x,y,z) \).

\[
\rho(x,y,z) = \frac{1}{V} \sum_{hkl} |F(h,k,l)| \exp[-2\pi i(hx + ky + lz) + i\alpha(h,k,l)]
\]

Here, the summation of all structure factors, will describe the electron density at a specific position \((x,y,z)\). By calculating the electron density on a fine grid the contents of the unit cell can be seen. If Friedel’s law holds, it can be shown that the electron density, unlike the structure factor, is a real number that can be given on an absolute scale as electrons/Å³. During data collection the intensity of each reflection is measured on a detector. Since the intensity is related to the structure factor in the following way

\[
I(h,k,l) \propto |F(h,k,l)|^2
\]

the structure factors can all be calculated on a relative scale based on the data collected. However, in order to calculate the electron density, we also need the phase, \( \alpha(h,k,l) \), for each structure factor. When collecting data, the detector cannot measure the phases, which leads to the so-called “phase-problem” within X-ray crystallography. The importance of solving this problem is shown by the fact that two Nobel Prizes in chemistry have been awarded, the first to Perutz and Kendrew in 1962, and the second to Karle and Hauptman in 1985, for techniques applied to solving the structures of macromolecules and small molecules, respectively.

Today, several methods can be used to solve the phase problem. With the single isomorphous replacement method (SIR), two data sets are collected on isomorphous crystals, which are crystals with identical unit cells. The only difference between the crystals is that one of them is formed by a heavy atom derivative of the protein. Since the heavy atom is a strong scatterer, its position can be determined by using the differences between the structure factors. This in turn can be used for calculating initial phases. To ease the determination of the correct phases, several different derivatives, with heavy atoms at different positions, can be used, and the method is then called multiple isomorphous replacement (MIR). One can also use the anomalous dispersion method. This method is based on the fact that heavy atoms, when exposed to X-rays of a wavelength close to their own absorption edge, absorb a portion of the X-rays, resulting in a scattering factor that is now a complex number instead of a real number. Due to this, there will be a break-
down of Friedel’s law, and $F(h,k,l)$ and $F(-h,-k,-l)$ will no longer show the same intensity. This difference between the Friedel pairs can again be used to derive the position of the heavy atoms, which enables the calculation of initial phases. When collecting data only at one wavelength, at the absorption edge for the heavy atom, in order to maximize the effect, the method is called single wavelength anomalous dispersion (SAD). When collecting data on the same crystal at different wavelengths, in order to vary the anomalous contribution, the method is referred to as multiple wavelength anomalous dispersion (MAD). This technique requires synchrotron radiation, which shows a more even distribution of wavelength intensities than a normal laboratory source. One can also combine the two techniques, using single or multiple isomorphous replacement together with anomalous dispersion, referred to as SIRAS and MIRAS, respectively. However, the method used for solving the structures presented in this thesis is that of molecular replacement (MR), which will be described in detail in the next chapter.

**Molecular replacement**

The principal of molecular replacement is to try to place an already known structure, homologous to the protein of interest, in the unit cell of the unknown structure. By doing this, initial estimates of the phases corresponding to the unknown structure can be calculated from the correctly positioned known structure. In order to understand the method of molecular replacement, it is important to first get acquainted with the Patterson function (Patterson, 1934), $P(u,v,w)$.

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

Just like the electron density equation, it provides a visualization of the three-dimensional density distribution, but here only by summing the intensities, leaving out all information about the phase angles. Patterson space has the same dimensions as real space, so in order to keep them apart, the Patterson space coordinates, which are in fractional units, are denoted $u$, $v$, and $w$. It can also be shown that this function can be written as

$$P(u,v,w) = \sum_{hkl} \rho(x,y,z) \cdot \rho(x + u,y + v,z + w)$$

This function will have peaks at $(u,v,w)$ when there are pairs of atoms separated by the same $(u,v,w)$ vector. By drawing all these vectors from a common origin in the unit cell, as illustrated in Figure 7, the Patterson space is created. Each peak, at the end point of each vector, is proportional to the
product of the atomic numbers of the atoms in-between which the vector is drawn. A unique peak exists for each pair of atoms, and since the vector goes in both directions between atoms, the Patterson space will always be centrosymmetric. In a molecule with \( N \) atoms there will be \( N^2 \) vectors in Patterson space, of which \( N \) peaks will be at the origin. The larger number of peaks in Patterson space than in the electron density increases the chances of overlaps. Also, due to symmetry within the crystal, certain planes, called Harker planes, will contain a large number of peaks. Some of these peaks can be used to find heavy atom positions, crucial for solving the structure by the previously mentioned methods, since they will provide initial estimates of the phases.

![Real space vs. Patterson space](image)

**Figure 7.** Here, a molecule, containing three identical atoms, is visualized in real space. Placed in Patterson space, are all the vectors between these atoms, starting from the origin. The peaks, at the endpoint of each vector, form a pattern, specific for the atom distribution in the unit cell. Since the Patterson function is the self-convolution of the molecule in real space, the triangle shape, formed by the three atoms in real space, can also be seen in Patterson space. The number of images is equal to the number of atoms in the unit cell, and each atom in the pattern is in turn placed at the origin.

As stated above, the molecular replacement method is based on correctly positioning an already known structure in the unit cell of the unknown structure. In order to find the orientation and position of a molecule in the three-dimensional unit cell, a search, with six degrees of freedom, has to be performed. Instead of trying to determine all parameters at the same time, the search can be divided into two steps, a rotation search followed by a translation search, each having three degrees of freedom (Hoppe, 1957; Rossmann et al., 1962). The method is most easily understood with the real space formulation based on the Patterson function. Two similar structures will give rise to approximately the same sets of interatomic vectors. However, these
sets will have different orientations since the molecules have different orientations in their respective unit cell.

In the rotation function, the Patterson map is calculated for the model structure, and the highest peaks are singled out. The vectors corresponding to these peaks are then rotated three-dimensionally over the Patterson map of the unknown structure. In order to increase the chances of finding a solution, the search is performed within a sphere, of such radius that most vectors included in the search are vectors that come from within one molecule, so-called self-vectors. The correctly oriented molecule will then have to be positioned relative to the symmetry axes of the space group, which is performed in the translation function. If several molecules are present in the asymmetric unit, their positions will be defined according to the origin defined by the first symmetry molecule. In order to do this, the cross-vectors, that is, the vectors between molecules, are explored. The two programs here used for solving the structures by molecular replacement are *AMoRe* (Navaza, 1994) and *Phaser* (McCoy et al., 2005), which both are based on a rotation and translation function. In order to find the best solution, the programs use different methods, which are both applied in reciprocal space. In *AMoRe*, the correlation coefficient between the observed structure factors, and the structure factors of the model, is used as an indicator of fit. In *Phaser*, the best solution corresponds to the maximal value for a log-likelihood function, based on the intensities for the model and unknown structure. When the solution is found, a rigid body refinement is often performed, in order to refine both the orientation and position of the model structure in the unit cell.

Clearly, for the molecular replacement method to work, the known structure used in the searches must be reasonably similar to the new structure one wishes to solve. It is known that proteins with significant levels of sequence identity most often have similar three-dimensional structures. This information can be used to find the most similar available structure, which in turn can be used for calculating initial phases for the unknown structure, when placed correctly in the new unit cell. How high the sequence identity between the so-called search model and the unknown structure has to be in order for the method to work varies. If the proteins are expected to have approximately the same size and where there is only one molecule in the asymmetric unit, it could be enough with 20-25% sequence identity, but results are likely to be better where the two proteins have an identity of at least 30-40%. In order to increase the chances of solving a structure by molecular replacement, the model structure may be modified in different ways. One method, which was used for Rv2870c, is to keep all side chains in the model structure that are identical to those in the unknown structure, and change the remaining residues to alanine. By doing so, the search will be biased toward vectors most similar in the two structures.
The first structure solved by molecular replacement, that of seal myoglobin, was reported in 1969 (Tollin, 1969), using the coordinates of the sperm whale myoglobin as a search model (Kendrew, 1958). Since then, the number of structures solved by this method has been constantly increasing, due to the rapid build-up of structures present in the Protein Data Bank (PDB) (Berman et al., 2000). Since the method is usually faster and less complicated than other methods of phasing, it is generally the first approach tested, provided that a suitable search model exists.

Model building and refinement

After solving a structure by molecular replacement, or building a model from an experimentally phased map, the first model will contain a lot of errors. In molecular replacement these come from the different sequences of the search model and the new structure, but may also arise from the presence of different ligands, for example. The process in which these errors are eliminated is called crystallographic refinement. This includes both a computer based optimization and interactive rebuilding of the model by the crystallographer, which are generally repeated in a cyclic process.

So, when a molecular replacement solution is found it is time to replace the structure of the search model with the first complete model of the new structure, so that the correct sequence is present. This can be performed with the program SOD (Kleywegt et al., 2001b). Here, the aligned sequences of the two proteins, as well as the coordinates of the search model placed in the new unit cell, are given as input. The output file, in the form of an O-macro, contains commands for sequentially mutating the residues in the search model into the residues corresponding to the unknown structure. By running the macro in the computer graphics program O (Jones et al., 1991), the first model of the new structure can be visualized. This model is then modified to better fit the electron density. In the studies presented here, O was used for model building and interpretation of the electron density. In the refinement that followed, a maximum-likelihood method was used along with restraints to maintain correct geometry. The goal of crystallographic refinement is to match as closely as possible the amplitudes of the structure factors calculated from the model, \( F_{\text{calc}} \), with those measured experimentally, \( F_{\text{obs}} \). In order to do this, the atomic parameters \( x, y, z \), and \( B \) of the model are slightly shifted in each refinement cycle. The number of observations depends on the resolution. As an example, using the space group and cell parameters for MtDXR solved in space group P2\(_1\), the number of unique reflections are ~ 15000, 46000, and 117000, for a 3, 2, and 1.5 Å resolution data set, respectively. This means that for a model containing ~ 6000 atoms, as in the case of MtDXR, the number of variables will be larger than the number of observations when using diffraction data to 3 Å. This can be overcome to a certain
extent by incorporating different restraints or constraints in order to increase the ratio between the number of observations and parameters. Restraints can be set, for example, to bias the refined model toward standard bond lengths, bond angles, and torsion angles. These are then allowed to vary around the standard values, and thereby effectively add to the number of observations that can be used. If using constraints, which for example again can be set for bond lengths and bond angles, these are taken as rigid, and the number of parameters to be refined is thereby reduced. The program used in the refinement steps in these studies was Refmac5 (Murshudov et al., 1997). In the case of successful rebuilding, each cycle will generate a better map, which can be used to further optimize the model.

In order to see if the refinement is improving the model or not, and how well the model actually matches the data, two different criteria are often checked after each set of refinement cycles. These are the R-factor and free R-factor, commonly denoted \( R \) and \( R_{\text{free}} \) (Brünger, 1992), respectively.

\[
R = \frac{\sum_{hkl} |F_{\text{obs}}| - k|F_{\text{calc}}|}{\sum_{hkl} |F_{\text{obs}}|}
\]

\[
R_{\text{free}} = \frac{\sum_{hkl \in T} |F_{\text{obs}}| - k|F_{\text{calc}}|}{\sum_{hkl \in T} |F_{\text{obs}}|}
\]

Since the observed structure factors are given on an arbitrary scale, which depends on the experimental setup (including the detector system, crystal size, radiation source, etc.) during data collection, and the calculated structure factors are given in real units of electrons, a scale factor has to be used in order to compare the different structure factors. The scale factor, here called \( k \), is often evaluated as a function of \( \sin \theta / \lambda \), and is recalculated in each run of Refmac5. The R-factor shows the percentage difference between the observed and calculated structure factor amplitudes, in relation to the values of the observed structure factor amplitudes. As long as the structure is being improved, \( R \) will decrease for each run. For a typical experiment, \( R \) can be as high as 0.5 (50%) for an unrefined (but correctly solved) structure, and is expected to be 0.15-0.25 (15-25%) for the final structure.

In each refinement, about 5% of the data is from the beginning set aside. \( R_{\text{free}} \) is calculated in the same way as \( R \), but only on this test set, \( T \), of
structure factors never used in the refinement. By comparing $R$ and $R_{\text{free}}$ the reliability of the model can be estimated, in a standard cross-validation procedure. The R-factor can always be lowered, even artificially, by introducing more parameters (Kleywegt et al., 1995). However, $R_{\text{free}}$ will not be affected by this and can therefore serve as a useful indicator for the progress of the refinement. For a well-performed refinement the difference between $R$ and $R_{\text{free}}$ is typically observed to be lower than 0.05 (5%) (Kleywegt et al., 2002). Of course, $R_{\text{free}}$ by itself also gives a very good estimate of the structure quality.

Validation

In order to determine the quality of a protein structure there are other criteria, besides $R$ and $R_{\text{free}}$, which should be considered. One of the most important is the resolution of the crystallographic data. The higher the resolution of the data, the more details are shown in the structure. When it comes to trying to learn more about how substrates or cofactors are bound, and in order to draw conclusions about a mechanism, the resolution is crucial. Also, the completeness of the data, that is the fraction of possible $F(h,k,l)$s measured, is important, as well as the multiplicity, the number of times each $F(h,k,l)$ is measured, to ensure the accuracy of the data. However, the fact that the crystal is being damaged by radiation during the experiment limits the number of observations that can reasonably be collected. The goal is to ensure that you have as much of the diffraction data as possible, and that the estimation of the structure factor amplitudes is as correct as possible. Therefore, the signal to noise in the experimental data is carefully evaluated during the image processing stage. The limit for $\langle I \rangle / \langle \sigma I \rangle$, that is the intensity over the standard deviation of the intensity, is usually chosen to be over 2 in the highest resolution shell.

Other criteria important to consider for the fully refined structure, are the r.m.s. deviations from ideal bond lengths and bond angles. According to commonly accepted limits these should preferably not be more than 0.02 Å and 2°, respectively. The $B$ factors can either indicate dynamics in the structure or errors in the model, or both. The overall $B$ factor usually reflects the quality of the crystallographic data. Low-resolution studies generally produce models with high overall $B$ factors. An exception is if only a low resolution data set was collected although the crystal diffracted further.

In the protein backbone, the angles of rotation around the bonds on either side of each $C_{\alpha}$ are called phi ($\Phi$) and psi ($\Psi$), for the N-$C_{\alpha}$ and $C_{\alpha}$-C bond, respectively. The value of the torsion angle for $\Phi$ is defined by looking from N-$C_{\alpha}$, and evaluating the angle that the C$_{\alpha}$-N bond makes to the $C_{\alpha}$-C bond. For $\Psi$ the value is instead defined by looking from $C_{\alpha}$-C, and evaluating the
angle that the N-C\textsubscript{α} bond makes with the C-N\textsubscript{−1} bond. A positive rotation is in a clockwise direction when looking from C\textsubscript{α} towards N, or from C\textsubscript{α} towards C. In a protein structure, due to collisions between different residues, some combinations of Φ and Ψ are not allowed. By plotting Ψ against Φ, going from −180° to +180°, a diagram, called the Ramachandran plot is created, showing the distribution of Φ and Ψ pairs. The quality of a structure can be evaluated by looking at the corresponding Ramachandran plot, and estimating how many of the pairs are positioned within allowed regions, and how many are positioned in the disallowed regions. In the strict-boundary Ramachandran plot (Kleywegt \textit{et al.}, 1996), presented for the structures included in this thesis, 95% of the pairs should be in the core regions, thereby allowing up to 5% of the pairs to be outliers, for a typical 2 Å resolution structure.
The tubercle bacillus

Robert Koch and the discovery of *M. tuberculosis*

In 1905, Robert Koch was awarded the Nobel Prize, mainly for his discovery that tuberculosis is an infectious disease caused by the pathogen *Mycobacterium tuberculosis*. The results of his studies were published in an article from 1882 entitled “The etiology of tuberculosis” (Koch, 1882), in which he describes the experiments leading up to this conclusion. The tests performed included showing that the bacteria could be found, mainly clustered, within host cells, that the bacteria could be isolated and cultured, and that this culture in turn was able to cause the disease in experimental animals. At this time, the main opinion within the medical community was that all diseases were caused by malfunction of our own cells (Virchow, 1871), which further highlights the importance of Koch’s findings. A few years later, between 1890-1891, Koch also tried to find a cure for tuberculosis. The material tested was described as a glycerol extract of tubercle bacilli; however it did not cure the disease, but instead had the potential of being used as a diagnostic test (Koch, 1890; 1891a; b; c). The so-called tuberculin test, which is a delayed-type hypersensitivity test, is still used today for the diagnosis of tuberculosis. A useful review may be found at (Kaufmann et al., 2005).

Bacillus characteristics and route of infection

The Gram-positive, rod-shaped, *M. tuberculosis* bacillus is characterized by its slow growth, its complex cell envelope, and its ability to lie dormant inside the host and reactivate decades after the initial infection (Brennan et al., 1995; Flynn et al., 2001). Outside the bacterium’s plasma membrane an envelope is found that is composed of a covalently linked complex of peptidoglycan, arabinogalactan, and mycolic acid (Brennan, 1995), see Figure 8. The long chains of mycolic acids are thought to form a layer around the bacteria, which functions as an extra permeability barrier. The outermost layer is composed of acyl glycolipids and mannose-capped lipoarabinomannan (Brennan, 1995; Karakousis et al., 2004; Riley, 2006). Under certain conditions an additional layer, referred to as the capsule, containing glucan, arabinomannan and smaller amounts of proteins and lipids, is present (Stokes et al., 2004). The hydrophobic cell envelope of *M. tuberculosis* explains why it
shows almost 100 times lower permeability than that observed for *E. coli* (Chambers *et al.*, 1995). The bacteria however, have a functional Sec pathway (Pugsley, 1993), capable of transporting proteins in an unfolded state, and it also seems to have a functional Tat pathway (Sargent *et al.*, 1998), which instead can transport proteins in a folded state.

![Figure 8. A schematic picture of the *M. tuberculosis* cell membrane.](image)

When a person with active pulmonary tuberculosis sneezes, coughs, or just talks, droplets containing bacilli can be transferred to another person. The bacilli then make their way into the lungs, where they are absorbed by pulmonary alveolar macrophages. After infection, the immune response is triggered and cytokines, such as tumor necrosis factor α (TNF-α), and different chemokines are produced, which in turn initiate the recruitment of lymphocytes, for example natural killer T-cells (NK T-cells) and CD4⁺ and CD8⁺ T-cells. These lymphocytes regulate the production of interferon γ (IFN-γ), which in turn regulates the immune response, and thereby the formation of a stable granuloma (Dheda *et al.*, 2005; Fenhalls *et al.*, 2002; Kaplan *et al.*, 2003; Tully *et al.*, 2005; Ulrichs *et al.*, 2005). A ring of lymphocytes has usually formed around the infected macrophage after a few weeks, when the immune response has reached its peak, for a summary see (Cardona *et al.*, 2003). When the specific immunity is triggered, foamy macrophages will form yet another layer around the infected macrophage (Cardona, 2003); the cellular mass so formed is called a “tubercle” or granuloma. The granuloma will fail to form in the absence of an INF-γ secreting CD4⁺ T-cell response (Saunders *et al.*, 2007), which can be observed in patients that are also infected by HIV, in which case the T-cells fail to control the *M. tuberculosis* infection (Saunders, 2007).

Even though pulmonary tuberculosis is the most commonly known form of tuberculosis, the bacteria can also disseminate to other areas of the body.
such as the lymph nodes, spleen, liver, skeleton, or the central nervous system. This is thought to occur either by entry into alveolar macrophages or by interaction with epithelial cells (Pethe et al., 2001).

The *M. tuberculosis* genome

Cole *et al.* published the complete genome sequence of the *M. tuberculosis* H37Rv circular chromosome in 1998 (Cole, 1998). Among the 4,411,529 base pairs, 3,924 putative protein-coding sequences were identified, for which 40% were predicted to have a specific function. By optimizing the search, by decreasing the minimal predicted gene length, and also by considering the codon usage, Cole and co-workers could in 2002 revise these numbers to 3,995 protein-coding sequences, for which now 52% could be assigned a function (Camus *et al.*, 2002). All of the coding sequences of the *M. tuberculosis* genome have been assigned an Rv number. In their original paper from 1998, Cole *et al.* showed that 16% of the genes are predicted to be specific for *M. tuberculosis*, since these do not resemble any currently known proteins. They also showed that the G+C content of the genome is as high as 65.6%.

Among the genes of the *M. tuberculosis* genome, 26% are defined as conserved hypothetical proteins, that is they are open reading frames that have equivalents in other species but little else is known, 22% participate in intermediary metabolism and respiration, 18% are thought to participate in the formation of the cell wall or in the cell wall processes, and 7% of the genes have an unknown function (Camus, 2002).

The genome of *M. bovis*, the bacterium causing tuberculosis in cattle, has also been sequenced (Garnier *et al.*, 2003). The single circular chromosome of 4,345,492 base pairs has a G+C content of 65.63%, very similar to that of *M. tuberculosis*, and in fact 99.95% of the genomes are similar at the nucleotide level. The largest variations are seen among genes coding for cell wall, and secreted proteins. From a phylogenetic standpoint, it has been proposed that *M. tuberculosis* either infected cattle, giving rise to *M. bovis*, or that they both have been derived from a founder strain, where *M. tuberculosis* is most similar to the ancestor (Brosch *et al.*, 2002; Gutierrez *et al.*, 2005; Mostowy *et al.*, 2002).

*M. lepra*, which causes leprosy, is another of the major human pathogens. When comparing the genome to that of *M. tuberculosis* (Cole *et al.*, 2001), it is easy to note the differences between them. *M. lepra*, with a genome of 3,228,203 base pairs, lack more than 1,100,000 base pairs which are present in *M. tuberculosis*. It also has more than 1,000 pseudogenes, that is, genes that are here present as inactive reading frames, but that have functional counterparts in *M. tuberculosis*. Due to these differences, the *M. lepra* genome can be used as a means of determining the degree of importance
among the *M. tuberculosis* genes. The genes still present, and active, in *M. leprae* are suspected to be essential in *M. tuberculosis*.

*M. smegmatis*, which is homologous to *M. tuberculosis*, but which does not function as a pathogen, provides a model system for the tuberculosis pathogen. The so-far incomplete sequencing of the *M. smegmatis* genome (http://www.tigr.org) is used for *in silico* comparisons, and the genes and pathways similar between the two bacteria can more easily be studied in *M. smegmatis*. This due to the shorter time required for growth of the bacteria, as well as the fact that the non-pathogenic nature of the organism simplifies handling procedures.

**Diagnostics, vaccines, and current drugs**

Several methods are today used in order to diagnose tuberculosis. The previously mentioned delayed-type hypersensitivity test, more commonly referred to as the tuberculin skin test (Jasmer *et al.*, 2002), is foremost used in diagnosing a latent infection (Mazurek *et al.*, 2001). This can also be accomplished by measuring the whole-blood INF-γ release. Other ways of detecting a current infection include searching for bacteria in a sputum sample (CDC, 2000), or performing a chest X-ray.

The only current vaccine towards tuberculosis, which has been in use since the 1920s, is *M. bovis bacillus Calmette-Guérin*, BCG, which is an attenuated strain of *M. bovis* (Calmette, 1927). The vaccine has shown to protect children against disseminated tuberculosis. However, it does not efficiently prevent pulmonary tuberculosis (CDC, 1996), which is the most common, and also most contagious form of the disease.

The five “first line” antituberculosis drugs are isoniazid, rifampin, ethambutol, streptomycin and pyrazinamide. Isoniazid and rifampin, presented in 1952 and 1966 (Maggi *et al.*, 1966; Robitzek *et al.*, 1952), respectively, were the first drugs ever to be used against tuberculosis, and even today the most common treatment consists of a regimen of these drugs in combination with pyrazinamide and either ethambutol or streptomycin (Bass *et al.*, 1994). Isoniazid and ethambutol both target the cell wall biosynthesis, by inhibiting the mycolic acid and arabinogalactan synthesis, respectively (Mikusova *et al.*, 1995; Takayama *et al.*, 1972; Winder *et al.*, 1970). Rifampin inhibits transcription (McClure *et al.*, 1978), and streptomycin has been shown to inhibit protein synthesis, by interacting with the 30S subunit of the ribosome (Garvin *et al.*, 1974). Pyrazinamide is a prodrug, which inside the cell is converted by the bacteria’s own enzymes, into its active form pyrazinoic acid (Zhang *et al.*, 2003). The active compound decreases the membrane potential and thereby inhibits the transport of nutrients. Two of the “second line” drugs are kanamycin and ethionamide, which inhibit protein and my-
coli acid synthesis, respectively (Morlock et al., 2003; Suzuki et al., 1998; Taniguchi et al., 1997).

*M. tuberculosis* has been shown to be naturally resistant to various antibiotics (Cole et al., 1995), most probably due to the permeability barrier presented by the bacteria’s cell envelope (Brennan, 1995). Multi-drug resistant tuberculosis (MDR-TB) is, on the World Health Organization (WHO) website, described as a disease showing resistance towards at least isoniazid and rifampin. Today, point mutations or deletions have been found that lead to resistance for all first line drugs (Ramaswamy et al., 1998). The Directly Observed Therapy-Short-course, DOTS, strategy (Bass, 1994; CDC, 1995) has in many cases been a success according to the WHO. Here, some of the aims are to strengthen the health care system, provide high quality services, and engage health-care providers. However, the increased prevalence of extensively drug resistant tuberculosis (XDR-TB), defined as a disease also showing resistance towards “second line” drugs, clearly emphasize the seriousness of resistance, and the acute need for new drugs.

The need for new drugs is also due to the most unpleasant side effects caused by the existing drugs (Addington, 1979). Rifampin can give flu-like symptoms and sometimes colors the urine, tears and saliva orange. Isoniazid can cause liver disease, dizziness, nausea and muscular twitching, and ethambutol can blur one’s vision. Needless to say, a drug should not make the patient feel worse, but better, so that the treatment is completed, which in turn helps to prevent the emergence of drug resistant strains.

**Essentiality and target selection**

The reasons for choosing a particular target enzyme can vary. Protein structures in particular in complex with different substrates, cofactors or inhibitors, provide not only structural information but also provide us with mechanistic insights. By comparing different complexes the conformational changes associated with catalysis can increase our knowledge of the basic biology associated with a particular function or pathway. A structure can also provide additional information, crucial when it comes to drug design. One may locate catalytic residues, specific interactions, cavities within the protein, and identify structural conservation to related proteins. With the goal of structure-based drug design aimed at eradicating diseases affecting mankind, several additional aspects have to be considered. Essentiality of the protein for the pathogen is perhaps one of the most important, in combination with drugability. This takes into account issues such as cellular location of the target and toxicity of the compounds as well as the existence of a good assay or potential transport pathway for uptake. It is also considered favourable if there are no human homologues.
So how can essentiality be determined? By performing knockouts, where the selected gene is silenced, and then complementation studies, the bacteria’s response can be observed. However, this is a difficult approach experimentally. The slow growing nature of the *M. tuberculosis* bacteria also adds to the complication. During the last few years, several other types of studies have been performed in order to answer the question of essentiality in *M. tuberculosis*. One approach includes simulating different stages of the pathogen’s life cycle and looking at the variation in gene expression. In a paper by Betts, a nutrient starvation model, thought to mimic the bacteria’s latent period, is presented (Betts et al., 2002). The bacteria are here found in a stage of no or little replication. The respiration rate is low, but the bacteria are still able to maintain long-term viability. The results were analyzed by 2D gel electrophoresis and microarray analysis, and the enzymes differentially expressed under these conditions were identified. In other studies, Sassetti and Rubin performed a transposon site hybridization (TraSH) experiment (Sassetti et al., 2001; 2003), in which high-density insertional mutagenesis was combined with microarray mapping in order to identify genes required for growth under different conditions. Here, the genes proposed to be essential were those that could not sustain the transposon insertion without killing the bacteria. These genes were therefore not present in the collection of genes presented on the microarray. The genes were all given a Sassetti score based on the level of expression, which can be used as a measure of essentiality under *in vitro* and *in vivo* conditions. The idea behind the study presented by Stewart (Stewart et al., 2002) is that heat-shock proteins are induced in both pathogen and host during infection. By identifying which genes were up regulated during heat shock, proteins thought to be essential for infection could be identified, using a combination of targeted mutagenesis and whole-genome expression profiling. Lamichhane also performed an *in silico* TraSH experiment in which statistical methods together with previous data were used to evaluate essentiality (Lamichhane et al., 2003).

These, and other studies, have been made with the goal of as accurately as possible mimicking different stages of the bacteria’s life cycle, and identifying genes that seem to be essential. However, when comparing the above studies it can be seen that very few of the top hits of each study can be found in any of the others. Therefore, my personal opinion is that one cannot blindly trust these “hits”, and that it is also important to focus on important pathways, in which both individual enzymes, or the pathway as a whole, differ from any human homologues.
Peptidyl-prolyl cis-trans isomerase A (Paper I)

*M. tuberculosis* PpiA and PpiB

As a part of our EU collaboration, a high throughput method for expressing and crystallizing *M. tuberculosis* proteins was set up in Marseille. However, the difficulty of expressing the proteins in a soluble form was soon made clear. Therefore, one of our first goals within the structural biology group at Uppsala University was to test different proteins to see which ones could be successfully expressed, and which proteins would exhibit problems. Our choices were also influenced by available data on essentiality and changes in expression during stress, see below. Two of the enzymes chosen at an early stage were peptidyl-prolyl cis-trans isomerases (Ppis) A and B (Rv0009 and Rv2582, respectively). Ppis (EC 5.2.1.8), also known as rotamases or cyclophilins, catalyze cis-trans isomerization of peptide bonds, preferring those preceding proline residues (Fischer *et al.*, 1989; Takahashi *et al.*, 1989), see Figure 9. The enzymes can be found in a large variety of organisms such as bacteria, plants, and mammals, sometimes as single domain proteins, but also as components in larger complexes (Galat *et al.*, 1995; Schmid, 2001). Ppis accelerate protein folding both *in vivo* and *in vitro*, they bind to and mediate the biological effects of the immunosuppressive agent cyclosporine A (Handscharumacher *et al.*, 1984), and they participate in cell surface recognition (Anderson *et al.*, 1993) and heat-shock response (Sykes *et al.*, 1993).

![Figure 9. A schematic cis-trans isomerization, catalysed by a Ppi, is here illustrated for the hexapeptide HAGPIA. The two residues involved in the isomerization are shown fully, while the others are only indicated by their three-letter code. The bonds indicating cis or trans for the peptide bond are shown in bold.](image-url)
As mentioned, *M. tuberculosis* has two different Ppis, *MtPpiA* and *MtPpiB*. Since *MtPpiA* lacks an obvious signal sequence or membrane spanning sequence it is presumed to reside in the cytoplasm. The sequence of *MtPpiB* on the other hand contains a membrane anchor, positioned in the ~140 residues long stretch preceding the catalytic domain, thereby most probably placing the active site on the extra-cellular surface. The *MtPpiB* sequence further shows that it is different from that of other Ppis. While the catalytic arginine is conserved, approximately one-third of the residues lining the active site are not, indicating that its specificity can be quite different from that of other Ppis, including the human enzyme. The difference between *MtPpiB* and *MtPpiA* regarding both their sequences, and most probably also structures, together with the different locations in the bacteria, indicate that they can have different functions, perhaps during different stages of the bacteria’s life cycle. The *MtPpiB* gene was reported to be essential in the transposon site hybridization study presented by Sassetti and Rubin (Sassetti, 2003), while *MtPpiA* was not. Instead, *MtPpiA* is up regulated during heat-shock, as reported by Stewart (Stewart, 2002), implying that it may be related to the heat shock response, and possibly virulence. *MtPpiA* is also down regulated during iron depletion (Wong *et al.*, 1999), indicating that it is iron-regulated.

In this thesis, the structure of *MtPpiA* is presented. Although several attempts were made to solve the structures of both *M. tuberculosis* Ppis, our efforts were fruitless in the case of *MtPpiB*. Several constructs were produced, containing different lengths of the enzyme, all of them totally excluding the predicted membrane spanning section. No soluble material was however produced.

**Cloning, expression, and purification**

The sequence corresponding to the full-length *MtPpiA* (molecular weight 19.2 kDa), originating from *M. tuberculosis* strain H37Rv (Cole, 1998) was amplified by PCR using the primers 5’-ATGGCAGACTGTGATTCCGTGAC-3′ (forward) and 5’-CTAGGAGATGGTGATCGACTCG-3′ (reverse), and *Taq* DNA polymerase (Roche). A second PCR was performed, using the product from the first PCR as a template, and now the forward primer 5’-ATGGCCCATCATCATCATCATCATGTTCTGTGAGACTGTGATTCCGTGAC-3′, thereby introducing an N-terminal His₆-tag. The pCR T7/CT TOPO TA® Exression kit (Invitrogen) was used for ligating the PCR product into the pCR®T7/CT-TOPO® vector. Cloning was performed in *E. coli* TOP10F’ cells (Invitrogen). Positive clones were selected for by growth on Luria agar plates containing 50 µg/ml ampicillin, and plasmids were isolated after the QIAprep® Spin Miniprep kit protocol (Qiagen). An analytical PCR was performed using the pCR T7/CT TOPO TA® Exression kit (Inviti-
trogen), with the v5, C-terminal, primer, and the His6 forward primer. The positive clone was verified by DNA sequence analysis (Uppsala Genome Center, Rudbeck Laboratory). Expression was carried out in BL21-AI™ cells (Invitrogen) at 37 °C. At OD550=0.7-1.0, the cells were moved to 22 °C, induced with 0.001% (w/v) arabinose, and growth was continued for 2 h. The cells were harvested and washed with 1×SSPE buffer (150 mM NaCl, 10 mM Na2HPO4 pH 7.5, 1 mM EDTA), and stored at -20 °C.

For purification, thawed cells were treated with lysis buffer (50 mM NaH2PO4 pH 8.0, 300 mM NaCl, 10 mM imidazole, 4% glycerol) with 0.01 mg/ml RNase, 0.02 mg/ml DNase and lysed with a Constant Cell Disruptor (Constant Systems Ltd) operated at 250 MPa. The cell lysate, containing soluble material of MtPpiA, was incubated for 30 min at 4 °C with nickel-NTA agarose slurry (Qiagen). The resin was washed, at room temperature, with ten column volumes of lysis buffer containing 20 mM imidazole and the protein was then eluted with four column volumes of the same buffer containing 250 mM imidazole. The protein was further purified by size-exclusion chromatography on a HiLoad 16/60 Superdex 75 column (Amersham Pharmacia Biotech) using a buffer containing 150 mM NaCl and 20 mM Tris-HCl pH 7.5. Fractions containing MtPpiA were pooled and desalted using a PD10 column (Amersham Biosciences) with a solution of 10 mM 2-mercaptoethanol and 20 mM Tris pH 7.5. The protein was concentrated to 29 mg/ml (based on the calculated absorbance of 0.252 for a 1 mg/ml solution at 280 nm) using a Vivaspin concentrator (Vivascience, molecular-weight cutoff 10 kDa).

### Activity determination

The cis → trans isomerisation catalyzed by MtPpiA can be evaluated in a spectrophotometric assay (Fischer et al., 1984), using the chromogenic peptide N-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide (Sigma), and α-chymotrypsin. α-chymotrypsin cleaves the peptide bond X-Pro only if it is in a trans position. When the peptide is cleaved the nitroanilide part will give rise to a coloured product, which can be detected at 390 nm, as an increase in absorbance.

In order to increase the fraction of the cis isomer, a peptide solution (7.8 mM) in trifluorethanol with 0.45 M LiCl, was prepared the day prior to the activity measurement. Each assay contained 910 μl 0.1 M Tris-HCl pH 8.0 (kept at 15 °C), 15 μl 600 μM α-chymotrypsin and 30 μl of MtPpiA (at 1.7 μM, 0.49 μM or 0.33 μM), which were mixed and pre-equilibrated in a cuvette at 15 °C for 2 min. The reaction was initiated by adding 10 μl of the peptide solution, resulting in a final concentration of 78μM. The cis → trans isomerization, both the spontaneous and that catalyzed by MtPpiA, was then followed by the increase in absorbance at 390 nm at 15 °C using a DU® 640
spectrophotometer (Beckman). Measurements were made every 0.5 s during a 3 min period.

The kinetic properties for many enzymes can be described by the Michaelis-Menten equation,

\[ v = \frac{V_{\text{max}} [S]}{K_m + [S]} \]

where \( v \) is the rate, \( V_{\text{max}} \) the maximal rate, \([S]\) the substrate concentration, and \( K_m \) the Michaelis constant, which is defined as the substrate concentration at which the reaction rate is half of its maximal value. In many instances \( K_m \) gives a good approximation of the dissociation constant \( K_D \), which describes the affinity of the substrate to the enzyme. If the substrate concentration is much lower than the Michaelis constant, that is \([S] \ll K_m\), the equation can instead be written as

\[ v = \frac{V_{\text{max}} [S]}{K_m} \]

By using the fact that \( v \) is the change in substrate concentration over time, the equation can also be written on the form

\[ v = \frac{V_{\text{max}} [S]}{K_m} = k_{\text{obs}} [S] = -\frac{d[S]}{dt} \]

where \( k_{\text{obs}} \) is a first order rate constant. By solving the integral equation the following expression is valid

\[ \ln [S] = -k_{\text{obs}} t + \ln S_0 \]

By plotting the natural logarithm of the substrate concentration against time, \(-k_{\text{obs}}\) is retrieved as the slope. By also knowing that \( V_{\text{max}} = k_{\text{cat}} [E] \), where \( k_{\text{cat}} \) represents the turn-over number, which shows the enzymes catalytic efficiency, and \([E]\) is the enzyme concentration, and inserting it in the equation above, the following equation is derived

\[ k_{\text{obs}} = \frac{k_{\text{cat}} [E]}{K_m} \]
By now plotting $k_{obs}$ against $[E]$, a measurement of the substrate specificity, in the form of a $k_{cat}/K_m$ value, can be estimated from the slope.

Since the assay used for MtPpiA not only measures the catalyzed cis $\rightarrow$ trans isomerization, but also the naturally occurring isomerization, this has to be taken under consideration when calculating the substrate concentration associated with the catalyzed reaction. The final absorbance, $A_{final}$, was estimated from each curve, and the absorbance at each time point $t$, $A_t$, was subtracted from this value, resulting in $\Delta A = A_{final} - A_t$, which always has a positive value. Since the change in absorbance is directly equivalent to the change in substrate concentration, $\Delta A$ can be used instead of $[S]$ in the following calculations. When plotting $\ln A$ versus time, the slope will be equal to $-k_{obs}$, in the linear region, which here was during the first 10 seconds of the reaction. Since $k_{obs} = (k_{cat}/K_m) \times [MtPpiA]$, a $k_{cat}/K_m$ value can be estimated from the slope, when plotting $k_{obs}$ versus the corresponding enzyme concentration, $[MtPpiA]$, for each reaction.

The $k_{cat}/K_m$ value for MtPpiA was estimated to be $2.0 \times 10^6$ M$^{-1}$s$^{-1}$, which is similar to the values of $7.9 \times 10^6$ M$^{-1}$s$^{-1}$ and $1.4 \times 10^7$ M$^{-1}$s$^{-1}$ previously published for Brugia malayi (Taylor et al., 1998) and the human PpiA (Liu et al., 1990), respectively. Our calculations were made under the assumption that $[S] \ll K_m$. While $K_m$ values are not generally available for Ppis, a $k_{cat}$ value of 9000 s$^{-1}$ has been estimated for the human PpiA, using a different peptide (Eisenmesser et al., 2002). If taken together with an available $k_{cat}/K_m$ value, estimated for the same enzyme (Liu, 1990), a $K_m$ value of $\approx 1$ mM is suggested. If the $k_{cat}$ value for MtPpiA is similar, it would also have a $K_m$ value in the mM range. Since the final concentration of peptide in the experiment was 78 μM, the $[S] \ll K_m$ condition is most probably fulfilled.

The MtPpiA crystals

The crystallization screens CryoI/II (Emerald BioSystems) and Index (Hampton Research) were set up with MtPpiA co-crystallized with a hexapeptide of sequence HAGPIA (Vajdos et al., 1997). Needles of crystal-line material were produced in conditions 13, 25, 77, and 95 of the CryoI/II screen. All of these contained either 30-50% of PEG200 or PEG400, and had a final pH between 6.1 and 8.7. After optimization, the best condition, number 13 (30% PEG200, 100 mM MES, 5% PEG3000, final pH 6.1), was chosen for further experiments. Sitting drops, containing 2 μl protein (29 mg/ml in the buffer used during concentration), with a final concentration of 1 mM of the peptide dissolved in dimethyl sulfoxide, and 2 μl of the crystallization
buffer, were set up at room temperature. Needle-like crystals, forming small “flowers”, see Figure 10, appeared within a few weeks.

The number of “flowers” was however high, preventing individual crystals from growing to a size that is preferred for data collection. Streak-seeding was therefore used in order to optimize the crystallization by reducing the number of “flowers” produced in each drop. Unfortunately, this treatment resulted in the same number of crystals as previously. By instead only dipping the acupuncture needle, with seeds, once into the drop, the number of nucleation sites could be reduced to only one or two in each drop. Crystals grew, now in hanging drops, to a size of 0.05 × 0.05 × 0.4 mm over a period of 2-3 months. To prepare for data collection, single crystals were isolated with the help of an acupuncture needle. Prior to flash cooling, the crystals were placed for 12 h in a drop of crystallization buffer complemented with 1 mM of peptide, in order to favor the peptide binding.

From data collection to final model

X-ray data were collected under cryoconditions at beamline ID14-2 at the European Synchrotron Radiation Facility. Diffraction data were indexed and integrated using MOSFLM (Leslie, 1999), and processed with SCALA (Evans, 1993), as implemented in the CCP4 program suite (CCP4, 1994). The data were processed in the primitive hexagonal space groups P31, P312, P3121, P61, and P622; P31 gave by far the best result. At this stage the data were also processed in space group P32, which gave similar results as that for P31. When instead using the program DENZO (Otwinowski et al., 1997), now also testing space groups P321 and P62, the results also indicated that P31 or P32 was the correct choice of space group. The Matthews coefficient suggested that there were either two (3.1 Å³Da⁻¹, with 60% solvent) or three (2.1 Å³Da⁻¹, with 40% solvent) molecules in the asymmetric unit. This first data set showed high anisotropy, and the selected resolution limit was 3.4 Å.
The structure was solved by molecular replacement in AMoRe (Navaza, 1994), using the structure of the human PpiA (PDB code 1AWR, 37% sequence identity) (Vajdos, 1997), as a search model. Both P3₁ and P3₂ were tested, which clearly showed that P3₁ was the correct space group. Two molecules, related by pure translational symmetry, were present in the asymmetric unit. The program SOD (Kleywegt, 2001b) was used together with the sequence for MtPpiA, in order to build the initial model. Structural refinement was performed with NCSREF and REFMAC5 (Murshudov, 1997), and rebuilding with the program O (Jones, 1991).

A second data set was collected, which still showed high anisotropy, but could be processed to a resolution of 2.6 Å. In the final rounds of refinement, using REFMAC5, non-crystallographic symmetry restraints were used, they were not used earlier, because differences were evident between the two molecules in the asymmetric unit. Water molecules were added after analyzing the results from the ARP/wARP (Lamzin et al., 1993) water building routine. Different weights were tested in the refinement in order to find the optimal balance between experimental data and stereochemistry. The final model, which has been deposited in the PDB with accession code 1W74, has an $R$ and $R_{free}$ value of 21.3 and 22.9%, respectively, and shows 3.4% outliers using a strict boundary Ramachandran plot (Kleywegt, 1996).

**Overall structure and active site**

The overall MtPpiA structure consists of an eight-stranded antiparallel β-barrel with one α-helix on each side, see Figure 11. This is consistent with previous Ppi structures, like the human PpiA (Vajdos, 1997) and the Ppi from B. malayi (Taylor, 1998). Previous studies have also shown that the active site is positioned on one side of the β-barrel (Fesik et al., 1991). Although MtPpiA was co-crystallized with the hexapeptide HAGPIA, which is derived from the HIV capsid protein sequence (Vajdos, 1997), no strong density could be seen for the peptide, except at the central proline residue. By super-positioning the human enzyme in complex with the HAGPIA peptide, the expected interaction in the MtPpiA active site could be determined. Four of the predicted active site residues, Phe78, Gln126, Leu137 and His141 are positioned in loop regions, while the catalytic arginine, Arg73, is placed in one of the β-strands in close proximity to the active site surface.
Figure 11. The overall structure of MtPpiA is illustrated in a ribbon drawing. The two α-helices are colored light grey, while the eight β-strands forming the antiparallel β-barrel are shown in a darker grey. Of these, three are shown in a still darker color, indicating the active site surface.

Concluding remarks

Although MtPpiA, with its extra insert and different N-terminal segment represents a variant of Ppi not found among the structures present in the PDB, it is still not suitable as a drug target in *M. tuberculosis*. Almost all residues in the active site are conserved in comparison to those in the human Ppis, and so the chances of finding selective inhibitors did not seem great. As stated previously, the initial goal was to solve the structure of both *M. tuberculosis* Ppis. The object was not only to compare the overall structures, in particular the active sites, but also to be able to learn how the differences in the sequence might be related to differences between the substrates that they can bind. Since problems were encountered in production of soluble MtPpiB, this goal was not achieved.
1-deoxy-D-xylulose-5-phosphate reductoisomerase (Paper II and III)

The mevalonate and non-mevalonate pathways

Isopentenyl diphosphate (IPP), as well as dimethylallyl diphosphate, are the precursors of the highly diversified group of isoprenoids, essential to all living organisms (Sacchettini et al., 1997). Among these are compounds serving, as reproductive hormones and photo-protective agents, and ones that participate in signaling networks, like steroid hormones, carotenoids, and ubiquinone. Some of the isoprenoids, of which one example is polyprenyl phosphate, also help to synthesize the M. tuberculosis cell wall (Crick et al., 2000). In eukaryotes and Archaea (Boucher et al., 2000), IPP is formed through the mevalonate pathway (summarized in (Spurgeon et al., 1981)), starting from acetyl-CoA, see Figure 12A. In plants, protozoa, green algae, and many bacteria, pyruvate and D-glyceraldehyde 3-phosphate are instead used for the production of IPP, within the so-called non-mevalonate, or deoxy-xylulose 5-phosphate/methylerythritol 4-phosphate, pathway (Lichtenthaler, 1999; Rohmer, 1999; Rohmer et al., 1993), see Figure 12B.

The first step in the non-mevalonate pathway is performed by the enzyme 1-deoxy-D-xylulose-5-phosphate synthase (DXS, Rv2682c), in which pyruvate and D-glyceraldehyde 3-phosphate are condensed to form 1-deoxy-D-xylulose 5-phosphate (DXP). In the second step, which is NADPH-dependent and requires the presence of a divalent cation such as Mg$^{2+}$, Co$^{2+}$, or Mn$^{2+}$ (Takahashi et al., 1998), the rearrangement and reduction of DXP is catalyzed by 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR, also referred to as IspC, Rv2870c, EC 1.1.1.267), resulting in the product 2-C-methyl-D-erythritol 4-phosphate (MEP). The following steps of the pathway are performed by the enzymes 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase (IspD, Rv3582c), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (IspE, Rv1011), 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase (IspF, Rv3581c), and 4-hydroxy-3-methylbut-2-en-1-yl-diphosphate synthase (IspG Rv2868c). The last step, which again involves an NADPH-dependent reduction of (E)-4-hydroxy-3-methylbut-2-en-1-yl diphosphate to form IPP and dimethylallyl diphosphate, is catalyzed, separately, by two enzymes, 4-hydroxy-3-methylbut-2-enyl-diphosphate reductase 1 and 2 (IspH1, Rv1110 and IspH2, Rv3382c).
Figure 12. The (A) mevalonate and (B) non-mevalonate pathway of isoprenoid biosynthesis. Abbreviations used: ADP, adenosine diphosphate; ATP, adenosine triphosphate; CDP-ME, 4-(cytidine 5’-diphospho) 2-C-methyl-D-erythritol; CDP-MEP, 2-phospho-CDP-ME; CMP, Cytidine monophosphate; CoA, Coenzyme A; CTP, Cytidine triphosphate; DMAPP, Dimethylallyl pyrophosphate; DPMD, Dipospho mevalonate decarboxylase; DXP, 1-deoxy-D-xylulose 5-phosphate; DXR, DXP reductoisomerase; DXS, DXP synthase; HMDP, (E)-4-hydroxy-3-methylbut-2-en-1-yl diphosphate; HMGR, 3-hydroxy-3-methylglutaryl-CoA reductase; HMGS, 3-hydroxy-3-methylglutaryl-CoA synthase; IPP, Isopentenyl diphosphate; IspD, 4-diphosphocytidyl-2-C-methyl-D-erythritol synthase; IspE, 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; IspF, 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase; IspG, 4-hydroxy-3-methylbut-2-en-1-yl-diphosphate synthase; IspH 1/2, 4-hydroxy-3-methylbut-2-en-1-yl-diphosphate reductase 1/2; MECDP, 2-C-methyl-D-erythritol 2,4-cyclodiphosphate; MEP, 2-C-methyl-D-erythritol 4-phosphate; MK, Mevalonate kinase; NADP+, Nicotinamide adenine dinucleotide phosphate, oxidized form; P_, Orthophosphate ion; PMK, Phosphomevalonate kinase; PP_i, Pyrophosphate ion.
Non-mevalonate (MEP) pathway

D-glyceraldehyde 3-phosphate

Pyruvate

CO₂

DXS

CO₂

NADPH

NADP⁺

DXR

MEP

CMP

CTP

CMP

PP₁

ATP

ADP

ispE

CDP-ME

CDP-MEP

ispF

CMP

MECDP

[FeS]₁

[FeS]₂

ispG

HMDP

NADPH

NADP⁺ + H₂O

ISPH₁

ISPH₂

IPP

DMAPP
In the transposon site hybridization (TraSH) experiments, where Sassetti and Rubin tried to identify genes required for growth (Sassetti, 2001; 2003), half of the enzymes participating in the non-mevalonate pathway (DXS, IspD, E, and F) were considered to be essential. Data were not shown for IspG and IspH1, but both IspH2 and DXR were listed as non-essential genes. Neither of the genes were, in the paper presented by Stewart, reported to be up-regulated during heat-shock (Stewart, 2002), nor were they mentioned in the in silico TraSH study performed by Lamichhane (Lamichhane, 2003). In the nutrient-starvation model presented by Betts, IspG was up-regulated after 96 hours, while IspD was actually down-regulated (Betts, 2002). When it comes to DXR, more recent studies performed by Tanya Parish and Amanda Brown shows that this enzyme indeed is essential for growth in vitro (personal communication). The chromosomal copy of the gene could only be deleted when a functional copy was present on an integrating vector. It should however be stated that in the TraSH experiment presented by Sassetti and Rubin, the reported ratio for DXR (0.25) was very close to the cut-off value used in the study (0.20), which indicates that their conclusion of non-essentiality can be a false negative. However, in order to really confirm essentiality, studies would also have to be performed in vivo.

Currently, 21 DXR structures are present in the PDB, representing enzymes from the bacteria E. coli, Zymomonas mobilis, and M. tuberculosis, see Table 1. Among the E. coli entries are two apo structures (PDB codes 1K5H and 1ONN) (Reuter et al., 2002; Steinbacher et al., 2003), structures containing a bisphosphonate and a sulfate, with and without Mg$^{2+}$ (PDB codes 1T1S and 1T1R) (Yajima et al., 2004), and two structures with Mn$^{2+}$ bound, one in complex with the inhibitor fosmidomycin (PDB code 1ONO and 1ONP) (Steinbacher, 2003). Four additional E. coli structures have NADPH present: one with bound sulfate, one in complex with the inhibitor fosmidomycin, one with fosmidomycin and citric acid and one in complex with the substrate DXP (PDB codes 1JVS, 1Q0L, 1Q0H, and 1Q0Q, respectively) (MacSweeney et al., 2005; Yajima et al., 2002). Two Z. mobilis DXR structures have been determined, one with an acetate ion (PDB code 1R0K) and one in complex with NADPH (PDB code 1R0L) (Ricagno et al., 2004). The eight M. tuberculosis DXR structures present in the PDB are all presented in papers II and III of this thesis. These represent five structures of the wild type enzyme, MtDXR-S, MtDXR-FMN, MtDXR-FN, MtDXR-MNS and MtDXR-MS, where F stands for fosmidomycin, M for Mn$^{2+}$, N for NADPH, and S for SO$_4^{2-}$. The remaining three structures, MtDXR$^{NO}$-S, MtDXR$^{NO}$-NS, and MtDXR$^{NO}$-FN represent a mutated enzyme containing the substitutions D151N and E222Q. Recently, a new E. coli DXR structure has been released, showing the enzyme in complex with fosmidomycin, Mg$^{2+}$, and NADPH (PDB code 2EGH) (Yajima et al., 2007).
Table 1. DXR structures currently present in the PDB represent enzymes from *M. tuberculosis* (*MtDXR*), *E. coli* (*EcDXR*), and *Z. mobilis* (*ZmDXR*). Here, F stands for fosmidomycin, M for Mn$^{2+}$, N for NADPH, and S for SO$_4^{2-}$. Presented is the PDB entry code, the resolution, and the reference. When comparing the DXR structures listed below one can see that there is a 40-41% amino acid sequence identity between *MtDXR* and the other DXR enzymes.

<table>
<thead>
<tr>
<th>Structure</th>
<th>PDB code</th>
<th>Resolution (Å)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>MtDXR</em>-S</td>
<td>2C82</td>
<td>1.90</td>
<td>(Henriksson <em>et al.</em>, 2006)</td>
</tr>
<tr>
<td><em>MtDXR</em>-MS</td>
<td>2JD2</td>
<td>2.15</td>
<td>(Henriksson <em>et al.</em>, 2007)</td>
</tr>
<tr>
<td><em>MtDXR</em>-MNS</td>
<td>2JD1</td>
<td>2.00</td>
<td>(Henriksson, 2007)</td>
</tr>
<tr>
<td><em>MtDXR</em>-FN</td>
<td>2JCV</td>
<td>2.20</td>
<td>(Henriksson, 2007)</td>
</tr>
<tr>
<td><em>MtDXR</em>-FMN</td>
<td>2JCZ</td>
<td>2.05</td>
<td>(Henriksson, 2007)</td>
</tr>
<tr>
<td><em>MtDXR$^{N_2}$</em>-S</td>
<td>2JCY</td>
<td>2.35</td>
<td>(Henriksson, 2007)</td>
</tr>
<tr>
<td><em>MtDXR$^{N_2}$</em>-NS</td>
<td>2JD0</td>
<td>2.30</td>
<td>(Henriksson, 2007)</td>
</tr>
<tr>
<td><em>MtDXR$^{N_2}$</em>-FN</td>
<td>2JCX</td>
<td>2.10</td>
<td>(Henriksson, 2007)</td>
</tr>
<tr>
<td><em>EcDXR</em></td>
<td>1K5H</td>
<td>2.50</td>
<td>(Reuter, 2002)</td>
</tr>
<tr>
<td><em>EcDXR</em></td>
<td>1ONN</td>
<td>2.60</td>
<td>(Steinbacher, 2003)</td>
</tr>
<tr>
<td><em>EcDXR</em>-S,bisphosphonate</td>
<td>1T1R</td>
<td>2.30</td>
<td>(Yajima, 2003)</td>
</tr>
<tr>
<td><em>EcDXR</em>-M</td>
<td>1ONO</td>
<td>2.50</td>
<td>(Steinbacher, 2003)</td>
</tr>
<tr>
<td><em>EcDXR</em>-MS,bisphosphonate</td>
<td>1T1S</td>
<td>2.40</td>
<td>(Yajima, 2003)</td>
</tr>
<tr>
<td><em>EcDXR</em>-FM</td>
<td>1ONP</td>
<td>2.50</td>
<td>(Steinbacher, 2003)</td>
</tr>
<tr>
<td><em>EcDXR</em>-NS</td>
<td>1JVS</td>
<td>2.20</td>
<td>(Yajima, 2002)</td>
</tr>
<tr>
<td><em>EcDXR</em>-FN</td>
<td>1QOL</td>
<td>2.65</td>
<td>(MacSweeney, 2005)</td>
</tr>
<tr>
<td><em>EcDXR</em>-FN,citric acid</td>
<td>1Q0H</td>
<td>2.20</td>
<td>(MacSweeney, 2005)</td>
</tr>
<tr>
<td><em>EcDXR</em>-N,DXP</td>
<td>1Q0Q</td>
<td>1.90</td>
<td>(MacSweeney, 2005)</td>
</tr>
<tr>
<td><em>EcDXR</em>-FN,Mg$^{2+}$</td>
<td>2EGH</td>
<td>2.20</td>
<td>(Yajima, 2007)</td>
</tr>
<tr>
<td><em>ZmDXR</em>-acetate ion</td>
<td>1R0K</td>
<td>1.91</td>
<td>(Ricagno, 2004)</td>
</tr>
<tr>
<td><em>ZmDXR</em>-N</td>
<td>1R0L</td>
<td>2.70</td>
<td>(Ricagno, 2004)</td>
</tr>
</tbody>
</table>

**Constructs, constructs, and more constructs**

After experiencing several problems with cloning and expression of the *M. tuberculosis* DXR enzyme, most probably caused by suboptimal primers and expression protocols, two template plasmids were provided by AstraZeneca India Pvt. Ltd, originating from *M. tuberculosis* strain H37Rv (Cole, 1998). These contained two constructs, both of which had extra stretches at the N- and C-terminus of around 20 amino acids. In addition, one of them had an N-terminal, and the other a C-terminal His-tag. Both of these expressed soluble protein, in an amount consistent with that documented by Astra-Zeneca. However, since the expression was performed over-night, the deci-
sion was made to optimize it so that it could be performed within one day, preferably providing a higher yield of pure and soluble protein. The optimization was a success, and the final protocol is described in the following section. Crystallization screens were set up, but as could be expected, no crystals were produced from these constructs, probably since extra, flexible parts can cause problems when it comes to crystallization. We therefore knew that the extra stretches present in the constructs had to be removed. However, we did not know how important these stretches were for the solubility of the protein, and this had to be considered as a potential pitfall. Following the new protocol, the AstraZeneca construct with an N-terminal His-tag gave a somewhat higher yield of soluble protein, and therefore, the new constructs were designed with an N-terminal His$_6$ affinity tag. At this stage, three different constructs were designed and tested. What distinguished them was the length of the extra C-terminal stretch that was kept in the construct. In the first, the entire stretch was kept, in the second approximately half of the residues were left, and the third construct corresponded to the actual coding sequence of the protein, with just an N-terminal tag. Since this last construct resulted in soluble protein, it was chosen for further analysis. Crystallization experiments were again set up and though crystalline material was produced, no well ordered crystals were obtained. When further examining the expression and purification results we found an explanation for this. At all stages, two protein products, of approximately the same size could be detected, indicating that the full-length protein was undergoing proteolysis. Since both forms of the protein bound to the chelating column, the degradation was at the C-terminus. Since a homogeneous protein sample is an important factor when it comes to crystallization, we decided at this stage to design yet another construct.

After comparing the $Mtx$DXR sequence with similar sequences and structures in a BLAST search (Altschul et al., 1997), analyzing secondary structure elements, and making a qualified guess of how much was being cleaved off at the C-terminus, the sequence corresponding to residues 1-393 out of 413 in the full-length protein (molecular weight 42.0 kDa) was amplified by PCR from the previously mentioned full length construct. The primers 5'-CACCATGGCTCATCATCATCATCATGTGACCAACTCGACGAC-3' (forward) and 5'-CTACGCCACCGAAGCCATACCAG-3' (reverse), used together with Pfu Ultra DNA polymerase (Stratagene) simultaneously introduced the N-terminal His$_6$-tag. In further discussion, this construct will be referred to as $Mtx$DXR construct number 1 ($Mtx$DXR-1). With this construct the first well ordered crystals were produced, which eventually resulted in the first $Mtx$DXR structure. This will be discussed more in detail in later sections. However, what is crucial to mention at this point is that when inspecting the three-dimensional crystal structure there was no density, or room for the last four residues in the construct, suggesting that some proteolysis had taken place. Based on this, and with the goal to further optimize
the design and improve crystallization, a new construct was designed where an additional four residues were cleaved off at the C-terminus. This construct, containing residues 1-389, which will be referred to as MtDXR construct number 2 (MtDXR-2), was amplified by PCR from MtDXR-1 using the same forward primer as before, but now with the reverse primer 5'-CTACATACCAGATACCGCGCTG-3'.

At this point in time we also decided to produce a mutant protein in which two of the three residues shown to participate in metal binding, Asp151 and Glu222 (Steinbacher, 2003), were mutated into their corresponding amides. This allowed us to test if these residues indeed are crucial for the activity, and to see how they would affect substrate binding. The primers 5'-GTGAGATCGTGCCGCTCAACTCCGAAACTCGCGCTG-3' (D151N) and 5'-CTGGTCAACAAAGGGACTTCAGGTCATCGAAACCCACCTG-3' (E222Q) used with the QuickChange multimutagenesis kit (Invitrogen) introduced the replacements of the MtDXR double mutant, MtDXR^NQ, using the clone of MtDXR-2 as template.

Cloning, expression and purification of MtDXR and MtDXR^NQ

For both MtDXR-1 and -2, the Champion™ pET101 Directional TOPO ® expression kit (Invitrogen) was used for ligation of the PCR product into the pET101D-TOPO vector. Cloning, now also for the mutant MtDXR^NQ, was performed in E. coli TOP10F’ cells (Invitrogen). Positive clones were selected by growth on Luria agar plates containing 50 μg/ml ampicillin, and plasmids were isolated following the QIAprep® Spin Miniprep kit protocol (Qiagen). For both MtDXR constructs, an analytical PCR was performed using the T7 forward primer from the Champion™ pET101 Directional TOPO ® expression kit (Invitrogen), and the respective reverse primers. For MtDXR^NQ, the analytical PCR was performed using the T7 reverse primer, and the analytical primers 5'-GTCAGATCGTGCCGCTG-3' and 5'-CTGGTCAACAAAGGGACTTC-3'. The three clones were verified by DNA sequence analysis (Uppsala Genome Center, Rudbeck Laboratory). Expression was performed in E. coli BL21-STAR™ (DE3) cells (Invitrogen) at 37 ºC. The cultures were induced with 100 mg/L isopropyl-β-D-thiogalactopyranoside (Sigma) at OD_{550} =0.6-1.0, and growth was continued for 3 h. The cells were harvested and washed with 1×SSPE buffer (150 mM NaCl, 10 mM NaH_{2}PO_{4} pH 7.5, 1 mM EDTA) and then stored at -20 ºC.

For purification, the thawed cells were treated with lysis buffer (50 mM NaH_{2}PO_{4} pH 8.0, 300 mM NaCl, 10 mM imidazole, 10% glycerol) with 0.01 mg/ml RNase, 0.02 mg/ml DNase and lysed with a Constant Cell Disruptor (Constant Systems Ltd) operated at 250 MPa. The cell lysate, containing
soluble protein, was incubated for 30 min at 4 °C with nickel-NTA agarose slurry (Qiagen). The resin was washed, at room temperature, with ten column volumes of lysis buffer containing 20 mM imidazole and the protein was then eluted with four column volumes of the same buffer containing 250 mM imidazole.

*Mt*DXR-1 was further purified by size-exclusion chromatography on a HiLoad 16/60 Superdex 75 column (Amersham Pharmacia Biotech) developed with a buffer containing 150 mM NaCl and 20 mM Tris-HCl pH 7.5. The fractions containing protein were pooled and concentrated to 5 mg/ml (based on the calculated absorbance of 0.907 for a 1 mg/ml solution at 280 nm) using a Vivaspin concentrator (Vivascience, molecular-weight cutoff 10 kDa) and stored at -80 °C.

*Mt*DXR-2 and *Mt*DXRNQ, were after the nickel-NTA column placed on a PD10 column (GE Healthcare), were the buffer was replaced with one containing 150 mM NaCl, 10% glycerol and 20 mM Tris-HCl, with a final pH off 7.5. The protein was then further purified by size-exclusion chromatography on a HiLoad™ 16/60 Superdex™ 200 column (Amersham Biosciences) developed with the same buffer. The pooled fractions were concentrated to 4.3 mg/ml, and the glycerol concentration was at the same time lowered to 2%.

For all three clones, the material was more than 98% pure as deduced from SDS/PAGE analysis (PhastSystem™, Amersham Biosciences). The yield of pure protein was 1.5 mg/L culture for *Mt*DXR-1, 1.4 mg/L culture for *Mt*DXR-2, and 1.7 mg/L culture for *Mt*DXRNQ.

Activity, and inhibition by fosmidomycin

Activity assay

Since NADPH is being used in the reaction catalyzed by DXR, the activity of the enzyme can be evaluated in a spectrophotometric assay (Kuzuyama *et al.*, 1998b; Takahashi, 1998), where the NADPH-dependent rearrangement and reduction of DXP to form MEP is monitored at 340 nm, using the absorption of NADPH ($\varepsilon_{340nm}=6220 \text{ M}^{-1}\text{cm}^{-1}$). Each assay contained 350 μl reaction buffer (129 mM NaCl, 2.9 mM β-mercaptoethanol, 5.7% Me2SO, 2.1 mM MnCl2, and 64 mM HEPES-NaOH, pH 7.5), 50 μl *Mt*DXR-1 (in 100 mM NaCl, 0.01% Brij-35 and 50 mM HEPES-NaOH pH 7.5), 50 μl NADPH (in dH2O) and 50 μl DXP (in dH2O). Six measurements were made where the DXP concentrations varied between 25 and 800 μM, at final concentrations of 357 nM and 204 μM for DXR and NADPH, respectively. Four reactions were also carried out were the NADPH concentration was varied between 3 and 25 μM, at final concentrations of 18 nM and 800 μM for DXR and DXP, respectively. The reactions were followed as the decrease in
absorbance at 340 nm at 22 °C (DU® 640 spectrophotometer, Beckman). Measurements were made every 5 s during a 3 min period. The slope of the linear phase of each reaction was used in calculating the initial velocity.

The $K_m$ value, the Michaelis constant, for DXP and NADPH, as well as $V_{max}$, the maximum rate of the enzyme, can be calculated from a Lineweaver-Burke or Hanes-Wolff plot. However, if the measurements made when varying DXP are not fully saturated with NADPH, and vice versa, the values obtained will be misleading. Since the $k_{cat}$ value, the turnover number for the enzyme, is calculated as the ratio between $V_{max}$ and the enzyme concentration, this value will not be trustworthy either, if $V_{max}$ is underestimated. Therefore, an equation, which will be derived below, was used, taking into account both substrates. The $k_{cat}$ value and $K_m$ values were then calculated in an iterative process.

The reaction below contains one enzyme, $E$ (DXR), and two different substrates, denoted by $N$ (NADPH) and $D$ (DXP). The full enzyme-substrate complex is denoted $END$, and the products $N^*$ and $D^*$, respectively. $K_s$ is the dissociation constant, and $k_{cat}$ is, as stated above, a catalytic rate constant, that describes how many substrate molecules are turned into product molecules per active site, and per minute.

$$E + N + D \xrightleftharpoons[K_s]{k_{cat}} END \xrightarrow{k_{cat}} E + N^* + D^*$$

Instead of using the dissociation constant $K_s$, equations can be written describing the two possibilities for the order in which the two substrates bind, to form the full ternary enzyme-substrate complex.

$$K_{iN} = \frac{[E][N]}{[EN]}, \quad K_N = \frac{[ED][N]}{[END]}, \quad K_{iD} = \frac{[E][D]}{[ED]}, \quad K_D = \frac{[EN][D]}{[END]}$$

The initial rate, $v$, of each reaction is related to $k_{cat}$ and the concentration of the enzyme-substrate complex in the following way

$$v = k_{cat} [END]$$

Knowing that $V_{max} = k_{cat} [E]_0$, where $[E]_0$ is the total enzyme concentration, and inserting this in the previous equation, the following expression for $k_{cat}$ is obtained

$$k_{cat} = \frac{V_{max}}{[E]_0}$$
, which in turn allows us to express \( v \) as

\[
v = \frac{V_{\text{max}} [\text{END}]}{[E]_0}
\]

By inserting the full expression for the enzyme concentration,

\[
[E]_0 = [E] + [EN] + [ED] + [\text{END}]
\]

, the expression for \( v \) can be expanded to

\[
v = \frac{V_{\text{max}} [\text{END}]}{[E] + [EN] + [ED] + [\text{END}]}
\]

By then dividing by \([\text{END}]\), multiplying by \([N][D]\), utilizing the fact that \( V_{\text{max}} = k_{\text{cat}} [E]_0 \), and inserting the equations for \( K_{iN}, K_N, K_{iD}, \) and \( K_D \), the final equation is obtained

\[
v = \frac{k_{\text{cat}} [E]_0 [N][D]}{K_{iN} K_D + K_D [N] + K_N [D] + [N][D]}
\]

The non-linear data from the activity measurements were fitted using Mathematica v.3.0 for Macintosh (Wolfram Research Inc.) to the equation above, where to summarize, \( v \) is the initial velocity, \([N]\) and \([D]\) are the substrate concentrations, \([E]_0\) the enzyme concentration, \(K_N\) and \(K_D\) are the Michaelis constants for the two substrates and \(K_{iN}\) is the dissociation constant for NADPH binding to the enzyme.

\(Mt\text{DXR-1}\) was estimated to have a \( k_{\text{cat}} \) value of 5.3 s\(^{-1}\), with \( K_m \) values of 7.2 \(\mu\)M for NADPH and 340 \(\mu\)M for DXP. These give rise to calculated \( k_{\text{cat}}/K_m \) values, which give a measure of the catalytic efficiency, of \(7.4 \times 10^5 \text{M}^{-1}\text{s}^{-1}\) for NADPH and \(1.6 \times 10^4 \text{M}^{-1}\text{s}^{-1}\) for DXP. These values are similar to those previously published for a full-length and a proteolytically cleaved enzyme truncated by 18 amino acids, where \(k_{\text{cat}}\) was 5.0 s\(^{-1}\) and the \(K_m\) values, for NADPH and DXP, were 3.3 \(\mu\)M and 100 \(\mu\)M, respectively, under similar conditions (Argyrou et al., 2004). Although our calculations were performed under the assumption that NADPH binds first, the values derived do not depend greatly on the kinetic model used, which is also in agreement with earlier studies (Argyrou, 2004).

For \(Mt\text{DXR-2}\), kinetic parameters were not calculated. However, from a smaller set of measurements it could be seen that the activity of this con-
struct is highly similar to that of MtDXR-1, indicating that the more extensive truncation at the C-terminus did not affect the activity. However, the double mutant, MtDXR^{NQ}, represents a completely inactive enzyme, emphasizing the importance of residues Asp151 and Glu222 in metal binding and catalysis.

**Fosmidomycin as an *M. tuberculosis* inhibitor**

Fosmidomycin is a hydrophilic compound that resembles the DXR enzyme’s natural substrate, DXP. It is a known inhibitor of the non-mevalonate pathway in plants and bacteria (Shigi, 1989; Zeidler *et al.*, 1998), and it has been shown to specifically inhibit *E. coli* DXR (Kuzuyama *et al.*, 1998a). The compound is active against *Plasmodium falciparum* in humans (Lell *et al.*, 2003; Missinou *et al.*, 2002), and *Plasmodium vinkei* in mice (Jomaa *et al.*, 1999). Although fosmidomycin inhibits MtDXR (see below), it does not affect the growth or viability of the tuberculosis bacteria (Dhiman *et al.*, 2005). Studies have been made by Amanda Brown and Tanya Parish (personal communication) in order to find the reason for this natural resistance. Many bacteria have a glycerol 3-phosphate transporter, which is responsible for the uptake of fosmidomycin. Mutants of *E. coli* that lack this transporter have also been shown to be resistant to the compound (Sakamoto *et al.*, 2003). The gene for a homologous transporter is not found in the *M. tuberculosis* genome, which then eliminates this means of uptake. In their studies, Brown and Parish investigated the fosmidomycin uptake both in *M. tuberculosis*, and also in the faster growing *M. smegmatis*. The results showed that fosmidomycin could only be found in the extra-cellular medium and not inside the bacteria. In order to see if the reason for this was the work of efflux pumps, studies were performed using efflux pump inhibitors. These did not show any effect. Tests were also done showing that the bacteria inside their cells were not modifying the inhibitor. Therefore, the explanation for the fosmidomycin resistance in *M. tuberculosis* is simply that the compound cannot make its way, across the membrane, and into the bacteria. A reason for this can be the phosphonate part of the inhibitor, since charged molecules do not readily diffuse across membranes.

**Inhibition studies**

The inhibition measurements were performed in the same spectrophotometric assay as described above. Each assay contained 350 µl of the reaction buffer, 50 µl of 3.85 µM MtDXR-2 (in 100 mM NaCl, 0.01% Brij-35, and 50 mM HEPES-NaOH pH 7.5), 50 µl of 1.1 mM NADPH (in dH₂O), 50 µl of 2.2 mM DXP (in dH₂O), and 50 µl of fosmidomycin (in dH₂O). One measurement was made in the absence of fosmidomycin, and six measurements were carried out with final fosmidomycin concentrations ranging be-
tween 25 and 200 nM. The reactions were again followed during 3 minutes; with time points recorded every 5 seconds. The slope of the linear phase (25-125 s) of each reaction was used in calculating the initial velocity. The velocities were then plotted against the corresponding fosmidomycin concentration. A line was fitted to the points in the approximately linear region of the curve, and the equation so generated was used to estimate the IC$_{50}$ value of fosmidomycin. The reaction rate in the absence of fosmidomycin was 0.0036 ΔA$_{340}$ s$^{-1}$, and therefore the inhibitor concentration associated with IC$_{50}$ was found at half this rate.

At an $Mt$DXR-2 concentration of 0.35 μM and substrate concentrations of 0.2 mM for DXP and 0.1 mM for NADPH, the IC$_{50}$ value of fosmidomycin was estimated to be 80 nM. This value is similar to that of 310 nM, previously published for $M$. tuberculosis DXR (Dhiman, 2005), and to that of 30 nM reported for the $E$. coli enzyme (Haemers et al., 2006).

Crystallization experiments
After setting up crystallization experiments of the $Mt$DXR full-length construct with an N-terminal His$_{6}$-tag, using the Cryo I/II (Emerald Biosystems), JBScreenHTS1 (Jena Biosciences), and Core96 (Page et al., 2004) screens, crystalline material was produced, at 22 °C, within days in condition 11D of the Cryo I/II screen. This crystallization buffer contained 50% PEG 400, 100 mM acetate, 0.2 M Li$_{2}$SO$_{4}$, and had a final pH of 5.1. In order to optimize the conditions, to see if well ordered crystals could be produced, several tests were performed. Different small PEGs were tested in the 11D buffer, and a batch experiment was also set up using a buffer of 50% 11D buffer and 50% Tris buffer (20 mM Tris, 150 mM NaCl, 10% glycerol, pH 7.5). In all experiments seeds were transferred from the original Cryo I/II screen, but unfortunately the crystalline material did not improve.

With $Mt$DXR-1, a microbatch experiment was set up of the CryoI/II screen, where the drops consisted of 50% protein (in the buffer used in the size exclusion chromatography) and 50% of the respective CryoI/II solutions. The bottom reservoir solution also contained a 50/50 solution, but without protein. The drops were streak-seeded from the crystalline material in condition 11D of the full-length construct, and crystals were now produced in condition 2A, containing 40% ethylene glycol, and 100 mM acetate, with a final pH of 5.0. The crystallization conditions were again optimized, and the inhibitor fosmidomycin was added to the crystallization setup to hopefully stabilize the protein structure. Used in the final crystallization experiment was a 2 μl sitting drop consisting of 1 μl protein (5 mg/ml in the buffer used in the size-exclusion chromatography, with the addition of 1.3 mM fosmidomycin) mixed with 1 μl crystallization buffer (40% ethylene glycol, 2 mM MgSO$_{4}$, 20 mM dithiothreitol, 0.2 mM EDTA and 100 mM
acetate pH 5.0), which was streak-seeded immediately with a horse hair. The final conditions in the drop (75 mM NaCl, 10 mM Tris-HCl pH 7.5, 20% ethylene glycol, 1 mM MgSO₄, 10 mM dithiothreitol, 0.1 mM EDTA, 50 mM acetate pH 5.0, 0.65 mM fosmidomycin, 2.5 mg/ml MtDXR, measured final pH ~5.1) were stabilized by vapor diffusion against the same solution without fosmidomycin and protein. Several different morphologies could be seen among the crystals produced. Some of them appeared as small triangles, or diamonds, while others had the appearance of small “ski-slopes” or boxes. Unfortunately, all of these resulted in diffraction patterns that indicated double lattices, and which could not be indexed properly. Some of the crystals even had a clear line in the middle, indicating that they had two parts, but when separating these crystals into the two “halves” the problem remained.

It was not until a few pyramid-shaped crystals appeared in one of the drops that a well ordered diffraction pattern could be recorded, and used for solving the first MtDXR structure. These crystals appeared within a few days and grew to dimensions of 0.1 × 0.1 × 0.3 mm in four weeks. Before the crystals were flash-cooled in liquid nitrogen, they were transferred to a drop of the reservoir solution complemented with an additional 10% ethylene glycol and 1 mM fosmidomycin.

MtDXR-2, co-crystallized with fosmidomycin, produced needle-like crystals within a few days, in condition 7H of the JCSG+Suite* screen (Qiagen). The sitting drop contained 0.9 μl of protein solution (3.3 mg/ml in the buffer used during concentration, plus 10 mM fosmidomycin, 3 mM NADPH, 0.1 mM EDTA, and 10 mM dithiothreitol) and 0.9 μl of reservoir solution (0.2 M (NH₄)₂SO₄, 25% PEG 3350, 0.1 M Bis-Tris, final pH of 5.7). Crystallization of both MtDXR, as well as of MtDXRNQ, was then optimized with the aid of seeding in a batch experiment. The 4-μl sitting drops consisted of 2 μl protein solution (3.3 mg/ml MtDXR or MtDXRNQ in the buffer used in the concentration with the addition of 0.1 mM EDTA and 10 mM dithiothreitol) and 2 μl of the previously mentioned crystallization buffer. The reservoir solution consisted of 0.1 M (NH₄)₂SO₄, 12.5% PEG 3350, 0.05 M Bis-Tris, 75 mM NaCl, and 10 mM Tris-HCL, final pH 5.9. MtDXR was co-crystallized with different ligands in the following combinations: DXP and MnSO₄; DXP and NADPH; DXP, MnSO₄ and NADPH; fosmidomycin and NADPH; fosmidomycin, MnSO₄ and NADPH. The MtDXRNQ mutant was crystallized both without ligands, as well as in the following combinations: DXP, MnSO₄ and NADPH; fosmidomycin and NADPH. In cocrystallization experiments, the following concentrations of the ligands were added to the protein solution: 20 mM DXP, 10 mM fosmidomycin, 12 mM MnSO₄, and 3 mM NADPH. The needle-like crystals, shown in Figure 13, appeared within a few days and grew to average dimensions of 0.2 × 0.05 × 0.02 mm in two weeks. A slight variation in crystal size could be observed between the wild type enzyme and the MtDXRNQ mutant. The MtDXR crystals, with different ligands, resulted in quite thin long needles, while the
mutant enzyme resulted in shorter needles, which instead had a greater thickness.

Before the crystals were flash-cooled in liquid nitrogen, they were transferred to a drop of cryo solution containing 0.1 M (NH₄)₂SO₄, 13% PEG 3350, 0.06 M Bis-Tris, 75 mM NaCl, and 25% glycerol, with a final pH of 5.9. Depending on which ligands were used for co-crystallization, the following concentrations of the ligands were present in the cryo solution: 10 mM DXP, 10 mM fosmidomycin, 12 mM MnSO₄, and 3 mM NADPH. This was designed to ensure the best possible binding of the ligands, in the respective crystals.

MtDXR and MtDXRᴺQ structures

Data processing, structure determination and refinement

Crystallographic data were collected under cryo conditions, at various beamlines at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France, and at MAX-lab in Lund, Sweden. For all data sets, the data were indexed and integrated with MOSFLM (Leslie, 1999) and processed with SCALA (Evans, 1993) as implemented in the CCP4 program suite (CCP4, 1994). The crystal symmetry for the first well diffracting MtDXR-1 crystal seemed to be either P₄₁ or P₄₃, and the Matthews coefficient (Matthews, 1968) suggested two molecules in the asymmetric unit (2.3 Å³Da⁻¹, with 46% solvent). A search model was produced using the program SOD (Kleywegt et al., 2001a), and in this model, the residues identical to E. coli DXR (PDB code 1Q0Q, 40% sequence identity) (MacSweeney, 2005) were preserved, while the others were exchanged for alanines. After running molecular replacement in Phaser (McCoy, 2005), with different potential space groups, the best fits were found for space groups P₄₁ and P₄₁22. The TF Z-scores for these were approximately 15 and 11, which differed from the remaining scores, which lay around 7. In Phaser, a TF Z-score below 5 indicates that a solution could not be found; while a score above 8 means that you have definitely solved the structure. Both of the remaining space groups were put through the ARP/wARP autobuilding routine (Lamzin, 1993), where P₄₁ again showed the best result. Refinement of the model obtained
from Phaser was carried out with the program REFMAC5 (Murshudov, 1997), and rebuilding was performed with the program O (Jones, 1991), exploiting the maps from the previously mentioned autobuilding in ARP/wARP. NCS restraints between the A and B molecule were not used in the early stages of the refinement as differences could be seen between the two molecules. However, addition of NCS restraints in the last rounds resulted in slightly better refinement statistics, and also improved the electron density in some places. Although the close similarity between the two molecules in the asymmetric unit, and the nearly perfect twofold symmetry relating them, indicated that the space group could be P4₁22, comparison of the statistics from the first rounds of refinement indicated that P4₁ must be the correct choice of space group. Waters were added after analyzing the results from the ARP/wARP water-building routine.

The crystals of MtDXR-2 and MtDXRNQ possessed the symmetry of space group P2₁, and the Matthews coefficient suggested that there were two molecules in the asymmetric unit (2.2 Å³Da⁻¹ with 44.5% solvent). The structures were solved by molecular replacement using AMoRe (Navaza, 1994) with the previously determined structure, described above, as a search model. Data processing, rebuilding, refinement, and the adding of waters were performed with the same programs as earlier.

Two important questions during the processing of crystallographic data, and structure determination, are: where should you cut the data, and when is the refinement completed? I believe that these are questions that when asked of several people, most probably will result in at least an equal number of different answers. I am here taking the opportunity of sharing my thoughts. When looking at the data collection and refinement statistics for the different MtDXR and MtDXRNQ structures, the values might trigger the feeling of inconsistency in the data processing. However, a quite consistent strategy was applied. When deciding where to cut the data I foremost look at the completeness, $R_{\text{meas}} \langle I \rangle / \langle \sigma I \rangle$, and the multiplicity. I prefer to have a completeness of over 90%, both overall and in the highest resolution shell, although some might think this is a bit stringent. I simply feel calmer knowing that I have a large amount of the data within the resolution range chosen. For $\langle I \rangle / \langle \sigma I \rangle$, I aim to be above 2 in the highest resolution shell, in order to ensure distinction from the background. For the multiplicity the same value is above 3, in order to minimize errors in individual measurements, and get a more “even” and reliable set of reflections. When it comes to $R_{\text{meas}}$, I prefer to be below 0.5 in the highest resolution shell, although I have been a bit flexible depending on the quality of the data. For some of the data sets though, $R_{\text{meas}}$, is much lower than my limit. This is the result of the decisions regarding the other mentioned parameters, but foremost the completeness, which in turn is defined by the data collection set up and strategy.
During the refinement, different weights were tested in order to find the optimal balance between the crystallographic data and the stereochemistry. I have always aimed at having no more than a 5% difference between $R$ and $R_{\text{free}}$, although this value fluctuates between 3.2 and 5.3 for the different data sets. Regarding the root mean square deviations from ideal bond lengths and bond angles I have aimed to keep below 0.02 Å and 2°, respectively. For all of the data sets the values have been well below these requirements with average values of around 0.01 Å and 1.17°, respectively. Another good indicator of structural quality is the percentage of Ramachandran plot outliers. The values here all lie between 0.6 and 1.5%, well below the recommended limit of 5% for the typical 2 Å resolution structure (Kleywegt, 1996). For several of the structures the temperature factors, the $B$ factors, had to be considered for some of the ligands. Instead of removing some parts of the ligands, all atoms were kept, but the occupancy for them was set to zero. This because we wanted, for example NADPH to look like NADPH, and not anything else, in the structures. The $B$ factors for the solvent atoms were also considered, and some waters were removed due to high $B$ factors compared to those of the protein atoms.

The remaining question is then, when is the refinement completed, or in other words, when is the model ready for deposition in the Protein Data Bank? To ensure deposition of an as accurate model as possible all the protein parts of the structure should fit the density as well as they can possibly be placed. Rotamers should be checked, as well as hydrogen-bonding interactions. Also, the correct number of waters to include in the final model deserves some consideration. I noted that after initially adding waters, the temperature factors for the solvents showed a slow but steady increase from lower to higher values, then a sudden rapid increase, see Figure 14A. During rebuilding, when manually checking the density and hydrogen bonding for each solvent, some waters were discarded and others added, and the average temperature factor was thereby lowered, see Figure 14B. However, by applying a cut-off at the inflection point, as performed in Figure 14C, the chances of including incorrect information were minimized. Finally, since the structure factors for all the structures presented in this thesis are available from the PDB, anyone who desires can him or herself perform additional refinement.
Figure 14. (A) The sorted temperature factor distribution for the water molecules in MtDXR-FMN (A) after adding waters with ARPwARP, (B) during mid-refinement, after manual inspection of the density, (C) and for the final structure. Number of water molecules are given on the x-axis, and the temperature factor on the y-axis.

Overall structure and domain movements

For all the structures presented here, the asymmetric unit contains one homodimer, see Figure 15. Each of the subunits is composed of three domains, an N-terminal NADPH-binding domain, a central catalytic domain, and a C-terminal α-helical domain. These are arranged in a V-shape placing the central catalytic domain at the vertex, and the N- and C-terminal domains at the two arms. The dimer is formed by the interactions between the β-sheet of each catalytic domain and also by the β-strand from the connecting region of each subunit. In this way, a ten-stranded β-barrel is formed, which is the core of the dimer interface.
Figure 15. The overall structure of the MtDXR homodimer is illustrated in a ribbon drawing (A). The B subunit is colored in black, while the A subunit is colored according to the domain structure, going from darker to lighter grey from the N-terminal NADPH-binding domain, the central catalytic domain, via the connecting region, to the C-terminal all α-helical domain. The dimer interaction is formed by an eight-stranded β-sheet, which is joined to an additional strand from each subunit to form a distorted ten-stranded β barrel, which is more clearly seen in panel (B), where the dimer has been slightly rotated.

When solving the structure of the MtDXR-1 construct it became clear that fosmidomycin was not present in the active site, but that instead a sulfate ion was placed at the site of the phosphate of the natural substrate DXP. Also, for all of the following structures crystallized in the presence of DXP, the substrate could not be modeled. Therefore, the structures presented in our papers are designated as MtDXR-S, MtDXR-FMN, MtDXR-FN, MtDXR-MNS, MtDXR-MS, MtDXR\textsuperscript{NQ}-S, MtDXR\textsuperscript{NQ}-NS, and MtDXR\textsuperscript{NQ}-FN, where F stands for fosmidomycin, M for Mn\textsuperscript{2+}, N for NADPH, and S for SO\textsubscript{4}\textsuperscript{2−}. These have been deposited in the PDB with accession codes 2C82, 2JCZ, 2JCV, 2JD1, 2JD2, 2JCY, 2JD0, and 2J CX.

The previously presented substrate and inhibitor complexes have shown the flexibility of the DXR enzyme. By analyzing all of the MtDXR and MtDXR\textsuperscript{NQ} structures solved in space group P2\textsubscript{1}, the flexibility for the M.


*tuberculosis* DXR within this crystal packing can be demonstrated, see Figure 16. The B molecule, as well as the central catalytic domain of the A molecule, are kept fixed by the crystal packing interactions. The active site of the B molecule is locked in an open conformation and does not show the complete set of substrate interactions. In the A molecule though, different substrates are able to bind, and the N- and C-terminal domains flex depending on what is present in the active site. Therefore the A molecule provides information about how the enzyme moves during the reaction.

![Figure 16](image)

*Figure 16.* The seven dimers (Cα positions) of the *MtDXR* and *MtDXRNQ* structures, solved in space group P2₁, have been superimposed in *O* (Jones, 1991) using a restrictive pair cut-off of 0.5 Å. The B chains, as well as the dimer-forming central catalytic domains of the A chains are well aligned, indicating that crystal contacts keep these parts fixed in the crystal packing. The N-terminal and C-terminal domains of the A chains however, move depending on the ligand state. Using the distance across the active site cleft (between the Cα atoms of residues Gly47 and Ala339), in the figure indicated by a black line, as a simple yardstick, the relative motions can be assessed. In the B molecules, the distance range between 24.5-25.0 Å, and in the A molecule it ranges between 20.3-22.3 Å. The black spheres indicate the center of the phosphonate part of the inhibitor (left sphere), as well as that of the NADPH cofactor (right sphere).
When comparing the *MtDXR* structures with the other available DXR structures from *E. coli* and *Z. mobilis*, one can see that there is a 40-41% amino acid sequence identity to the other DXR enzymes. One can also see that the identity is higher in the central catalytic domain (45-50%) than in the N-terminal (40%) and C-terminal (35%) domains. And although, as seen in Figure 17, the overall three-dimensional structures are similar, *MtDXR* has a few shorter segments, placed in the N-terminal and at the start of the central catalytic domain, compared to the others.

Depending on the relative position of the N-terminal domain to the C-terminal domain, as reflected in the relative yardstick distance across the active site, the different *E. coli* and *M. tuberculosis* DXR structures, can be grouped into closed, open and super-open forms. The only representatives of

![Figure 17. A structural comparison where the *M. tuberculosis* structures *MtDXR*-S (red), *MtDXR*-MS (magenta), *MtDXR*-MNS (brown), *MtDXR*-FN (black), and *MtDXR*-FMN (cyan) and the *E. coli* structures *EcDXR*-N,DXR (green), *EcDXR* (blue, PDB code 1ONN), and *EcDXR*-FM (yellow) have been superimposed based on their N-terminal domains (Cα atoms) with the *lsq-explicit* command in O (Jones, 1991). The *M. tuberculosis* enzyme displays a smaller movement of the domains than the *E. coli* enzyme upon the binding of different substrates and cofactors.](image-url)
the super-open conformation are the A and B molecules of EcDXR (PDB code 1K5H). Among the structures showing an open conformation are those \textit{E. coli} structures binding a single substrate or inhibitor, with or without a metal ion present, along with all B molecules of the \textit{M. tuberculosis} structures, as well as the A molecule of \textit{MtDXR-S}. The remaining A molecules of the \textit{MtDXR} structures show a closed conformation, although not quite as closed as that seen for \textit{EcDXR-FN} (with and without citric acid) and \textit{EcDXR-N,DXP}. This could of course be a consequence of the species difference, and the larger variation among the \textit{EcDXR} structures can be the result of different crystallization conditions and space groups used for structure determination. In order to illustrate the variation, a super-position of \textit{MtDXR} and \textit{EcDXR} structures, based on their N-terminal domains, is shown in \textit{Figure 17}.

Active site

At the center of the \textit{MtDXR} subunit, as seen in \textit{Figure 18}, an active site flap is placed. The flap is formed by a loop region from the central catalytic domain, and underneath the flap, a charged pocket can be found.

\textit{Figure 18}. The A subunit of \textit{MtDXR-FMN} illustrating the position of NADPH, lying along the N-terminal domain, and of fosmidomycin and the Mn\textsuperscript{2+} ion underneath the active site flap of the central catalytic domain.
The pocket has previously shown to interact with the phosphate/phosphonate moieties of DXP and fosmidomycin (MacSweeney, 2005), which can also be seen for our structures. In MtDXR-S, four residues interact with the sulfate ion, Ser177, Ser213, Asn218, and Lys219. When comparing to the A molecule of MtDXR-FMN one can see that the interaction to Ser213 has been lost, while instead His200 has changed its position, now enabling it to interact with the phosphonate part of the inhibitor. The MtDXR-FMN structure also shows, again see Figure 18, the position of the Mn$^{2+}$ ion underneath the flap, and the close contact between the amino end of the inhibitor and the nicotinamide ring of the NADPH substrate. The remaining part of the NADPH molecule lies stretched along one edge of the N-terminal domain.

The advantage of having several structures of the same enzyme in complex with different substrates and inhibitors, especially when the structures are crystallized in the same space group, is that it increases the possibilities of drawing conclusions about dynamics and interactions associated with catalysis. The results presented in the remainder of this section are only valid for the A molecule, which moves depending on different interactions, and not for the B molecule, which is locked in an open conformation due to the crystal packing.

In MtDXR-MS, an octahedral coordination is seen around the Mn$^{2+}$ ion, formed by the interactions with three water molecules, and with the residues Asp151, Glu153, and Glu222, see Figure 19A. This coordination is not changed when NADPH is also present in the structure, as in MtDXR-MNS. However, it does slightly effect the interactions with the sulfate, by indirectly stabilizing the active site flap, and thereby showing the interaction to His200. The density for the nicotinamide ring, as well as for the sugar, of the NADPH substrate is weak. The density for this substrate is stronger in the MtDXR-FN structure, indicating the importance of inhibitor, and possibly also substrate binding, for an optimal binding of NADPH. This structure also provides a good view of the position of fosmidomycin and NADPH in the absence of a metal ion, see Figure 19B. When comparing the interactions to these molecules with those seen in the EcDXR-N,DXP structure (PDB code 1Q0Q), it can be seen that the residues interacting with fosmidomycin are strictly conserved, while only 70% of those interacting with NADPH in the MtDXR-FN structure can be found in the E. coli enzyme. When moving on to the MtDXR-FMN structure, which was mentioned at the beginning of this section, this structure provides the most complete view of the interactions in the active site, see Figure 19C. Here the close contacts both between NADPH and fosmidomycin, but also to the Mn$^{2+}$ ion, are clearly seen. The electron density is also very strong for all ligands. When comparing the structure to MtDXR-FN, the movements of foremost Asp151 and fosmidomycin upon metal binding are clearly seen, see Figure 19D. A most interesting observation is also that when fosmidomycin binds, the octahedral coordination around the metal, which was seen in the MtDXR-MS structure, is
lost. The two oxygens of the (N-formyl-N-hydroxy) amino group of the inhibitor only replace two of the three water molecules, leaving one of the coordination sites empty. This could be due to the close contact between fosmidomycin and Glu153. It is also worth mentioning that though the metal binding residues, and the fosmidomycin oxygens, move in order to coordinate the metal, the position of the inhibitor’s phosphonate part, and the position of NADPH remain fixed. The density for the NADPH substrate is here very strong, as is that for the inhibitor, the active site residues, and the metal.

Figure 19. (A) The active site of MtDXR-MS shows the octahedral coordination of the Mn$^{2+}$ ion by Asp151, Glu153, Glu222, and three water molecules. (B) MtDXR-FN shows the position of fosmidomycin in the absence of a metal ion. A water molecule is placed between the inhibitor and the metal binding site. (C) The electron density in the active site of MtDXR-FMN is strong for all interacting residues and ligands. The SIGMAA-weighted (Read, 1986) $2|F_O| - |F_C|$ map was contoured at 1 σ (blue, 0.4 e/Å$^3$) and 5 σ (red, 2.0 e/Å$^3$). (D) Superpositioning of MtDXR-FN (pink) and MtDXR-FMN (gold), illustrating the movement of foremost Asp151 and fosmidomycin in the absence or presence of Mn$^{2+}$. 
When comparing the mutant enzyme structure \( \text{MtDXR}^{\text{NO}}-\text{S} \), containing the substitutions Asp151Asn and Glu222Gln, with the \( \text{MtDXR}-\text{S} \) structure, one can see that while the residue corresponding to Glu222 is unchanged, Glu153 is moved towards the metal binding area, and the residue corresponding to Asp151 is moved away from this area. The close contact between Asp151 and Glu222 in \( \text{MtDXR}-\text{S} \) suggests that one of these residues is protonated. In the mutant enzyme this interaction is lost, which can explain the change in conformation of the mentioned residues. The \( \text{MtDXR}^{\text{NO}}-\text{NS} \) and \( \text{MtDXR}^{\text{NO}}-\text{FN} \) structures show that although the mutant enzyme has lost its ability to bind the metal, which was also suggested in the assay by the loss of activity, it can still bind both fosmidomycin and NADPH. The density for the nicotinamide portion of the NADPH substrate is very weak in the \( \text{MtDXR}^{\text{NO}}-\text{NS} \) structure, though slightly stronger in \( \text{MtDXR}^{\text{NO}}-\text{FN} \). Also, for \( \text{MtDXR}^{\text{NO}}-\text{FN} \), Asp151 is moved even further away from the metal binding site.

By comparing all the \( \text{MtDXR} \) structures a few conclusions can be drawn. First of all, when comparing the corresponding wild type and mutant structures one can see that the density for NADPH is always worse in the mutant, despite the fact that there is no real contact between NADPH and the mutated residues. This suggests that the close contact between the metal binding residues, probably especially between Asp151 and Glu222, does have an effect on the enzyme conformation, in turn affecting the NADPH binding. All the structures in complex with NADPH also show that the adenosine part of the substrate, in particular the \( 2' \)-phosphate is better defined than the nicotinamide part. This is consistent with previous studies, which suggest that the \( 2' \)-phosphate is important for binding of the substrate (Argyrou, 2004).

**Mechanism**

The reaction catalyzed by \( \text{MtDXR} \), the production of MEP from DXP, is performed in two steps, an isomerization followed by an NADPH-dependent reduction. The intramolecular rearrangement, which leads to the formation of the intermediate 2-C-methylerythrose 4-phosphate, is thought to occur either via a retroaldolization or an \( \alpha \)-ketol rearrangement (Hoeffler *et al.*, 2002), see *Figure 20*. The retroaldolization mechanism starts with a proton abstraction at the O4 hydroxyl of DXP, followed by a deprotonation of the O3 hydroxyl. The latter step is also involved at the beginning of the \( \alpha \)-ketol rearrangement. \( \text{MtDXR} \) has also been proposed to be a class B dehydrogenase (Proteau *et al.*, 1999). Whether an enzyme is defined as a class A or B depends on which of the hydrogen atoms at the C4 of the NADPH nicotinamide ring is transferred in order to form the product, see *Figure 21*. How
then does all this information correspond to the interactions we can see both in our own, and in the other published DXR structures?

Possible reaction mechanisms for DXR

Retroaldolization

$\text{DXP}$

$\text{2-C-methylerythrose 4-phosphate}$

$\text{NADPH} + \text{H}^+ \rightarrow \text{NADP}^+$

$\text{MEP}$

$\alpha$-ketol rearrangement

$\text{DXP}$

$\text{H}^*$ hydrogen of MEP is derived from the NADPH cofactor. For clarity, the carbon numbering follows the convention for DXP. For the equivalent carbon numbering for the product MEP, see Figure 23A.

Figure 20. The DXR catalyzed production of MEP from DXP is carried out in two steps, an isomerization followed by an NADPH-dependent reduction. The intramolecular rearrangement of DXP to form the aldehyde intermediate 2-C-methylerythrose 4-phosphate is thought to occur via a retroaldolization or an $\alpha$-ketol rearrangement. The $\text{H}^*$ hydrogen of MEP is derived from the NADPH cofactor. For clarity, the carbon numbering follows the convention for DXP.
The conformations of fosmidomycin seen in MtDXR-FN and EcDXR-FN (PDB code 1Q0L) are equivalent. However, when comparing MtDXR-FMN with the E. coli structure in complex with fosmidomycin and Mn$^{2+}$ (PDB code 1ONP) (Steinbacher, 2003), it is clear that the conformations of the inhibitor seen in these two structures are different, see Figure 22A,B. In MtDXR-FMN the N-hydroxyl of fosmidomycin is placed opposite to residue Glu153, and the N-formyl oxygen sits opposite residue Glu222, while the coordination site opposite Asp151 is empty. In the E. coli structure, the residue corresponding to Glu222 still sits opposite the N-formyl oxygen of fosmidomycin, while the N-hydroxyl oxygen now is placed opposite the residue corresponding to residue Asp151, and the residue corresponding to Glu153 sits opposite a water molecule. The difference between the two structures can be due to the lack of NADPH in the 1ONP structure, i.e. it can be a real effect, or there can be a misinterpretation of the crystallographic data. The 1ONP structure was refined to a resolution of 2.5 Å, with rather high $R$ and $R_{\text{free}}$ values of 25.5 and 33.0%, respectively. The $B$ factors were also high, 69.3 Å$^2$ for the protein atoms, 64.7 Å$^2$ for the metal ion, and 77.7 Å$^2$ for the inhibitor. In their paper, the authors show a simulated annealing $|F_o| - |F_c|$ omit map covering the active site, illustrating the density for the metal and inhibitor. However, the density shown does not allow unambiguous assignment of the inhibitor conformation. Also, since the structure factors for the 1ONP structure have not been deposited it is impossible for us to evaluate which mode of ligand, and also water, binding, is in best agreement with their experimental data. Very recently, an E. coli structure in complex with Mg$^{2+}$, fosmidomycin and NADPH was deposited in the PDB with accession code 2EGH (Yajima, 2007). This structure was refined to a resolution of 2.2 Å with $R$ and $R_{\text{free}}$ values of 21.8 and 22.5%, respectively. The $B$ factors were 39.5 Å$^2$ for the protein atoms, while those for the ligands were much higher, 74.8 Å$^2$ for NADPH, 59.5 Å$^2$ for the metal ion, and 77.4 Å$^2$ for the
Figure 22. Fosmidomycin conformation, and metal coordination in (A) MtDXR-FMN and (B) EcDXR-FN (PDB code 1ONP), shown in stereo view. (C) Stereo view of the active site of MtDXR-FMN with models of DXP (magenta) and MEP (cyan) superpositioned onto our experimentally observed fosmidomycin structure. Golden bubbles show the observed interactions with the Mn$^{2+}$ ion. Magenta and cyan bubbles indicate the close contacts between the ion and the O4 and O3 hydroxyl groups of DXP and MEP, respectively. Indicated is also C4 of the NADPH cofactor.
inhibitor. The fosmidomycin conformation in this structure is equivalent to that seen in the 1ONP structure. However, when inspecting maps calculated using the crystallographic data for 2EGH, the weak density in the active site makes one question if the inhibitor in this case should have been modeled at all. Therefore, this structure will not be mentioned in further discussions.

The coordination around the metal ion has important mechanistic consequences. Three different studies of the DXRs from the cyanobacterium *Synechocystis* (Proteau, 1999), *E. coli* (Radykewicz et al., 2000), and *M. tuberculosis* (Argyrou, 2004) agree on that the pro-R hydrogen of C1 of MEP is derived from the 4S-H of NADPH in the reduction step. (For the naming convention for MEP, and also fosmidomycin and DXP, see Figure 23A.) However, when super positioning the substrate onto the fosmidomycin in the 1ONP structure, the mode of binding is not suitable for this stereochemistry. Instead, in order to explain the hydrogen transfer from NADPH, the authors propose a rearrangement of the intermediate around the metal, so that the N-hydroxyl oxygen moves into the position of the water molecule (Steinbacher, 2003). This post-rearrangement conformation is actually directly equivalent to that we believe already exists based on modeling both DXP and MEP binding using the fosmidomycin conformation seen in MtDXR-FMN, see Figure 22C and Figure 23B. Here, the hydrid transfer from C4 of NADPH can take place so generating the stereochemically correct product, and this without major rearrangement around the metal ion. The hydrogen atom is also transferred from a position of the C4 in the nicotinamide ring that would be appropriate if DXR is a class B dehydrogenase, as proposed earlier, and which is in agreement with the experimentally determined course of the reaction (Argyrou, 2004; Proteau, 1999; Radykewicz, 2000).

![Figure 23. (A) Carbon atom numbering for fosmidomycin, DXP and MEP. (B) The conformation of fosmidomycin around the metal is illustrated as well as the modeled conformations of DXP and MEP. The H* hydrogen of MEP, which protrudes from the O1-C1-C2 plane, is derived from the NADPH cofactor.](image)
In addition to showing that our model for substrate binding is suitable for the hydrid transfer from NADPH, our studies also provide other mechanistic insights. First of all, the O4 hydroxyl of our modeled DXP comes in close contact with the unoccupied Mn$^{2+}$ coordination site, suggesting that all six coordination sites are occupied during the reaction. The O4 hydroxyl also comes in close contact with the carboxylate group of Glu153, which makes this residue ideally positioned for the proton abstraction that leads to the formation of the intermediate via the retroaldolization mechanism. Further, Glu222 is suitably positioned for deprotonation of the DXP O3 hydroxyl group, which is required at some stage for both of the proposed mechanisms.

Studies have been performed, with the goal of determining the mechanism, where the O3 and O4 hydroxyl, one at a time, were excluded from the composition of the DXP substrate (Hoeffler, 2002). Unfortunately though, these compounds all acted as inhibitors rather than substrates, and therefore no conclusions could be drawn.

Likelihood of proton abstraction by Glu153

Molecular dynamics simulations and free energy calculations were used to evaluate the likelihood of the proton abstraction step by Glu153, which would lead to the formation of the intermediate via the retroaldolization mechanism. The first step in this process was to dock the substrate DXP into the MtDXR-FMN structure, from which fosmidomycin had been removed; hydrogen atoms had been added to the protein, and the Mn$^{2+}$ ion had been replaced by a Mg$^{2+}$ ion. During the docking process, the protein was kept rigid, while the substrate was allowed to flex. A genetic algorithm was cycled during the docking. Here, the top solutions were combined at various stages throughout the cycling, and in this way the program tried to find the optimal centre of mass and torsions associated with the molecule. As output, a number of suggested conformations were given for the substrate. These were ranked depending on the energy associated with their specific interactions with the protein. For example, the number of hydrogen bonds, and also their distances, is taken under consideration in the scoring process. The calculated energy scores combined with available experimental data were used to select a protein-substrate complex.

To run a molecular dynamics simulation, that will give a picture of how the substrate and protein move as a function of time, several things are required. First of all, the protein-substrate complex has to be solvated, and the force field, which describes the energy of the system, has to be defined. The calculations of the force field energy include terms for covalent bonds, angles and torsion angles, electrostatics, and van der Waals interactions. Once defined, the system is computationally "heated up" and then equilibrated at a lower temperature in order to prepare for the free energy calculations. During this project, the free energy profiles were calculated, both for the proton
abstraction by a single glutamic acid in water, and also for Glu153 in the
MtDXR structure. In this process, the different states describing the proton
transfer are simulated, and the corresponding free energies are calculated at
each step. If the energy barrier for the catalyzed reaction, that is the one cal-
culated for the glutamic acid within the protein, is lower than that for the free
glutamic acid in water, the protein is predicted to exert a catalytic effect.

In the case of MtDXR and Glu153, the transition state for the reaction
was calculated to be stabilized by approximately 10 kcal/mol compared to
the uncatalyzed reaction. Therefore, Glu153 does have a distinct catalytic
effect on the reaction, which in turn emphasizes the possibility of a retroal-
dolization mechanism.

To learn more about the true route of the reaction, molecular dynamics
simulations and free energy calculations could be performed for all steps of
both reactions, to see if one could see a significant difference between them.
However, this would take a lot of time and effort, and it is likely that one
still might not be able to obtain a clear answer.

Thoughts about inhibitor design

By investigating the interactions in the active site to different substrates or
inhibitors, as well as the space around them, many clues that will be impor-
tant in the design of new inhibitors can be found. Fosmidomycin has been
shown to be active against Plasmodium species (Jomaa, 1999; Lell, 2003;
Missinou, 2002), but has not shown any activity against M. tuberculosis
(Dhiman, 2005), which we now know is due to the lack of uptake (Amanda
Brown and Tanya Parish, personal communication). It has been suggested
that the only possibility of producing extended fosmidomycin analogues
would either require a displacement of the NADPH nicotinamide ring, or
targeting of the open conformation of DXR (MacSweeney, 2005). By ana-
lyzing the structures presented here, we suggest otherwise. We believe that
one of the fosmidomycin-like compounds that has already been produced,
and which is more active than fosmidomycin against P. falciparum
(Haemers, 2006), does not extend into the NADPH binding-site. Instead we
believe that the compound, which has a 3,4-dichlorophenyl substitution at
the C4 atom of fosmidomycin, extends into a hydrated cavity close to the
phosphonate part of fosmidomycin. In the MtDXR-FMN structure, several
well-defined residues line this cavity, which traps four water molecules. In
order to visualize the interactions of the analogue in the active site, a model
was built and superimposed onto fosmidomycin, and then regularized, see
Figure 24. The analogue has, as far as we know, not been tested on myco-
bacteria. However, the presence of the phosphonate group makes it unlikely
that it would be taken up by the bacteria, and hence it will probably not show
biological activity.
Figure 24. Active site of MtDXR-FMN, shown in gold, with a super-positioned model of the fosmidomycin analogue with a 3,4-dichlorophenyl substitution at the C4 atom of fosmidomycin, shown in pink. The calculation of the pocket present in the MtDXR-FMN active site was performed with the *asa* (accessible surface area) command in *O* using a probe with a radius of 1.4 Å. Water molecules lining the pocket are shown as small red spheres. The analogue was also built in *O* using the *qds* (Quick and Dirty Sketcher) commands. In panel (A) one can see that the analogue fits well to the conformation of our experimentally determined fosmidomycin, and in panel (B) one can see the residues lining the pocket.
So what should one aim for when designing a new inhibitor towards \( Mt \)DXR? Well, first of all one could, of course, try to produce a fosmidomycin analogue where the phosphonate group has been substituted for a group more compatible with uptake. However, since the enzyme shows a lower conservation in the residues involved in NADPH binding, I believe that a good approach would be to target this area. Others have already performed initial attempts along these lines, which have produced mimics that bind with \( \sim 200 \) nM affinity (Sem et al., 2004). A more sophisticated approach would be to design an inhibitor that would span both the DXP and NADPH binding sites at the same time, an approach that does not seem to have been attempted to date.

Future perspectives

Together, the \( Mt \)DXR structures presented here both provide us with insights into the dynamics associated with the binding of substrates, and give us new information about the catalytic mechanism. However, when it comes to the question of drug design, the information that these structures provide is merely the first step on the challenging road to finding a new candidate drug.

One method, which has been used during the last ten years in the search for new inhibitors, is fragment-based drug design (Hajduk et al., 2007). Here, thousands of low molecular weight compounds are screened in order to find low affinity fragments that interact within a specific area of the active site. In order to evaluate the binding, two different methods can be used. The first is two-dimensional isotope-edited nuclear magnetic resonance spectroscopy, where the chemical shifts caused by ligand binding are identified. The second method is X-ray crystallography, which can show the different interactions in detail. By combining different fragments that bind in the active site, a new inhibitor can be designed. Structural studies on the new compound may confirm the mode of binding and suggest new modifications to optimize interactions in the active site. So far, for \( Mt \)DXR, we have only tried this method on a pilot scale with approximately 30 compounds, which were chosen based on pure intuition and experience. Unfortunately, due to solubility problems associated with the compounds, the experiments were inconclusive, and due to time limitations, these efforts are for the present put on hold. Co-crystallization has yet not been attempted, however, the binding was tested in a thermodfluor experiment. In this approach, binding is measured by following the increase in melting temperature. In all the measurements that showed a change in melting temperature, the change was within the margin of error, and consequently no conclusions could be drawn. Although our focus for the time being is directed elsewhere, fragment-based drug design is a method well worth testing further.
Interesting to note here is that most MtDXR structures crystallized in the same space group, even though different substrates and inhibitors were bound; conformational differences in one of the molecules in the asymmetric unit could be accommodated without disrupting the crystal packing. This shows that the crystals formed under these conditions are suitable for future soaking experiments on a relatively high throughput basis. This will greatly facilitate work with fragment libraries, as well as with designed inhibitors and those resulting from screens (see below).

The chemists at Uppsala University, from the medicinal (Daniel Muthas and Anders Karlén) and computational (Jens Carlsson and Johan Åqvist) sections have performed a virtual screen in order to identify new MtDXR inhibitors. In such a screen, several million compounds are tested in order to find a suitable set to be used in further experiments, see Figure 25. One advantage, especially when combining structure-based drug design with virtual screening, is that the number of compounds that need to be tested before identifying one that causes inhibition can potentially be reduced. This approach has been widely used in the pharmaceutical industry, see (Perola, 2006) and references therein. For MtDXR, the original set of compounds chosen by the medicinal chemists consisted of a library of chemicals that are commercially available. These were passed through a filter where the selection was made on molecular weight and the number of rotatable bonds. In the next selection filter, physico-chemical traits like polarity, size, and the number of hydrogen bond donors and acceptors were taken under consideration. The compounds were then compared to the conformations of the inhibitor fosmidomycin, and parts of the NADPH substrate, as positioned in the MtDXR-FMN structure. Those that showed the best similarity were passed on through the screen. At the next stage, a high-throughput docking was performed, in which the test set was expanded to include different enantiomers, tautomers, and protonation states for the compounds. During docking, the compounds were placed in the active site of the MtDXR-FMN structure, see Figure 26. The compounds were allowed movements by translations and rotations, their torsion angles were allowed to flex; several different predicted conformations were saved for each compound. These were then put through a pharmacophore filter, where they again were inspected, looking for certain desired interactions in the active site, for example in this case interactions to a metal ion. Several compounds were excluded on this basis, and for those that passed, the top conformation was chosen for molecular dynamics simulations and free energy calculations, performed by the computational chemists. The predicted energy associated with placing the compound in solution was here compared to that calculated for the compound placed in the MtDXR-FMN active site. The results were passed back to the medicinal chemists, who made a manual selection of the final set of compounds, based on for example solubility calculations; only one compound was selected, where several highly similar compounds were present in
the list. The remaining compounds will now be tested for inhibition, and if a hit is found, co-crystallization with the \( Mt \)DXR enzyme will be the main goal for us, as structural biologists.

The entire process of finding a new lead compound can take up to a decade, and sometimes even longer, and many compounds are excluded at different stages. One critical point is moving from studies performed on test animals, to phase I, II or III studies, where the compounds are tested on humans. At the time of writing, approximately ten anti-mycobacterial compounds are either in preclinical or clinical trials (information obtained from http://www.stoptb.org/globalplan/, Global Plan to Stop TB: 2006-2015, Strategic Plan).

\( \text{Figure 25. Flowchart illustrating the process of virtual screening here used for MtDXR.} \)
The results presented and discussed in this thesis indicate that MtDXR is an excellent drug target in *M. tuberculosis*. The structural work thus provides a solid basis for the design of a new drug with anti-mycobacterial activity.
Populärvetenskaplig sammanfattning


Idag används flera mediciner mot tuberkulos, av vilka de flesta kan ge upphov till obehagliga bieffekter såsom leverskador, illumående och yrsel. En av medicinerna kan till och med fårga urin, tårar och saliv oranga. Detta, i kombination med den ökande delen av resistenta, multiresistenta och även extremt-multiresistenta tuberkulosbakterier, gör behovet av nya läkemedel stort.


ett protein, kan det vara svårt att, på grund av den volym som alla aminosyror upptar, få en ordentlig bild av hur det egentligen ser ut. Man kan därför välja att bara visa aminosyrorna som ”pinnar” som i Figur 27A. Trots att detta redan är en förenklad bild ser den fortfarande ganska grötig ut. Därför brukar man visualisera proteinet genom att visa dess α-helixar och β-flak, se Figur 27B.

*Denna illustreringen sker under anslutning till texten och visar protein i två olika former - fig. 27A och 27B.*

**Figur 27.** Om alla atomer i ett protein visas (A) kan det vara svårt att se dess generella form. Genom att istället illustrera proteinet genom att visa dess α-helixar och β-flak (B), kan detta underlättas. De två α-helixarna visas som ljusgrå spiraler, medan β-strängarna visas, i en mörkare grå ton, som pilar, vilka går från N-terminalen till C-terminalen.


RAPID (Rational Approaches to Pathogen Inhibitor Discovery) programmet vid Uppsala universitet startade officiellt i januari 2003, med målet att ta fram nya läkemedelskandidater mot främst tuberkulos och malaria. Inom programmet kombineras strukturologi och medicinsk kemi med datorbaserade modellerings och simuleringar av kemiska processer och bindningsenergier mellan proteiner och de ämnen de binder. Vi ingår även i ett EU samarbete om tuberkulos och samarbetar med AstraZeneca i Bangalore, Indien, som har ett stort egenintresse av denna typ av forskning.
Figur 28. I (A) visas en mycket förenklad bild av ett enzym som är omgärdat och även binder till sitt substrat. I (B) visas hur ett läkemedel, vilket kan ses i svart, binder till enzymet, trots att substratet finns i dess omgivning och hindrar det därmed från att fungera normalt.

Tuberkulobakterien karakteriseras bland annat av sitt komplexa cellmembran och sin mycket långsamma tillväxt. Den har även förmågan att kunna ligga latent inuti sin värdorganism för att sedan reaktivera decennier efter den initiala infektionen. Bakteriens genom, som finns i en enda cirkulär kromosom, innehåller cirka 4000 gener. Bland dessa finns gener som ger upphov till proteiner, som är viktiga för att sköta bakteriens metabolism, respiration och som deltar i uppbyggnaden av cellmembranet.


Det övergripande målet med vår forskning är att bidra till utvecklingen av nya hämmerare som binder mer effektivt än de som finns idag och som förhoppningsvis ska vara betydligt skonsammare som läkemedel.

Det kan finnas olika anledningar till varför man vill få tillgång till en viss proteinstruktur. Kanske finns det sedan tidigare ingen liknande känd struktur, vilket gör att den kommer att bidra till kunskapen om hur proteiners aminosyrasekvens påverkar dess struktur. Man kan även få kunskap om hur en viss process utförs. I de fall funktionen är okänd kan denna försöka utredas genom närmare analys av strukturen och speciellt av det område som förmodas vara proteinet aktiva yta. Då det handlar om att ta fram nya läkemedelstyper vill man naturligtvis försöka inhibera proteiner som är nödvändiga för bakteriens överlevnad. Det är dessa proteinstrukturer vi är intresserade av. I denna avhandling presenteras strukturer av två olika tuberkulosproteiner, peptidyl-prolyl cis-trans isomeraser A (MtPpiA) och 1-deoxy-D-xylulos-5-fosfat reduktoisomeras (MtDXR).

Hos tuberkulosbakterien finns två Ppi protein, MtPpiA och MtPpiB. Uppgiften för dessa proteiner är att sköta isomeriseringen av peptidbindningen i proteinet på positioner före aminosyran prolin. Denna funktion är viktig för att bakteriens egna proteiner ska kunna veckas rätt, det vill säga upptaga rätt tredimensionell struktur för att kunna utföra sin uppgift. MtPpiA finns inuti...
bakterien medan MtPpiB sitter fast i cellmembranet och har sin funktionella del, den med aktiva ytan, utanför bakterien. I studier som har gjorts har man fastställt att MtPpiB är nödvändig för bakteriens överlevnad och att MtPpiA är viktig vid infektion. Målet med detta projekt var att ta fram strukturerna för båda dessa proteiner, för att kunna titta på deras bindningsyta och för att kunna jämföra dem. På grund av att vi inte lyckades uttrycka lösligt protein av MtPpiB, vilket är nödvändigt för kristallisering, presenteras i denna avhandling enbart strukturen av MtPpiA. Förutom att beskriva de experimentella steg som är nödvändiga för att erhålla en struktur visade vi även att proteinet är aktivt, det vill säga att det är Funktionellt och kan utföra sin reaktion. Vi visar även att proteinets struktur består av 8 β-strängar som tillsammans liknar en ”tunna”, med en α-helix på vardera sidan, se Figur 27B. Den aktiva ytan är här positionerad på ena sidan av ”tunnan”. Om aminosyrorna i denna yta jämförs med de som kan ses i Ppi-strukturer från andra organismer kan man se att de är väldigt lika. Detta gäller även om man jämför med det humana proteinet. Detta betyder att MtPpiA inte är lämpligt som målmolekyl för ett läkemedel. Om man istället tittar på aminosyrasekvensen för MtPpiB så skiljer sig denna både från MtPpiA och de humana proteinen. På grund av detta lämpar sig detta protein bättre för läkemedelsdesign och om tid fanns skulle det vara högst intressant att åter igen försöka uttrycka ett lösligt protein.

MtDXR är ett enzym som i bakterien deltar i en process som leder till bildandet av en mängd ämnen, av vilka en del hjälper till att bygger upp cellmembranet. Hos oss människor sköts motsvarande process av helt andra proteiner, som inte alls liknar bakteriens. Detta gör att alla bakteriens proteiner, som deltar i denna specifika process, är lämpliga mål för nya läkemedelskandidater. MtDXR binder två substrat, 1-deoxy-D-xylulos 5-fosfat (DXP) och NADPH. Dessutom binder proteinet en metalljon som behövs för dess aktivitet. Under reaktionen kommer bland annat NADPH att lämna ifrån sig en väteatom till DXP intermediatet och på så sätt formas produkten 2-C-methyl-D-erythritol 4-fosfat (MEP).

I denna avhandling presenteras åtta MtDXR strukturer där enzymet kan ses i olika kombinationer med substraten och metallen. Tre av strukturerna representerar även ett protein där två av de tre aminosyror som binder till metallens har muterats. Detta för att kunna se om de verkligen är viktiga för aktiviteten, vilket de visade sig vara. I två av strukturerna finns även fosmidomycin bundet. Fosmidomycin är en hämmare som har visat sig inhibera MtDXR, men som trots detta inte är intressant som läkemedelskandidat eftersom den inte kan ta sig igenom tuberkulosbakteriens cellmembran. Strukturerna kan däremot fortfarande ge mycket värdefull information när det gäller designen av nya läkemedel. En proteinmolekyl av MtDXR, även kallat en subenhet, består av tre så kallade domäner, vilka kan beskrivas som tre individuellt rörliga områden av proteinen, se Figur 30A.
Metalljonen och fosmidomycin (eller DXP) binder i den katalytiska domänen under en loopregion, medan NADPH är positionerad längs med den N-terminala domänen, se Figur 30B.

I två artiklar visar vi att vårat MtDXR enzym är aktivt och att det även inhiberas av fosmidomycin. Vi visar även att i vår struktur av enzymet i komplex med fosmidomycin, metall och NADPH, så skiljer sig positionen av hämmaren från det som tidigare har beskrivits i en struktur av E. coli enzymet i komplex med fosmidomycin och metall. Detta innebär att man på nytt kan titta på hur reaktionen går till, t.ex. hur väteatomen från NADPH kommer i kontakt med intermediet. Baserat på E. coli strukturen så måste intermediet flippa runt metaljonen för att rätt kontakt ska kunna uppstå. I vår struktur behövs dock inte detta utan positionen är redan rätt för att kunna bilda produkten på det sätt som tidigare studier har fastställt som korrekt. Då diffractionsdatat resulterar i en elektrondensitet och inte en exakt position av varje atom kan resultatet misstolkas om densiteten är svag och atomer kan placeras fel. I våran struktur är densiteten stark i aktiva ytan och positionen av både fosmidomycin, metall och NADPH är tydliga. I dagsläget finns det två olika föreslagna reaktionsmekanismer som beskriver hela förloppet av bildandet av MEP från DXP. Från vår struktur kan vi se vilken aminosyra som är lämplig för att initiera den ena av dessa mekanismer. Möjligheten av detta steg undersöktes med molekyldynamik och energiberäkningar, vilka visade att MtDXR faktiskt katalyserar detta steg, vilket i sin tur innebär att mekanismen är möjlig.
MtDXR projektet lever vidare även efter det att denna avhandling skrivs. Nästa steg är att på olika sätt få fram nya potentiella läkemedelskandidater för vilka graden av inhiberingen ska fastställas. Om en bra inhibitor hittas kommer denna att kristalliseras tillsammans med proteinet och på så sätt ge ytterligare information om interaktionerna i aktiva ytan.
Acknowledgement

First of all I would like to thank my supervisors and “the boss” for their inspiration, knowledge, for sending me to so many interesting courses and conferences, and for just being terrific people. Torsten, thank you for all the help in the lab and for buffers prepared, for being an “angel”, and for managing to combine a never-ending enthusiasm with a never-ending stream of ideas. Sherry, I have really enjoyed all the great discussions we have had throughout the years! Thank you for knowledge shared, for being a “word-magician” and for helping me correct the slightly unspherical waters. To Alwyn, thank you for clogs in the corridor, for bringing me updated versions of $O$ when we wanted to try something new, and for good discussions and tricky questions, even if they sometimes made me sink lower and lower into my chair!

I would like to thank Tanjore Balganesh, Srinivasa Bachally and Sunita DeSousa at AstraZeneca in Bangalore, India, and Amanda Brown and Tanya Parish, Queen Mary’s School of Medicine and Dentistry, for a good collaboration during the DXR project.

To collaborators at Uppsala University, for help with the same project, I would like to express my gratitude to Jens Carlsson for performing molecular dynamics simulations and free energy calculations, and especially for explaining it all. To Johan Åqvist for discussions concerning the reaction-mechanism, and to Daniel Muthas (DMT) and Anders Karlén for performing, and also discussing the principles of virtual screening.

I also want to thank Patrik Johansson for helping me solve my first structure, that of PpiA, Christofer Björkelid for optimizing the expression of DXR, and Anna Suárez Larsson for continuing the DXR work, while I was writing my thesis.

To the other members of “Tottes angels”, also known as my very nice roommates. Alina, thank you for dessert-buffets and swimming pools, for crumbles in bed and New York stairs, and for being able to talk about absolutely everything. To Annette, thank you for sound effects sometimes joined by little jumps, for color coordinating yourself with your posters, for keeping our plants alive, and for being there for others. I can genuinely say that this experience would not have been the same without the two of you!
Many thanks to both former and present members of xray: Jimmy for taking such good care of me during my under-graduate project and for having such a nice attitude, Nina for all the help and sanity checks in the lab (I believe I owe you a lot of small amounts of various buffers), and for all sorts of excursions, Patrik for answering a lot of questions, for nice lunches and times at the ESRF, and for together with Alina bringing me climbing for the first time, Seved for “gel-swapping”, singing in the lab, and for always dancing at parties, Daniel for being an excellent teaching companion, for “arga leken” and for preparing very cool pictures, Henrik for many nice lunches and for really being one of these who know how to enjoy life, Evalena for not seeing the men in black, for “shit-pom-frit” and for nice times at the beamline, Anna for self-invited “parmiddagar”, for “apparently”, and for always finding the good comments in time, Magnus for entertaining e-mails and for Christmas-decoration enthusiasm, Wojciech for “the damn structure” and for being able to smell when it’s cake, Christofer for taking care of “glassklubben”, Nisse for excel-sheets and for having such a cool suit, Pavel for making a virus-lamp, Adrian for salsa dancing, Anna for preparing me for assay questions, Erling for solving all my acute computer problems, Tex for crystallization tips and for “as we speak”, Maria for updates on house-shopping and for having a great dance-style, Fariborz for skating and his interest in the “red-heads”, Mark for answering MolRay questions, Karin, Margareta, Lars, Jerry, Inger, Ulla and Sara for songs and glitter in the annual Lucia-tåg, “Åqvist-grabbarna”: Johan, 3×Martin, Jens, Sinisa, Stefan, Viktor, Lars, Göran and Johan for being such nice guys, and for finishing all cakes, Alexandra for hours spent on the dance floor, and late weekend study times, Ulrika, Linus, Martin H, Lotta, Martin, Louise,
Agata, Gösta, Hasse, Gerard, Mats, Christer, Sanjee, Wimal, Jonas and all other prior and present members of alla@xray: Thank you for “fika”, lunches and for creating such a nice workplace!

Thank you Sigrid, Ingrid, Sofia, Christina, Susanne and Åsa for solving economical issues, registering points, and ordering books.

Stort tack till gamla gymnasiegänget: Sofia, för att vara min ”KS”, tänk vad mycket vi har upplevt tillsammans, Marcus, för att vara den kramgoa gitarr-killen, Johan för att annordna annandagsfika och Andreas, Tobias, Sven, Samuel m.fl. för att ni fortfarande dyker upp!

För utmärkt sällskap, intensiva diskussioner och en hel del guacamole, tack till giraff@fri.nu: Maria och Dan för middagar, shoppingturer, utflykter, skavsarplaster och för att vara helt underbara personer och vänner, Ulrika och Robert för otroligt roliga stunder i Forskå, för ”den blondaste”, ”den var närmast inverterad” och för express-shopping på coop (”give me a påse”), Solveig för att låna ut ”reservböckerna” och för att nu senast ha guidat mig runt i England, Kalle och Ylva för att veta hur man åter paj, tittar på fotboll och spelar spel, Nisse för att helt enkelt vara Nisse och även en räddare i nöden, Gustav och Katta för flipperspel och för att lyckas chocka Matte, Nina och Martin för att vara så mysiga, Erik för att vara en riktig akademi-ker och Susanna för lån av Schnappi, Henrik för att figursåga sin dator, Hattne och Malin för att visa oss rund i Hamburg, Martin och Malin för Danmarksutflykt, Ola och Jonas för ”das kleine grüne kaktus”, Catharina och Hanna för att vara så fina tjejer och Jocke för spännarlinnen.

Till “knitting kittens”: Nina, Maria, Ulrika, Solveig, Nina and Johanna. Ett stort tack för att ni verkligen förgyller min fritid, för utsökt fika och naturligtvis - bra skvaller!

Till familjen Andersson/Dahlström/Lundh, tack för många trevliga middagar med både utsökt mat och sällskap och tack för att ni har välkomnat mig in i er familj! (Tack även för ”sådärja” och ”basic untrust”.)

Familjen Ekström/Faming, tack för vandringar, “uriner – jag menar rui-ner”, Viktorialopp, köttgrytor och sockerkakor, hamstern som hängde sig, Möllewood filmer och för genuint trevligt och underhållande sällskap!
Till mormor Majken och morfar Egon. Tack morfar för alla vackra tavlor och trähantverk och för att vara den mest genomgoda person jag någonsin känt. Tack mormor för virkade dukar, kroppkakor, telefonsamtal och för att du är den mest envisa person jag känner. Det kan vara så att jag har ärvt lite av det, eller vad tror du? Tack båda för all tid ni har tillbringat med oss, för svälta räv och yatzy och för att vara helt underbara!

Systrarna: Sara, tack för “sugar pie, honey bun”, för träor med matchande accessoarer, för “ge gubben kolan” och för att du korrekturläste delar av avhandlingen. Lisa, allas vår lilla solstråle, tack för att du spelar spel med mig, för att “vara cool”, för all härlig musik och för att du låter mig sjunga till den. (Naturligtvis också tack till Trisse och Slangen för morrhår till mina kristallisationsexperiment.)

Mamma Ingrid och pappa Anders: Tack mamma för alla skjutningar, alla telefonsamtal, att du har lärt mig sy, sticka och virka och för att du vet hur man fnittrar. Tack pappa för alla enhetscirklar, ormsalvor, påhittade låttexter och för att du alltid kollar att jag inte stressar för mycket. Till er båda, tack för 90 liter lingon, för jeansshoppar, för att ni alltid har uppmuntrat mig att prova och lära mig nya saker och för att ni helt enkelt är så bra föräldrar! Bara det faktum att erat vardagsrum numera har förvandlats till ett musikrum säger allt!

Till Mattias, min älskling och “the love of my life”. Tack för att du verkligen är en riktig vän, en som inte bara ställer upp för mig utan även för andra. Tack för glass till kvällsmat, för att du inte tycker att jag pratar för mycket och för att du får mig att känna att jag kan klara av det mesta. Jag älskar dig, puss puss!
References

Addington, W. W. (1979) The side effects and interactions of antituberculosis drugs Chest 76, 782-784


Calmette, A. La Vaccination Preventive Contre la Tuberculose (1927) Masson et Cie, Paris


Evans, P. R. (1993) in Proceedings of CCP4 Study Weekend on Data Collection and Processing 114-122, Daresbury Laboratory, Warrington, England


berculosis growth at the cavity surface: a microenvironment with failed immunity Infect. Immun. 71, 7099-7108


Koch, R. (1890) Weitere Mittheilungen über ein Heilmittel gegen Tuberculose Dt. med. Wochenschr. 16, 1029-1032

Koch, R. (1891a) Fortsetzung der Mittheilungen über ein Heilmittel gegen Tuberculose Dt. med. Wochenschr. 17, 101-102


Koch, R. (1891c) Weitere Mittheilung über das Tuberkulin Dt. med. Wochenschr. 17, 1189-1192


98


Virchow, R. *Die Cellularpathologie in Ihrer Begründung auf Physiologische und Pathologische Gewebelehre* (1871) 4August Hirschwald, Berlin


A doctoral dissertation from the Faculty of Science and Technology, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology”.)