Structural studies of R2 and R2–like proteins with a heterodinuclear Mn/Fe cofactor and enzymes involved in *Mycobacterium tuberculosis* lipid metabolism.
Structural studies of R2 and R2–like proteins with a heterodinuclear Mn/Fe cofactor and enzymes involved in *Mycobacterium tuberculosis* lipid metabolism

Charlotta Selina Andersson
Till pappa
Abstract

Tuberculosis is a notorious disease responsible for the deaths of 1.4 million people worldwide. A third of the world’s population is infected with *Mycobacterium tuberculosis*, the bacterium causing the disease. The increase of multi drug–resistant strains worsens the situation, and the World Health Organization has declared tuberculosis to be a global emergency. The bacterium envelopes itself with a unique set of very long-chain lipids that play an important role in virulence and drug resistance. Therefore enzymes involved in lipid metabolism are putative drug targets.

To allow entry into different metabolic pathways and transmembrane transport, fatty acids have to be activated. This is done primarily by fatty acyl–CoA synthetases (ACSs). We identified an ACS possibly involved in the bacterium’s virulence and solved its structure. Structural interpretation combined with previously reported data gives us insights into the details of its function. This enzyme is known to harbor lipid substrates longer than the enzyme itself, and we now propose how this peripheral membrane protein accommodates its substrates.

Some of the most chemically challenging oxidations are performed by dinuclear metalloproteins belonging to the ferritin–like superfamily. We show that the ferritin–like protein, R2lox, from *M. tuberculosis* contains a new type of heterodinuclear Mn/Fe cofactor. This protein cofactor is capable of performing potent 2–electron oxidations as demonstrated by a novel tyrosine–valine crosslink observed in the protein.

Recently a new subclass of ribonucleotide reductase (RNR) R2 proteins, was identified in the intracellular pathogen *Chlamydia trachomatis* containing the same type of Mn/Fe cofactor mentioned above. The RNR R2 proteins use their metal site to generate a stable radical, essential for the reduction of ribonucleotides to their deoxy forms, the building blocks of DNA. With this work, we were able to characterize the architecture of this metal cofactor.
List of Papers

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals.


IV  Andersson C.S., Berthold C., and Högbom M. A dynamic C-terminal segment in the *Mycobacterium tuberculosis* Mn/Fe R2lox protein can assume a helical structure with possible functional consequences. *Submitted*

Reprints were made with permission from the publishers.
Contents

1 Introduction ................................................................. 13
2 Short Summary ......................................................... 15
3 Metals in Biology ....................................................... 17
   3.1 Metalloproteins .................................................. 17
   3.2 Challenges of the field ......................................... 20
   3.3 Studying metalloproteins ...................................... 22
   3.4 Ribonucleotide reductase ....................................... 24
   3.5 R2 – like proteins .............................................. 30
4 Tuberculosis .............................................................. 35
   4.1 Mycobacterium tuberculosis ..................................... 36
   4.2 Lipid Metabolism of M. tuberculosis ......................... 38
   4.3 A very long–chain fatty acyl–CoA synthetase from Mycobacterium tuberculosis is a peripheral membrane protein (Paper IV) .......... 41
5 Concluding Remarks and Future Perspectives ................. 43
6 Svensk populärvetenskaplig sammanfattning .................... 45
7 Acknowledgements ...................................................... 47
Bibliography ................................................................. 51
1. Introduction

Biochemistry describes the chemistry of life, defined by its properties of self-replication and catalysis. The biochemistry of a living organism is diverse and complex, with as many as 1,000 – 2,000 chemical reactions involved only in cellular self-replication [1]. A living organism is built up from a few classes of molecules: nucleotides, carbohydrates, proteins and lipids. Together with organic and inorganic compounds, these molecules are responsible for maintenance of the cellular functions. Lipids are for example important structural components of the cell membranes, shielding the cellular interior from the environment. Carbohydrates are diverse molecules, which may serve for energy storage or as structural components. Proteins are macromolecules that facilitate biological function. A group of proteins called enzymes are responsible for catalyzing the chemical reactions of a cell. A special group of enzymes converts the building blocks of RNA into corresponding elements for DNA by using radical chemistry. As we will discover, this type of challenging chemistry requires the shielding environment of a protein.

Life as we know it is based mainly on DNA genomes with proteins functioning as catalysts [2]. It is believed that in early life RNA fulfilled the function of both DNA and proteins, being able to hold genetic information and to perform catalytic activity. Introduction of variation by means of natural selection is what drives evolution according to the Darwinian theory [3]. The genes that contribute to a high individual "fitness" will have a higher chance of replicating than those that give deleterious effects. Divergent evolution occurs when environmental changes force a common ancestor to take different paths and evolve into gene products with different functions. Biological systems have adapted from an oxygen-free world to being capable of functioning in the presence of oxygen. As a result, evolution has generated divergent protein structures that reveal similarities between families where function may differ. For example, the four-helical bundle of the ferritin-like superfamily is found in proteins of diverse functions, ranging from iron-storage to proteins which perform challenging substrate oxidations [4].
Pathogens

Pathogenic bacteria may cause serious disease in humans, sometimes with death as the only outcome. Of all microbial pathogens *Mycobacterium tuberculosis*, the causative agent of tuberculosis, has the highest disease burden with over 2 billion people infected [5]. Another common pathogen is *Chlamydia trachomatis*, responsible for ocular and sexually transmitted infections in humans [6]. The two intracellular pathogens *M. tuberculosis* and *C. trachomatis* can be studied *in vitro* or *in vivo*, by the use of cell cultures or in mouse/rat cell–lines. However, the complexity of the interplay between pathogen and host are not easily interpreted. All knowledge on bacterial activities provides important information about determinants involved in virulence and pathogenicity.

Most antibacterial substances present to date are harmful and toxic to the environment and host [7, 8, 9]. Severe side–effects can lead to prolonged treatment and hospitalization of the patients. It is therefore important to identify drugs specifically targeted against the infectious pathogen. There is also an alarming emergence of multi drug–resistant bacterial strains, creating a demand for new drugs. Antibiotics are often targeted against the bacterial proliferative system or its lipid metabolism, as inhibitors of enzymatic activity. Understanding of a protein structure facilitates development of new substances that can specifically inhibit protein function.
2. Short Summary

This doctoral thesis describes my work and the main contributions to our knowledge of well characterized as well as more poorly understood biochemical pathways. I aimed to describe and understand structural features of enzymes with known or proposed function. My main focus was the structural characterization of enzymes that appear to be important to the survival of pathogenic microorganisms.

The work was initially focused on enzymes involved in the lipid metabolism of *Mycobacterium tuberculosis*. I studied the enzyme FadD13, which activates very long-chain fatty acids. We identified a patch that is important for membrane interaction, and propose a function for the catalytic activities. This provided new insights into how interaction with membranes may allow for acceptance of substrates larger than the protein.

Additionally, I have characterized a novel group of enzymes, the R2lox family, suggested to take part in the fatty acid metabolism of several extremophiles and pathogens. This work was also the first to determine the arrangement of a heterodinuclear cofactor from an Mn/Fe containing enzyme. The work on R2lox lead us to investigate the true nature of another Mn/Fe enzyme, involved in the *de novo* synthesis of deoxy-nucleotides, the building blocks of DNA. The Mn/Fe proteins studied show great diversity in how they perform their challenging biochemistry. Despite going into this large scientific field we have still managed to contribute novel results. The results produced in this thesis have provided new insights on radical generation and how evolutionary adaptions contribute to a higher diversity among bacterial and archaeal species.

The main method used throughout my work has been X-ray crystallography. With the use of advanced crystallographic methods we were able to elucidate the architecture of metal cofactors. The ultimate goal was to structurally characterize proteins in pathogens that are important for the pathogenicity and might therefore be used as drug targets.
3. Metals in Biology

Metals could be responsible for the origin of life. Their role in early evolution is highly debated in the scientific community, but there is no question regarding the importance of metal ions in the biochemistry of living organisms. Due to its unique properties a metal ion can both accept and donate electrons, therefore metals play a central part in many challenging chemical reactions and in energy metabolism [10, 11].

Cells acquire a number of metal ions, that are structurally constrained by ligands and ultimately used in catalysis and regulation. Metals such as iron, copper, zinc, and manganese are thereby essential nutrients in all living organisms. The metal ions are often referred to as trace elements, but studies of individual cells reveal that inside cells they are concentrated by several orders of magnitude compared to the concentration in a regular growth medium [12]. For instance, *Escherichia coli* cells contain iron and zinc at a total concentration of 0.1 mM, or $\sim 2 \times 10^5$ atoms per cell [12]. Iron is found in all types of cells and is required in many proteins and enzymes such as oxidases and oxygenases [13]. In cells free metal ions are kept at low concentrations and are released under tight control by internal physiological metal–buffering systems [14]. They are protected and guided throughout the cell by metallochaperones [12].

3.1 Metalloproteins

Metals are used as inorganic cofactors in proteins. A protein with a bound metal ion is denoted metalloprotein. It is estimated that 30–50% of all proteins require metal ions for their function [15, 16]. Metals are important in biological chemistry as they are responsible for structural features in folded proteins, participate in catalysis by orienting substrates for reactions, functioning as Lewis acids, and as mediators in oxidation–reduction reactions. There are several signaling pathways within the cell that require metal ions and consequently these are handled by metalloproteins. Many proteins coordinate the metal ions directly through charged side chains, others bind metal ions indirectly through small organic compounds, for example via protoporphyrins such as heme or the porphyrin like corrin, which is found in vitamin B$_{12}$. 
Before entering a cell or a cellular compartment, metals have to pass through different selectivity filters, very specific ion transport channels. The metal concentration inside the cells is tightly regulated, and many organisms keep the cellular level of the most common metal ions in the range of micro- to millimolar concentrations [12, 14]. Metal incorporation is proposed to take place into the already folded protein, delivered by a metal insertase or chaperone [15, 17]. However, it is not likely that there is a specific metal transfer protein for every single metalloprotein. The specificity of metal insertion will be influenced by the atomic radius of the metal, coordinating ligands, and competing proteins. Compartmentalization may also affect the availability of different metal along with changes of oxidation states, due to reducing environment in the cytosol as compared to the oxidizing environment of the periplasm [15]. Most metal centers are deeply buried in the proteins, disturbing the otherwise hydrophobic core, and it is likely that the apo–protein (without any metals bound) will be partly destabilized, as the coordinating ligands are usually charged residues. Also, metal ions are associated with specific redox potentials, and the enzymatic function is dependent on insertion of the correct metal. In a given protein the metal specificity is generally high, but many questions still remain on how the correct metal ion is incorporated.

3.1.1 Iron dependent proteins

Iron is one of the most abundant and versatile metals found in proteins [15]. It may assemble into hemes, iron–sulphur clusters or none–heme mono– or dinuclear cofactors. Heme–containing proteins are responsible for the storage and transport of oxygen, and others such as the cytochromes play an important role in oxidative metabolism, transferring energy within the mitochondria. Non–heme diiron–carboxylate proteins, present in eukaryotes and bacteria, are capable of an impressive array of redox transformations [18]. Some of the most challenging reactions are performed by diiron–carboxylate proteins belonging to the ferritin-like superfamily. These include the radical-forming ribonucleotide reductases (RNRs) [19, 20], plant soluble \( \Delta^9 \) stearoyl-acyl carrier protein desaturase [21] which is responsible for fatty acid desaturation, the carbon fixing monooxygenases that convert methane to methanol [22, 23], and the iron storage proteins e.g. ferritins [24]. The diiron–carboxylate group of proteins continues to grow with the increasing number of genomes sequenced. The bacterial multicomponent monooxygenases (BMMs) and RNR R2 proteins perform 2 or 1 electron oxidations using their diiron–carboxylate cofactor. Ribonucleotide reductase R2 proteins use their metal site to generate a stable radical on a nearby tyrosyl residue, which
is essential for the reduction of nucleotides (NTP) to their deoxy (dNTP) forms. This family of proteins will be discussed in more detail in section 3.4.

Iron is mainly present in two different oxidation states, Fe$^{II}$ (ferrous iron) and Fe$^{III}$ (ferric iron). However, transient high–valent Fe$^{IV-V}$ states have been observed in biological systems, identified as active intermediates for a number of enzymatic systems [25, 26]. The highest observed oxidative state of iron, the ferrate ion (Fe$^{VI}$), has so far only been observed in non-biological systems [25]. The high–valent ferrate state may potentially exist in biological systems as short–lived intermediates, but there is no evidence of such existence to date. Ferrous iron is relatively loosely bound in proteins, but is readily oxidized upon coordination and bound more stably as Fe$^{III}$ [15].

3.1.2 Manganese dependent proteins

In contrast to iron, manganese is rarely found in redox enzymes and also much less abundant in cells, commonly kept at a tenth of the concentration of iron [12]. Superoxide dismutase (SOD) is an essential antioxidant enzyme that utilizes its metal cofactor for redox reactions [27]. SOD acts as front–line defense in protecting organisms from reactive oxygen species (ROS), by conversion of superoxide radical anions into hydrogen peroxide [27, 28]. Manganese superoxide dismutase contains one manganese ion per monomer, but it seems only modestly selective for manganese and may also bind iron [27]. Another family of antioxidant enzymes is constituted of the catalases that utilize their metal cofactor to catalyze the conversion of hydrogen peroxide into water and molecular oxygen. Catalases act to neutralize the hydrogen peroxide from e.g. the SOD pathway, and are likewise present in all organisms with an oxidative metabolism. There are different groups of catalases and among them a group of non–heme Mn–catalases, which are found in several bacterial species [29].

Another abundant complex that utilizes manganese is Photosystem II (PSII), a thylakoid membrane protein complex found in plants. The catalytic center contains a Mn$_4$Ca–cluster, and it utilizes light to oxidize the metals to produce molecular oxygen from water. In 2011 a high resolution structure of the PSII complex was published, providing necessary details of the catalytic site to enable mechanistic interpretations [30].

A few years ago a subclass of the dNTP synthesizing ribonucleotide reductases was found to utilize a dinuclear manganese cofactor, in place of the earlier identified diiron cofactor [31, 32, 33].

Manganese can reach the same oxidation states as iron, Mn$^{II}$–being the most stable oxidation state. It is proposed that manganese has higher
affinity for carboxylate coordinations than iron [15]. However, metal binding in proteins is highly complex and the arrangement or affinity of metals is not easily deduced.

### 3.1.3 Iron and Manganese dependent proteins

There are many examples of homodinuclear proteins using diiron or dimanganese metal centers. The two metals possess very similar charge properties and ionic radii [34]. Consequently, they appear to be coordinated in a similar way by the same type of ligands, histidines and carboxyl residues (asparates and glutamates). Despite their similarities it appears that iron and manganese are specifically used for different cellular functions. Recently a couple of studies have provided insight in metal diversity, where the same enzymatic system may use different cofactors dependent on the species in which the system is present [19, 27].

The existence of a heterodinuclear iron–manganese site was proven by spectroscopic methods for two enzymatic systems; purple acid phosphatase from sweet potato [35] and the ribonucleotide reductase from *Chlamydia trachomatis* [36, 37] (presented in more detail under section 3.4.1). A third enzyme, the amine oxygenase AurF, has been suggested to contain a heterodinuclear Mn/Fe cofactor, but there is contradictory data supporting a diiron active site [38].

The Fe$^{III}$Mn$^{II}$ centre of purple acid phosphatase was presented in 2001 by Schenk et al.[35]. The metal ion content was measured by inductively coupled plasma mass spectrometry (ICP–MS) and the catalytically active enzyme was reported to contain a heterodinuclear metal center [35]. The structure of the enzyme was presented a couple of years later, providing details of the metal centre [39]. However, the architecture of the metal site is based only on calculated thermal motion of the individual atoms, and the coordination chemistry, but the exact arrangement has not been properly assessed.

### 3.2 Challenges of the field

When working with metalloproteins there are some essential facts to keep in mind. The same type of metal site may bind different metals depending on organism, presence of metal in cellular sub–compartments, or uptake from its surrounding environment. A native metal cofactor can be defined as the metal ion that will provide properties essential for the main function of a protein *in vivo* [4]. Therefore, metalloproteins must be properly assessed when studied outside their natural system, as is the case in heterologous expression systems when proteins are taken
from their natural environment and internal buffering systems. When protein is produced in common expression hosts, the cells are usually grown in enriched medium sometimes with supplements of metals, an environment that is different than that of the native host. Enzymes are often tested for activity, but the native cofactor would not necessarily provide the highest specific activity. It has been demonstrated in several cases that non–native metals generate enzymatically active systems, sometimes with even higher activity than of the native metal cofactor [14, 40].

It is also unclear how the native metal cofactor is incorporated into the protein in vivo. Metalloproteins may in vitro show high specificity for the appropriate metals that are acquired via diffusion, collision and substitutions, but what happens in the cytosol is more complex. Many metalloproteins are charged with the metal(s) by a chaperon [15], a protein that has a certain metal specificity and that recognizes the target protein(s) to deliver its cargo. If every metalloprotein were to acquire its metals from chaperons, each cell would then need a large array of chaperons to efficiently deliver the metals. A protein may find the right metal(s) by simple collisions, inside the cytosol even at very low concentrations, due to the relatively rapid diffusion of metal ions and attraction by chemical specificity to its ligands. However, "free" metal ions would potentially cause a problem as they can be toxic to the cell. The intracellular concentration of free copper for example, is estimated at only one free copper ion per cell [41].

As stated above, metal incorporation seems to be promiscuous in certain cases. One way of rationalizing metal binding is based on the empirically derived hard and soft acid–base (HSAB) principle. Provided that certain ions act as hard and others as soft Lewis acids or bases, chemical elements prefer to pair up with the corresponding ligands. Metals are classified as soft or hard acids according to their electronegative power and secondly their radii [34]. Importantly, this principle is influenced by the oxidation state; the higher the oxidation state the harder the acid. Thus the oxidation state will affect the coordination and binding strength to the protein. For example, it has been shown that divalent manganese has a higher affinity for carboxylate–coordinating ligands than iron, due to being a "harder" transition metal [15].

The answer to whether a protein has high metal specificity or if it is promiscuous may be found in its metal coordination, but the underlying principles are poorly understood.
3.3 Studying metalloproteins

X-ray crystallography reveals protein structures at atomic resolution, however it does not provide any information on the oxidation state of bound metal ions. Several powerful methods are available that can provide information on identifying particular metal ions and their oxidation states within any given protein. These methods utilize the spectroscopic or anomalous properties of metals. A commonly used spectroscopic technique is Electron Paramagnetic Resonance (EPR). This method generates unique spectra for unpaired electrons of different transition metals, which appear from certain metals at different oxidation states. Another spectroscopic method, Mössbauer spectroscopy, uses gamma rays and their absorption effects from gamma–active isotopes. Surrounding ligands will cause measurable affects on the absorption and provide spectra unique to the protein environment. There are also several of X-ray based methods for studying metals in proteins, namely Extended X–Ray Absorption Fine Structure (EXAFS) and anomalous dispersion. In EXAFS monochromatic X–rays are used to provide information on oxidation state and coordination of the metal centre. In this work we have used a crystallographic technique called anomalous dispersion for determination of the precise positions of metals in two different enzymatic systems. The method takes advantage of the ion's ability to absorb energy and alter the dispersed X–rays. It will be described in more detail below.

3.3.1 Anomalous diffraction experiments

The outcome of a crystallographic experiment is a diffraction pattern. The images produced are then used to calculate the electron density of the crystal. The quality and accuracy of the retrieved information is dependent on crystal quality and the resolution gained. The electron density can be represented by a Fourier sum, and the equation used to describe the density $\rho(x,y,z)$ from the diffraction pattern is:

$$\rho(x,y,z) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| e^{i\alpha(hkl)} e^{-2\pi i(hx+ky+lz)}$$

This equation tells us that each atom from the unit cell contributes to every reflection of the diffraction pattern, but from a regular diffraction pattern we will only obtain information on the intensities $I(hkl)$. These are in turn proportional to the structure factor amplitude $F(hkl)$ of the reflections $(hkl)$. However, diffraction patterns contain no information of the phase $\alpha(hkl)$ of the wave function. The "phase–problem" can be solved by phases retrieved from a homologous structure, using a method called molecular replacement, or through the use of anomalous scatterers.
Figure 3.1: Argand diagram of the structure factors from an anomalous scatterer. (a) Showing the relationship among the vectors $F_{PH}$, $F_P$, and $F_H$. Where $F$ represents the structure factor, $P$ indicates the overall protein contribution, and $H$ the heavy atom contributions. The length of the vector represents the amplitude of the structure factor and the angle ($\varphi$) represents the phase. (b) Summary diagram, showing the resulting scattering terms when Friedel’s law is broken.

Each structure factor, or rather each term in the fourier sum is a complex number and can therefore also be represented by vectors in an Argand diagram see figure 3.1a. According to Friedel’s law all structure factors will have symmetry elements that have the same amplitude $|F(hkl)| = |F(-h-k-l)|$ but opposite phases $\varphi(hkl) = -\varphi(-h-k-l)$, these are called Friedel pairs. Therefore $F^+$ will have a mirror image in $F^-$. If the energy of an incident X–ray wave is close to the absorption edge of an element this will cause an electron to enter a higher shell. When this electron returns to a lower shell it will emit radiation and cause retardation of the wave. This phenomenon will add an anomalous component to the structure factor, and as a consequence Friedel’s law will no longer be valid. This gives measurable effects in intensity differences of the symmetry–related reflections. Absorption occurs at specific X–ray wavelengths for each element. Every element has an absorption edge, only most of the light atoms have edges too far away for the X–rays of today’s modern synchrotrons. For example common substituents of proteins such as carbon, hydrogen, oxygen and nitrogen will not contribute to the anomalous scattering. However, metal ions will add an extra term to the structure factor, described in figure 3.1. The two contributions are perpendicular and appear as one real and one imaginary part. The difference between the structure factors can then be used to calculate the phases, but it can also be used to calculate anomalous difference maps for the individual metal ions, a method we used in Paper I and II.
As seen in figure 3.2 there is a clear difference in the anomalous properties of manganese and iron, which makes it possible to differentiate between these metals. This is important when trying to elucidate the arrangement, specificity and nature of a metallocofactor.

![Figure 3.2: Edgeplot over the anomalous scattering coefficients $f'$ and $f''$. Anomalous scattering terms are plotted for: manganese (purple), and iron (green). Theoretical X-ray absorption edges were calculated using the skuld-webserver [42].](image)

### 3.4 Ribonucleotide reductase

In the early 60’s Reichard and colleagues discovered how cells synthesize deoxyribonucleotides, the building blocks of DNA that are essential to all living organisms [43, 44, 45]. They found that the enzyme ribonucleotide reductase (RNR) is responsible for the conversion of ribonucleotides, RNA building blocks, into the corresponding deoxynucleoside 5’-di– or triphosphates (dNDPs or dNTPs). This enzyme plays a central role in nucleotide metabolism, being responsible for de novo synthesis of all four DNA precursors. It also controls the intracellular concentrations and relative ratios of dNTPs. The correct concentration of each dNTP is kept and regulated by an interplay of several mechanisms: gene transcription, protein degradation, allosteric regulation of the catalytic subunit, and specific inhibitors.

The reduction of NTPs into dNTPs is chemically a very challenging reaction. We now know that the system is capable of performing
one- or two-electron oxidations to produce a stable radical [46, 47]. The substrate-binding catalytic subunit of all RNRS generates a transient thyl radical ($S^\cdot$) in the active site. The different classes of RNRS differ in how they generate this $S^\cdot$, discussed in more detail below. The catalytic protein is also controlled by allosteric regulation, affecting substrate specificity along with catalytic rate.

Due to its importance in proliferating cells the RNRS are potential drug targets, in the fight against pathogenic bacteria, as antiviral agents and for treatment of cancer.

### 3.4.1 Classes of ribonucleotide reductases

All RNRS are believed to have the same evolutionary origin, but are divided into different classes based on oxygen dependency, sequence similarity and the cofactor utilized to generate the $S^\cdot$ [15, 46, 47, 48]. In spite of the overall low sequence identity, all three classes contain common structural elements that allows for allosteric mechanisms, such as regulation of substrate specificity.

<table>
<thead>
<tr>
<th>Class I</th>
<th>Class II</th>
<th>Class III</th>
</tr>
</thead>
<tbody>
<tr>
<td>O$_2$-dependency</td>
<td>aerobic</td>
<td>O$_2$-independent</td>
</tr>
<tr>
<td>Subunit structure</td>
<td>$\alpha_2\beta_2$</td>
<td>$\alpha$ or $\alpha_2$</td>
</tr>
<tr>
<td>Cofactor</td>
<td>Dimetal</td>
<td>AdoCbl$^a$</td>
</tr>
<tr>
<td>Primary Radical</td>
<td>Tyrosyl/metal centered</td>
<td>AdoCbl$^a$</td>
</tr>
<tr>
<td>Occurrence</td>
<td>Eukaryotes, Bacteria</td>
<td>Archaea, Bacteria</td>
</tr>
</tbody>
</table>

$a)$ adenosylcobalamine, $b)$ adenosylmethionine.

Class I RNRS generate a stable tyrosyl radical ($Y^\cdot$) using a dinuclear metal cofactor in an accessory subunit to the catalytic subunit. Class II RNRS utilize electrons from thioredoxine, and generate the $S^\cdot$ radical with an adenosylcobalamine (AdoCbl) cofactor directly bound in the catalytic subunit [49]. Both of these classes operate in the presence of oxygen, class I being dependent on oxygen for generation of its $Y^\cdot$, and class II having two independent pathways for the AdoCbl synthesis of which one is oxygen dependent [50]. Class III RNRS are anaerobic and use an iron-sulfur [4Fe4S] cluster to cleave $\text{S-adenosylmethionine}$ (AdoMet), which in turn will generate a glycyl radical [49, 51]. In class III enzymes, the radical is generated by a separate protein, belonging to a group of radical SAM–enzymes. Many organisms possess multiple RNRS, with expression dependent on growth conditions [49].
Class I RNRs

Ribonucleotide reductases belonging to class I have two homodimeric subunits. The larger α (R1) subunit is responsible for the catalytic activity, containing substrate and regulatory binding sites, whereas the smaller β (R2) subunit generates the necessary radical. The R2 subunit has a conserved structural fold and they all contain a dinuclear metal cluster. The first structure of the radical generating subunit from *Escherichia coli* was reported in 1990, providing structural details of its di-iron centre [19]. It showed an α–helical structure with a novel eight-helix bundle conformation (figure 3.3a). This fold and dimerization is characteristic for the family of R2 proteins. A four helical bundle provides the ligands for the metal coordination, which consists of three glutamates, one aspartate and two histidines. The tyrosine residue that harbors the radical, sits in close proximity to the metal site. The radical is thereby stabilized by the protein and metal centre (figure 3.3b).

![Figure 3.3: The ribonucleotide reductase R2-subunit from *E. coli*. (a) The radical generating subunit from class I RNR in *E.coli*, one monomer is colored in light blue and the second monomer in rainbow, tracing from the blue N- to C-terminal in red. The monomers each contain two iron atoms, visible as orange spheres. (b) The di–metal centre and its radical harboring tyrosine (Y122). The figures were generated by PyMol [52] using PDBid: 1RIB.](image)

The class I RNRs are divided into three subclasses, Ia, Ib, and Ic, which are structurally homologous. The different subunits are encoded by the *nrd* genes. *NrdA* and *nrdB* encode the class Ia R1 and R2 proteins, respectively, and class Ib is encoded by *nrdE* and *nrdF* encoding the R1E and R2F proteins. Class Ib R2 proteins have identical coordination sites to those of class Ia, and they were long believed to contain the same set of metals. However, it was recently shown that this group
of proteins contains a di-manganese cofactor in place of the classical di-iron site [31, 32, 33]. Additionally, the class Ib RNRs require two more enzymes, one for reduction of the metallocofactor to enable cycling of the radical generation, by the protein NrdH encoded by \textit{nrdH} [15, 48]. The second enzyme is a flavodoxin-like protein encoded by \textit{nrdI}, which is important for the radical generation and is involved in the oxidation of the manganese metal centre [15, 53]. Furthermore, a new subclass Ic has been identified [54], but this group will be described in more detail in the next section.

The reduced metal cofactor of class Ia is oxidized, by cleavage of molecular oxygen, and in turn the hydroxyl group of the nearby tyrosine to a stable radical. This tyrosine is essential for activity and thus conserved throughout the R2 proteins of class Ia and Ib [55]. The mechanisms of radical generation is still being explored. The radical is delivered to the active site in R1 via a $>35$ Å proton-coupled electron transfer (PCET) pathway. The residues of the pathway are highly conserved throughout the groups of class I RNR [56].

\section*{The class Ic RNR}

The RNR subclass Ic was first identified by Högbom and colleagues in \textit{C. trachomatis} [54]. There are several chlamydial species, \textit{C. trachomatis} being one of the most common sexually transmitted pathogens, is an intracellular parasite. As an intracellular pathogen, \textit{Chlamydia} are able to utilize the host cell machinery for some of its biosynthetic pathways, a strategy that has yielded an unusually small genome size [1]. It appears that the bacteria obtain ribonucleotides from the host cell, but no transport proteins that are specific for deoxyribonucleotides have been identified so far [57]. Subsequent synthesis of deoxyribonucleotides is performed by ribonucleotide reductases, specific to \textit{C. trachomatis} [58, 59].

The genome of \textit{C. trachomatis} contains no other RNR gene beside a class I variant with a sequence similarity of around 25\% to class Ia RNRs [58]. Most of the residues in the active site are conserved when compared to the other subclasses. However, there are few crucial exceptions. One of the residues that are not conserved, but of great importance for enzymatic activity, is one of the carboxylate ligands, being a glutamate instead of an aspartate. Furthermore and importantly the otherwise highly conserved radical harboring tyrosine is replaced by a phenylalanine [58]. A substitution of this tyrosine yields an inactive enzyme in other R2s [60, 55]. The recombinantly produced \textit{chlamydia} protein however is active \textit{in vitro} and sensitive to hydroxyurea [58]. Hydroxyurea inhibits RNR activity by quenching the radical [48, 61]. A tyrosine is found only two residues downstream from the phenylalanine and it was long believed to be harboring the radical. However, when the crystal structure of the \textit{Chlamydia} R2 was solved a few years later, it showed that the pheny-
Figure 3.4: **Superposition of R2 protein metal sites.** Structures of the oxidized metallocofactors of the class Ia (cyan) and Ic (pink) ribonucleotide reductases. Metals are shown as gray spheres, and solvent molecules from the oxidized structure of class Ic in red spheres. Important differences between the two structures are the depicted phenylalanine (F127) and glutamic acid (E89), with *Chlamydial* numbering. The figure was made using PDB files: 1MXR *E. coli* and 1SYY *C. trachomatis*.

L-alanine is indeed replacing the radical harboring tyrosine [54, 62] (figure 3.4). It was proposed that the diiron center is oxidized to a radical equivalent Fe$^{IV}$/Fe$^{III}$ state [54, 63], in place of the tyrosyl radical. Later, activity measurements also suggested that the activity did not correlate directly with the iron–content [37]. When the enzyme instead was reconstituted with equal amounts of iron and manganese activity increased 50–fold and maximum activity was observed for a ratio of 1:1 of the two metals [37]. It was shown that the most active cofactor actually consists of a manganese and an iron ion [36, 37]. The heterodinuclear cofactor was shown to be oxidized to Mn$^{IV}$/Fe$^{III}$ [37, 64, 65] before releasing the radical to the R1 subunit.

The *Chlamydial* class Ic protein represents the first RNR protein with a Mn/Fe cofactor and the first heterodinuclear redox cofactor within the ferritin–like superfamily. To better understand the chemistry of this cofactor, structural information is needed. Provided that several enzymes seem to be promiscuous when it comes to their manganese incorporation [27], the question arises how the heterodinuclear site is arranged. Structural details of the metal site would provide the field with new insights in radical generation. Therefore we set out to investigate the metal specificity of the two metal sites.
3.4.2 Architecture of the Mn/Fe cofactor in *Chlamydia trachomatis* (Paper II)

Recently, a newly identified subclass of RNR R2, R2c, was shown to have a Mn/Fe heterodinuclear cofactor, as described above. With this study, we have been able to provide structural information on the architecture of the heterodinuclear metal site. Thus, our results have provided the field with new insights on a Mn/Fe heterodinuclear cofactor used in radical generation.

We set out to answer the question whether the metal positioning is specific, and if so, in what position the individual manganese or iron ion is located? We used X-ray crystallography utilizing anomalous dispersion to show the metal arrangement. We were able to provide data supporting a specific arrangement of this cofactor. We found that the manganese ion occupied only metal position 1, and that iron is mainly found in position 2 (figure 3.5). By calculating double–difference maps of the protein, we were also able to assess the contribution of iron in position 1. The absence of any other anomalous scatterers, than manganese and iron, was verified by data collected at the low energy side of manganese.

The study provided insights on the metal architecture of the *Chlamydial* R2c, along with important insights on a novel type of heterodinuclear metal cofactor in R2 proteins. The work is of great importance e.g. theoretical chemists as a correct model system, that in turn will provide us with further understanding of radical chemistry and oxygen activation of a heterodinuclear site.

![Anomalous difference maps of manganese and iron](image)

*Figure 3.5: Anomalous difference maps of manganese and iron* (a) The anomalous difference map at $\lambda=1.85$ Å, displaying the anomalous scattering of only manganese. (b) Anomalous difference map at $\lambda=1.7$ Å, where both manganese and iron contribute to the anomalous scattering.
3.5 R2 – like proteins

The ribonucleotide reductase R2 proteins are dinuclear metalloproteins belonging to the ferritin–like superfamily. In Paper I we reported on an R2 homologue from Mycobacterium tuberculosis, which is one of the ten most upregulated genes in the pathogenic strain compared to the avirulent bacillus Calmette-Guérin (BCG) vaccine strain [66]. Initial sequence alignments suggested it belonged to the RNR subclass of Chlamydial R2c, figure 3.6. However, the protein showed no RNR activity. More importantly, the crystal structure showed a non–protein ligand directly coordinated to the cofactor, projecting away from the metal site through a hydrophobic tunnel. The presence of the ligand combined with a, previously undescribed, covalent valine–to–tyrosine crosslink suggested that the protein may perform two–electron oxidations and potentially be involved in fatty acid metabolism. The enzyme was proposed to be an ligand oxidase, and therefore named R2lox.

Figure 3.6: Phylogenetic tree of R2 homologues. The phylogenetic tree shows R2 homologues where the canonical radical harboring tyrosine is replaced by phenylalanine. The tree can be divided into two regions (highlighted in orange and purple), where the locations of C. trachomatis R2c and M. tuberculosis Rv0233 are indicated.

Recently, another group of R2–like proteins was also discovered in eubacteria, the aldehyde decarbonylases, that convert saturated and unsaturated fatty aldehydes to alkanes and alkenes, respectively [67]. These enzymes show a high structural similarity to the class I RNR R2 proteins, and they contain the same conserved four–helical bundle architecture seen in the ferritin–like superfamily. The first aldehyde decarbonylase structure from Prochlorococcus marinus was solved by the Joint Center of Structural Genomics (pdb-id: 2oc5). It shows a high structural similarity to the M. tuberculosis protein [68]. The aldehyde decarbonylases are smaller than regular R2s, containing only around 250 residues as compared to 300–400 residues in normal R2s [67], but the core of the protein is conserved. The overall chemical reaction is formally neutral, but it has been proposed that the enzymes utilize peroxide to perform alkane or alkene synthesis [69]. Furthermore, they employ a reduction system sim-
ilar to the ones used in other O$_2$–dependent diiron enzymes [67, 69]. The biochemical and structural observations made on this group of proteins suggest that they are involved in fatty acid metabolism, possibly using the same type of heterodinuclear redox cofactor as the *M. tuberculosis* R2–like protein. However, the metal cofactor of the decarbonylases has not been conclusively identified.

3.5.1 An R2–like ligand oxidase from *Mycobacterium tuberculosis* (Paper I)

While the ribonucleotide reductase R2 proteins are known to generate radicals necessary for subsequent reduction of ribonucleotides, the R2c–like protein Rv0233 from *M. tuberculosis* showed no RNR activity when tested with the catalytic R1 subunit. This R2 homologue is up–regulated about 7–fold in the pathogenic H37Rv *M. tuberculosis* strain compared to the BCG vaccine strain and therefore a possible virulence determinant [66]. Identification of regulatory differences between the attenuated and pathogenic strains, are of great interest from a drug development perspective and possibly for the characterization of novel metabolic pathways.

In this paper we provided structural information on the Rv0233 and details of its metallocofactor. The protein structure revealed a new type of heterodinuclear manganese–iron cofactor, earlier proposed for the class Ic RNR [36, 37], but which had not been structurally characterized previously. With this work, we were able to depict the architecture of the metal site and position the two metal ions. The protein cofactor supports potent 2–electron oxidations, as demonstrated by a novel tyrosine–valine crosslink observed in close vicinity of the active site (figure 3.7a). In 2011, there was a report of a crystal structure of another protein (symerythrin) containing a similar crosslink, valine–to–phenylalanine, which was implicated in stabilization of the enzyme’s diiron center [70].

The protein structure of Rv0233 also revealed a bound ligand, found in a continuous cavity protruding from the metal site towards the protein surface (figure 3.7b). This large cavity shows similarities to those found in the bacterial multicomponent monooxygenases [71]. The ligand was modeled as a myristic acid, but the true nature of the molecule is yet unknown, though the ligand implicates that the enzyme may be a new type of oxidase, i.e. R2–like ligand oxidase (R2lox). Continued studies of this new group of R2lox enzymes are aimed at defining its enzymatic function.
3.5.2 A dynamic C–terminal helix of R2lox shields a positively charged patch on the surface (Paper III)

In our first crystal structure of the *M. tuberculosis* R2–like ligand binding oxidase (*Mt*R2lox), presented in Paper I, the structure had a disordered C–terminus. Likewise, all currently available structures of R2 homologues also contain a disordered C–terminal region. In an attempt to find new crystallization conditions we identified two new space groups with different crystal packing. From these new crystal forms we were able to trap a structure of the *Mt*R2lox with an ordered C–terminus.

The overall surface potential of the *Mt*R2lox protein revealed a pronounced concave positively charged patch on the surface (figure 3.8a). In the structure with an ordered C–terminus, the C–terminus forms a helix that folds over the patch and shields the positive charge (3.8b and c). Based on sequence analysis, we found that the positively charged patch is likely conserved within the R2lox family. While the C–terminal helix only appears to be conserved among the *Mycobacterium* genus and a few strains of *Rhodococcus* species. Other organisms contain a conserved C–terminal segment of varying sequence and length.

Given the ability of altering between an ordered and disordered state, we hypothesize that the dynamic C–terminal segment may be of regulatory importance.

*Figure 3.7: R2lox metal site.* (a) The crosslink between Valine 71 and Tyrosine 162. Iron and manganese are colored in orange and purple, respectively. The green mesh shows the calculated OMIT map for the two residues. (b) The ligand is modeled as a myristic acid in the calculated OMIT map.
Figure 3.8: Model of the *M. tuberculosis* R2lox monomer Overall surface contact potential, red and blue indicate negative and positive potential, respectively. (a) Shows the surface contact potential of the previously published structure (3EE4). With a disordered C-terminus the protein display a positive patch. (b) Visualization of the C-terminal helix position. (c) The surface contact potential calculated for the structure with the ordered C-terminal helix.
4. Tuberculosis

Pathogenic mycobacteria constitute a large group of microorganisms responsible for bacterial infections. There are several strains of pathogenic environmental mycobacteria that can be found in water supplies. They mainly lead to disease in immunosuppressed people, especially those also infected with HIV [72]. The microbial pathogen with the highest disease burden is \textit{M. tuberculosis}, the causative agent of the disease tuberculosis (TB). In 2010 TB was responsible for 1.4 million deaths [5]. Despite the existence of antibiotics and an attenuated vaccine, the tubercle bacilli is still among the worst global killers known, and the World Health Organization (WHO) estimates that about one third of the world’s population is infected with tuberculosis. Mycobacterial infections are increasing with a growing population, and it is therefore important to find ways to treat the subsequent disease.

Tuberculosis is an airborne, chronic and infectious disease that typically affects the lungs, but it can also reach other parts of the body. The bacilli enter the human alveolar passages where they are ingested, primarily by macrophages. The bacteria are able to multiply within the macrophage, but this is not completely understood. Inside the host, the bacteria experience an environment of low pH, ROS, low availability of oxygen, presence of toxic fatty acids and peptides [8]. This enclosed infection is termed as persistent or latent TB. Thus, persons with a good cell–mediated immunity are substantially more efficient in arresting the infection at this stage, compared to immunosuppressed patients. Generally, people infected with \textit{M. tuberculosis} have a relatively small risk of developing TB during their lifetime. Approximately 10% develop disease and the other 90% remain with an asymptotic infection [5]. The mortality rate among infected patients, testing positive in smear–tests, that do not receive any treatment is estimated to around 70% within 10 years [73]. However, the probability of developing the disease is much higher in immunosuppressed patients, and TB is the leading cause of death in people with HIV [5]. Uncontrolled growth of the bacterium will ultimately lead to extensive tissue damage and cause death by suffocation due to an insufficient supply of oxygen.
4.1 M. tuberculosis

The tubercle bacillus was first identified and described by Robert Koch in 1882 and was primarily characterized by its slow growth and complex cell wall. It is a gram positive bacterium with a cell wall exceptionally rich in unique lipids, glycolipids and polysaccharides. The bacterium can adapt to long periods of dormancy in tissues [74], and the cell wall may be the explanation of this long–term latent and asymptotic infection. M. tuberculosis can be revived when a person’s immune system is weakened, and the bacteria then start to divide [8]. Treatment is complicated as antibiotics usually act against growing bacteria, targeting metabolic pathways. Within the dormant phase of the bacterium, it will not be susceptible to drugs, and thus treatment of infected patients is prolonged. This type of resistance is called phenotypic resistance [75].

Dormant M. tuberculosis is susceptible to isoniazid [76], a drug targeted against cell division. Hence, the bacterium must be able to divide even in its dormant phase. Biochemical evidence suggests that the tubercle bacillus can switch its lipid synthesis from carbohydrate–metabolites, in an anaerobic atmosphere, to utilizing lipids when under aerobic conditions [8].

The complete genome sequence and annotation has been made available for a number of M. tuberculosis strains, but importantly the most commonly used lab strain, H37Rv, was sequenced by Cole et al. in 1998 [77]. This work provided insights into specific features of the tubercle genome, such as its richness in guanosine and cytosine. The genetic homogeneity results in a bias of the amino acid content of the translated proteins. The greatest difference to other bacteria is that a large portion of the TB genome is devoted to production of proteins involved in lipid metabolism, which is therefore one of the most elaborate lipid metabolisms known. It is likely that M. tuberculosis uses unique biosynthetic pathways to generate cell–wall components such as mycolic acids, lipids and araginogalactan.

4.1.1 Drug Resistance

The introduction of rifampicin in 1963 was the last time a completely new drug specifically targeted against tuberculosis was introduced [78]. Now, 50 years later, there is an alarming need of new drugs against TB. Over the years there has been a rapid development of drug–resistance among the identified TB cases, and multi drug–resistant (MDR), extensively drug–resistant (XDR), and totally drug–resistant (TDR) [79] strains are increasing all over the globe. MDR–TB is caused by strains of M. tuberculosis resistant to at least isoniazid and rifampicin (two of the so called first–line drugs). XDR bacilli are additionally resistant to in-
jectable, second–line drugs (amikacin, capreomycin or kanamycin) and to
fluoroquinolones. Out of 12 million prevalent TB cases WHO estimates
that 650 000 cases are MDR–TB [5]. The estimated average (median)
cost per treated patient implied by reported data from 27 high MDR–
TB burden countries is US$ 8.200 [5]. TDR–TB cannot be treated with
any of today’s commercially available drugs. Treatment of drug–resistant
strains is more difficult, and requires substances that are expensive, toxic
and less effective, thus leading to a prolongation of the treatment along
with increased costs and side effects. Therefore there is a great demand
for new drugs that are efficacious and allow for significantly shorter treat-
ment courses.

4.1.2 Virulence Determinants

Bacterial pathogens have unique biological properties that enable them
to grow in the environment of the host, inhibit its defense mechanisms
and damage the host, thus causing disease [80]. Typical determinants
of pathogenicity are toxins, poisonous substances produced and secreted
by microorganisms. Known pathogens whose toxins cause severe effects
in humans are for example *Chlostridium botulinum* which produces bo-
tulinum neurotoxin that causes botulism, *Vibrio cholerae* whose cholera
toxin is responsible for diarrhea, and *Corynebacterium diphtheriae* whose
toxin causes diphtheria by inhibition of protein synthesis [80, 81]. How-
ever, *M. tuberculosis* does not possess such classical virulence determinants.
Instead, its virulence determinants are rather described by factors that
are important for the progression of the disease [8].

To better understand the pathogenesis of the tubercle bacillus it is im-
portant to elucidate the factors associated with the disease. An important
step in characterizing the putative virulence determinants of tuberculo-
sis is to identify gene products unique to the pathogen, or regulatory
differences between the pathogenic and attenuated or non-pathogenic
mycobacterial strains.

Much work has been done in search of the virulence determinants of
*M. tuberculosis*, and a variety of targets has been identified. A significant
number of these targets are enzymes involved in the synthesis of the
mycobacterial cell wall and secreted proteins [8, 82, 83, 84].
4.2 Lipid Metabolism of *M. tuberculosis*

The lipid metabolism of *M. tuberculosis* is one of the most elaborate known, with about 250 enzymes putatively assigned to it from the genetic sequence [77, 85]. The large number of enzymes may be related to its unique cell envelope and its ability to grow intracellularly [74], where it has been suggested that *M. tuberculosis* uses fatty acids as its major carbon source [8]. The molecules produced in mycobacteria range from simple fatty acids, isoprenoids, to very long chain fatty acids. All mycobacterial strains synthesize a few types of unique fatty acids used in their outer cell wall, the mycolic acids [86].

The cell wall of *M. tuberculosis* is essential for the bacterium’s survival. This unique and complex cell envelope consists of an essentially impermeable layer of 60-90 carbon long mycolic acids that in turn are covalently attached to a peptidoglycan layer via the arabinogalactan network [87, 88, 89]. Mycobacteria utilize systems otherwise found in both eukaryotes and other bacterial species for fatty acid biosynthesis, e.g. the fatty acid synthase I and II [8, 86]. The type I fatty acid synthase system (FAS I) of mycobacteria is similar to the eukaryotic system, i.e. one polypeptide containing all functions required for *de novo* synthesis of C\textsubscript{24} and C\textsubscript{26} fatty acids [8, 90]. The *M. tuberculosis* type II FAS system is in turn analogous to the bacterial FAS II system, built up from four independent enzymes. The FAS II system does not perform *de novo* synthesis, but utilizes the coenzyme A (CoA) activated short-chain fatty acid precursors from FAS I to synthesize long and complex fatty acids and mycolates used in the outer cell wall [84, 90].

Due to the number of annotated genes and the complexity involved in the mycobacterial lipid metabolism, it is very likely that there are novel pathways yet to be identified.

4.2.1 Fatty acid activation

The mycobacterial genome contains an array of proteins involved in lipid metabolism. This includes a large family of annotated *fadD* genes [77]. The FadD proteins show high similarity to the fatty acyl-CoA synthetases (ACSs) known to be involved in the activation of fatty acids (FAs). Fatty acids must be activated before they are introduced into different metabolic pathways. The universal activation mechanism involves the conversion of an *n*-fatty acid into its corresponding *n*-fatty acyl-CoA, a reaction catalyzed in two steps by the ACSs (figure 4.1). The formation of acyl-CoA is a prerequisite for the biosynthesis of phospholipids and triacylglycerols [91, 92]. The *fadD* genes have been shown to be important for the mycobacterial biological activity by several inactivation experi-
ments [93], and their gene products have been suggested as potential drug targets in several human pathogens [94].

![Figure 4.1: Fatty acid activation by ACS proteins.](image)

This schematic illustration shows the general mechanism for fatty acid activation by ACS proteins. The binding of ATP induces a conformational change allowing entry of the hydrophobic substrate. Formation of an acyl-adenylate intermediate is followed by a 140° rotation of the smaller C-terminal domain, and enables binding of CoA for production of the final product, a fatty acyl-CoA thioester.

About twelve of the annotated FadDs use an alternate mechanism, only performing the first half of the reaction e.g. activation of long chain fatty acids into acyl-AMPs, before transferring them onto polyketide synthases (PKSs) [93, 94]. The combined action of fatty acyl-AMP ligases (AALs) and PKSs produces unusual fatty acyl chains that are used to build up unique mycobacterial lipids [93]. A small structural substitution has been shown to be responsible for turning an ACS into an AAL. The acyl-CoA synthetases and acyl-AMP ligases use the same fatty acid pools, but channel them to different fates, acting as node points for the lipidic metabolites at distinct stages of the mycobacterial life cycle. ACSs are responsible for degradation of carbon sources during the latent phase of the mycobacterial cycle, whereas AALs are essential when the bacteria are growing [93].

Homologues of ACSs are divided into subgroups dependent on their substrate size; short-, medium-, long- and very long-chain fatty acyl-CoA synthetases. The group of very-long-chain ACSs (ACSVL) has been mainly characterized in eukaryotes [95, 96]. Enzymes of this group appear to be integral membrane proteins [96], as compared to other ACSs that are soluble proteins with a substrate binding pocket matching the size of the preferred substrate [97]. The ACSVL enzymes have the capacity to accept and activate substrates longer than C_{22}. Activation of these long-chain fatty acids, in the cytosol, is shown to lead to a net influx of long chain fatty acids across the membrane [92, 98, 99].
4.2.2 The operon \textit{mymA} and the enzyme FadD13

One ACS enzyme of \textit{M. tuberculosis}, i.e. FadD13, was biochemically characterized and showed the highest activity for the longest substrate tested, C\textsubscript{26} \cite{100}. The gene encoding FadD13 is found in the operon \textit{mymA}. The operon was named after its first identified gene, Rv3083, which codes for a mycobacterial monoxygenase. When the bacillus is grown \textit{in vitro} under acidic conditions, mimicking the macrophage environment, the operon is induced 17–33 fold \cite{101}. Several other studies also imply that \textit{mymA} is important for mycobacterial survival within the macrophages. Knock–out or down–regulation of the operon leads to an altered composition of the mycolic acids in the cell wall, along with increased bacterial sensitivity to acid stress and detergents \cite{84, 101, 102}. Importantly, the operon is not essential for mycobacteria when grown in normal cell media, but it appears to be part of their survival strategy inside the macrophages. The signal for induction of the operon is not known, but transcription of the operon is regulated by the divergent \textit{virS} (Rv3082c) \cite{84}. This implies that \textit{virS} is possibly important for the up–regulation of the \textit{mymA} operon under acidic conditions. The putative enzymes of the \textit{mymA} operon are likely to be involved in fatty acid metabolism \cite{84}, see figure 4.2. Homologues of these genes are also found in \textit{M. leprae}, \textit{M. avium} and \textit{M. smegmatis}, although as individual genes and not in an operon arrangement \cite{102}. Interestingly, when \textit{M. tuberculosis} is growing at low pH, \textit{mymA} is up–regulated, and in turn the afore–mentioned FAS II system is down–regulated \cite{84}. The accumulated C\textsubscript{24} and C\textsubscript{26} fatty acids could be modified and directed by \textit{mymA} components into appropriate pathways for further modification, before being transferred to the mycobacterial cell envelope.

\textbf{Figure 4.2: Schematic representation of the \textit{mymA} operon.} The operon consists of seven open reading frames (ORFs), Rv3083 to Rv3089. All ORFs are oriented in the same direction with a maximum distance of 38 base pairs, and the total length is 8 kb. Annotations of the putative functional proteins are adapted from Tuberculist \cite{103}.

The last gene of the \textit{mymA} operon is Rv3089, encoding the FadD13 ACS homologue. FadD13 has been extensively characterized biochemically and tested for a variety of substrates lengths, showing the highest activity for the longest substrate tested, C\textsubscript{26} \cite{100}. Hence, this en-
zyme clearly belongs to the class of very long–chain fatty acyl–CoA synthetases. Previously identified ACSVL enzymes are found to be integral membrane proteins, directly associated with the plasma membrane, whilst this mycobacterial homologue was isolated from the soluble fraction after recombinant expression in *E. coli* BL21(λ DE3) [100]. Given the size of the protein the very long–chain fatty acids are expected to be longer than the binding cavity of the enzyme. In our work, we aimed to find an explanation on how the FadD13 enzyme may accept a substrate longer than itself.

4.3 A very long–chain fatty acyl–CoA synthetase from *Mycobacterium tuberculosis* is a peripheral membrane protein (Paper IV)

Unlike other ACSVL enzymes, the FadD13 protein sequence suggested a soluble protein, but with the knowledge of the FadD13 chemistry we hypothesized that the protein may not be able to accommodate the entire substrate inside the protein. This prompted us to investigate if FadD13 could possibly be a peripheral membrane protein.

To test our hypothesis we over expressed the protein in *E. coli*, isolated a membrane fraction from which we were able to acquire large amounts of the protein. The membranes were resuspended and extensively washed with buffer with different salt concentrations and pH. When membranes were washed with high salt concentration, phosphate buffer or at a high pH, the amount of isolated FadD13 was reduced by ∼50%. These results suggested a non–integral interaction between the enzyme and the plasma membrane.

The overall structure of *M. tuberculosis* FadD13 protein showed a two domain arrangement (figure 4.3a) similar to previous structures of soluble ACS enzymes [97, 104]. However, the structure revealed a large arginine and aromatic rich patch, on the otherwise negatively charged protein (figure 4.3b). The protein structure also revealed a hydrophobic tunnel protruding from the active site, ending by the positively charged patch (figure 4.3c). The channel is capped at the surface by an arginine–rich lid loop, shown in figure 4.3a and c. We proposed the patch as the site for membrane interaction. Our hypothesis was further tested by comparing the interaction between a model membrane and of wild–type protein as compared to a variant, which contained a number of alanine substitutions within the patch. Surface plasmon resonance (SPR) analysis was performed to monitor the binding of the protein to the membrane. We could show that both wild–type and the variant bound to the model membrane surface, and the increased number of hydrophobic residues led
Figure 4.3: Model of the FadD13 ACS homologue (a) Overall structure shows a conserved two-domain fold, the large N-terminal domain in pink and the smaller C-terminal domain in light green. Protruding the enzyme from the ATP binding site, we find a long hydrophobic tunnel seen in cyan. The tunnel is closed off by a lid-loop region marked in red. (b) Top view, showing the calculated surface contact potential. The partly basic/hydrophobic patch stands out in the otherwise acidic protein. (c) The arginine-rich lid loop capping the top of the hydrophobic channel at the protein surface.

to an increase in total binding to the surface. The fact that an altered patch lead to a stronger membrane binding supports its involvement in membrane interaction.

We propose, based on structural and biochemical information, that FadD13 utilize the membrane to partly accommodate lipids or fatty acids. The enzyme is proposed to be peripherally bound to the membrane through charged and moderately hydrophobic interactions, commonly found in interactions between protein and lipid head groups [105].
5. Concluding Remarks and Future Perspectives

Function of a protein is to a large extent determined by its three-dimensional structure. Small evolutionary changes of a protein structure can lead to an altered function, which seems to be the case for the R2lox proteins. The R2lox of *M. tuberculosis* was first identified as an R2 subunit of ribonucleotide reductase. However, we produced and analyzed the protein structure by using X-ray crystallography along with a biochemical activity assay and we were able to show that the protein is in fact an R2-like ligand oxidase. With this work, we were also able to identify and characterize a new type of heterodinuclear Mn/Fe cofactor of the R2lox protein. We also characterized this new type of cofactor in the functional R2 subunit of *C. trachomatis*. Thus determining the metal architecture of this heterodinuclear Mn/Fe metal cofactor in both protein groups.

The same type of Mn/Fe centre is active in both the R2lox and RNR R2c protein groups, but they utilize the cofactor for different kinds of chemistry. The cofactor has proven its potential of performing 1- and 2-electron oxidations, by being active in the R2c and from a covalent valine-to-tyrosine crosslink of the R2lox protein. The identification of R2lox and structural findings of a long hydrophobic substrate indicated that the R2lox protein is part of the *M. tuberculosis* lipid metabolism, and possibly a member of an undescribed metabolic pathway. This agrees well with the elaborate lipid metabolism found in the mycobacterial species. The *in vivo* function of the R2lox group is still unknown.

The metal cofactor needs further attention regarding its redox potential, coordination chemistry as well as the metal specificity of the active site. Clearly the redox potential is affected on the metal combination and from its coordination geometry, but question is what the reason is for utilizing this type heterodinuclear center? It would be interesting to elucidate what kind of chemistry the cofactor may perform. Diiron proteins have shown a great diversity and possibly we may discover that the Mn/Fe cofactor is present also in other important chemical pathways.

Both the two functional protein groups, R2lox and R2c, have unstructured C-terminal segments. These C-terminal segments are seemingly
important for the protein’s catalytic activity. Peptides mimicking the C-terminus of the R2 subunit have an inhibitory effect on the catalytic activity of the R1 subunit [106]. In R2lox we have shown that the helix is partly conserved within the family and that it may alter its conformation from a disordered to an ordered state. The helix in R2lox shields a positively charged patch that may be important for interactions with other cellular components, something that needs to be further studied.

Another example of the unusual lipid metabolism of *M. tuberculosis* is the FadD13 enzyme. This protein belongs to the family of very long-chain acyl-CoA synthetases (ACSVL), that are normally found to be integral membrane proteins (IMPs) in eukaryotes. The chemical function of this soluble protein was shown by Khare et al. [100] and their results indicated that the biochemical solution was different compared to the eukaryotic IMPs. Here we could show that mycobacteria produces a peripheral membrane protein utilizing the membrane to partly accommodate its very long substrates. Regarding the work done on FadD13 there is still a lot to do. It would be of great interest to be able to trap substrate, ligands and intermediates in the enzyme. Another interesting subject to pursue is how the substrate might affect the lipid–interactions. For this we have initiated mutational studies, to gain more data on the lipid–protein specificity.
6. Svensk populärvetenskaplig sammanfattning

Enligt världshälsoorganisationen (WHO) uppskattas en tredjedel av jordens befolkning vara infekterade av sjukdomen tuberkulos, som orsakas av tuberkelbakterien *Mycobacterium tuberculosis*. Trots befintliga läkemedel samt vaccin, skördar sjukdomen tuberkulos 1.4 miljoner människors offer varje år. Under årens lopp har bakterien hittat sätt att förhindra verkan av medicinsk behandling och utvecklat resistens mot dagens tillgängliga antibiotika. Detta har lett till en markantökning av resistenta stammar runt om i världen. Behandlingen av tuberkulos är idag ineffektiv, kostsam och mycket långdragen. Härav finns det ett stort behov av nya läkemedel i kampen mot tuberkulos.


En bakteries förmåga att framkalla sjukdom hos sin värd kallas virulens. Nyckeln till att skapa nya mediciner i kampen mot sjukdomsfiramkallande bakterier är att identifiera faktorer som är involverade i dess virulens. Hos tuberkelbakterien är en sådan faktor dess speciella oljerock. Många av de mest effektiva läkemedlen är sådana som hindrar tillverkningen av fetter som bygger upp denna yttre barriär.

I min avhandling har jag, tillsammans med andra forskare, tittat på proteiner som är viktiga för tuberkelbakteriens överlevnad. Om dessa proteiner inaktiveras, leder det till att bakterien får förändringar i sin oljerock, blir känsligare mot antibiotika och får svårare att överleva. Förståelsen för hur proteinerna ser ut och fungerar, är nödvändigt för att utveckla nya substanser som kan hindra deras funktion. Dessa substanser skulle kunna användas som nya läkemedel i kampen mot tuberkulos.


7. Acknowledgements

Last but not least, this thesis had not been possible to write if it were not for a number of people. You have all made my life so much easier with your individual contributions, some scientific, others more of a social and fun character. There are many people not mentioned here, but this does not mean you are forgotten!

First of all I would like to thank my brilliant supervisor Martin. You were brave to take me on as your first student into a (physically) non–existing lab. I am very happy that you had the courage to do this. Throughout the years we have had quite a trip together. You started off as a young eager researcher, growing more confident and professor–like each year. Thank you for always keeping your enthusiasm throughout my stumbling attempts of doing science!

I would also like to thank the two Wieslander groups, Åke and Lars, and people from biophysics and the Gräslund group for rewarding collaborations. I would especially like to thank Pål for handing me the "just–do–it" project. I am quite sure that the atmosphere during "fika" would not be the same without your politically incorrect comments. You are a fantastic resource in many ways, but foremost a great crystallographer. Jag uppskattar att du nästan alltid har spenderbyxorna på.

Speaking about great colleagues. What would I have done if it wasn’t for all of you fantastic people? The first member to join the "Högborn group", but crazily hard to kick out, was our own Dr. Phil (Mike). You taught Catrine how to curse and me how to drink lukewarm/British beer. I don’t know which of the two was most appreciated, but I am equally grateful. Next came my physically closest and longest–lasting office friend Karin. We have shared many beds, but please let us avoid the allergy–inducing rooms in the future. Dr. Banina–Banana for keeping our spirits up on a nearly hopeless project. We all miss your "pappa–skämt", especially teddy–Dan whom tries his best to keep up a steady pace of bad jokes. The ever so calm Henryk and your energetic wife–to–be Honorata! If it
weren’t for you, I would still be studying for the BIG-exam. The
one person in the group that earns my eternal gratitude for her
impeccable English and great muffins is Julia. It felt nice to finally get
some company in "RNR-träsket". I want to thank all past, present
and future members of the Högbom/Stenmark group: Mike, Karin,
Catrine, Dan, Auður, Henryk, Daniel M.M., Linda, Julia, Ronnie,
Alex, Camilla, Matthew, other visitors of the groups and of course
the missing-link, Ann-Louise, whom made us connect to people in
the other corridor. Some of you have grown to dear friends over the years.

Further thanks goes to people of the C4-corridor. We have
had many sweet Friday cake clubs, intense lunch discussions and bubbly
AfterWorks. Great discussions over the daily "fika" often involved:
Maria R., Kristoffer, Linnea, Pär, Daniel N., Jens, and Katta.

In fact, all the people of DBB have contributed to the fantas-
tic atmosphere here at the department. A special thanks to Stefan,
Peter, Ann, Maria, Lotta, Heidi, Håkan, Torbjörn and all others that
put in so much good work into running this department and for making
a student’s life so much easier.

Jag lämnade Linköping för många år sedan, trots detta har
jag det senaste året spenderat fler helger där än vad jag har hemma i
lägenheten på Kungsholmen. Tack Fredrik M., L., N., Gilljam, Malin
och Anders som gjorde och gör Linköping lite roligare, men kanske
mest av allt för all galenskap, kalasande och för ert fantastiska öl– och
teknik–snobberi.

Vännerna i Finland. Patrik och Maria. Ni är långt borta fast
ändå alltid närvarande. Jag saknar er massor.

Jesper för alla sköna frukostar då du lärt mig om livet
allmänhet men kanske om livets njutningar i synnerhet. I sommar vill
jag följa med och segla! Allas vår finlandssvenske och bastu-badande
Dennis. Också Hanna, Carina, Lingon, Fredde B. och Lisa vad vore
besöken till Tynningö utan ert härliga sällskap?

Tack till Åsa för att du kämpar för en bättre värld! Jag upp-
skattar verkligen att vi pratar om annat än naturvetenskap när vi ses.
Vi är några som väntar med spänning på Mads–Eriks sommarprojekt:
"Den glade dansken". Kärlek.

Min vän Aron. Du har alltid trott på mig, även gånger jag
varit övertygad om motsatsen. Oftast eller kanske nästan alltid har du
också haft rätt. Jag uppskattar verkligen att du finns i mitt liv.


The tuberculist database.
http://genolist.pasteur.fr/TubercuList/.

