Targeting Infectious Disease

Structural and functional studies of proteins from two RNA viruses and Mycobacterium tuberculosis

ANNA M. JANSSON
Abstract


The recent emergence of a number of new viral diseases as well as the re-emergence of tuberculosis (TB), indicate an urgent need for new drugs against viral and bacterial infections.

Coronavirus nsp1 has been shown to induce suppression of host gene expression and interfere with host immune response. However, the mechanism behind this is currently unknown. Here we present the first nsp1 structure from an alphacoronavirus, Transmissible gastroenteritis virus (TGEV) nsp1. Contrary to previous speculation, the TGEV nsp1 structure clearly shows that alpha- and betacoronavirus nsp1s have a common evolutionary origin. However, differences in conservation, shape and surface electrostatics indicate that the mechanism for nsp1-induced suppression of host mRNA translation is likely to be different in the alpha- and betacoronavirus genera.

The Modoc virus is a neuroinvasive rodent virus with similar pathology as flavivirus encephalitis in humans. The flaviviral methyltransferase catalyses the two methylations required to complete 5’ mRNA capping, essential for mRNA stability and translation. The structure of the Modoc NS5 methyltransferase domain was determined in complex with its cofactor S-adenosyl-L-methionine. The observed methyltransferase conservation between Modoc and other flaviviral branches, indicates that it may be possible to identify drugs that target a range of flaviviruses and supports the use of Modoc virus as a model for general flaviviral studies.

1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR) is part of the methylerythritol phosphate (MEP) pathway that produces essential precursors for isoprenoid biosynthesis. This pathway is used by a number of pathogens, including Mycobacterium tuberculosis and Plasmodium falciparum, but it is not present in humans. Using a structure-based approach, we designed a number of MtDXR inhibitors, including a novel fosmidomycin-analogue that exhibited improved activity against P. falciparum in an in vitro blood cell growth assay. The approach also allowed the first design of an inhibitor that bridge both DXR substrate and co-factor binding sites, providing a stepping-stone for further optimization.

Keywords: RNA virus, coronavirus, alphacoronavirus, nsp1, TGEV, flavivirus, Modoc virus, NS5, methyltransferase, mRNA capping, Mycobacterium tuberculosis, tuberculosis, 1-deoxy-D-xylulose 5-phosphate reductoisomerase, DXR, fosmidomycin analogues, MEP pathway, drug development, xray-crystallography

Anna M. Jansson, Uppsala University, Department of Cell and Molecular Biology, Structure and Molecular Biology, 596, SE-751 24 Uppsala, Sweden.

© Anna M. Jansson 2013

ISSN 1651-6214
ISBN 978-91-554-8618-1
urn:nbn:se:uu:diva-196623 (http://urn.kb.se/resolve?urn=nbn:se:uu:diva-196623)
List of Papers

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


Reprints were made with permission from the respective publishers.

* These authors contributed equally to this work.


Contents

Introduction .......................................................................................................................... 9
  Aim of the thesis .................................................................................................................. 9

Coronaviruses ..................................................................................................................... 11
  Genome organization ......................................................................................................... 11
  Taxonomy ............................................................................................................................ 12
  Coronavirus diseases ......................................................................................................... 12
  Coronavirus nsp1 ................................................................................................................ 13

TGEV nsp1 (Paper I) .......................................................................................................... 14
  The structure ...................................................................................................................... 15
  Conservation within alphacoronavirus nsp1 ..................................................................... 16
  Nsp1 conservation between the genera ........................................................................... 17
  The function of alphacoronavirus nsp1 .......................................................................... 19

Flaviviruses ......................................................................................................................... 21
  Genome organization ........................................................................................................ 21
  mRNA capping ................................................................................................................... 22
  Diversity of methyltransferases ....................................................................................... 24
  The NS5 protein ................................................................................................................. 24
  Development of antivirals ................................................................................................. 24
  Targeting the flaviviral NS5 methyltransferase ............................................................... 25

Methyltransferase domain from Modoc virus (Paper II) ...................................................... 27
  Modoc virus MTase ............................................................................................................ 27
  The MTase_{MODV} structure ........................................................................................... 28
  Binding sites ....................................................................................................................... 28
    MTase_{MODV} SAM-binding site .................................................................................. 29
    GTP and RNA binding sites .......................................................................................... 30

Targeting isoprenoid synthesis in *Mycobacterium tuberculosis* ....................................... 32
  Tuberculosis .................................................................................................................... 32
  *Mycobacterium tuberculosis* .......................................................................................... 33
  Isoprenoid biosynthesis *via* the MEP pathway ............................................................... 33
  1-deoxy-D-xylulose 5-phosphate reductoisomerase ......................................................... 35
  Previous work on MtDXR ............................................................................................... 36
Structural and functional studies of fosmidomycin analogues as MtDXR inhibitors (Paper III)........................................................................................................................................37

Structural details of fosmidomycin analogues..........................................................37
MtDXR enzymatic assay ..............................................................................................40
Crystallographic data of MtDXR-inhibitor complexes ..............................................40
In vitro assays: Mt cells and Pf infected blood cells..................................................43
Inhibition of MtDXR and antimalarial activity .........................................................43

Populärvetenskaplig sammanfattning .................................................................45

Acknowledgements ...............................................................................................48

References ..............................................................................................................51
**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>α-CoV</td>
<td>alphacoronavirus</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>β-CoV</td>
<td>betacoronavirus</td>
</tr>
<tr>
<td>β-CoV&lt;sub&gt;A&lt;/sub&gt;</td>
<td>betacoronavirus lineage A</td>
</tr>
<tr>
<td>β-CoV&lt;sub&gt;B&lt;/sub&gt;</td>
<td>betacoronavirus lineage B</td>
</tr>
<tr>
<td>BCG</td>
<td>Bacillus Calmette-Guérin</td>
</tr>
<tr>
<td>γ-CoV</td>
<td>gammacoronavirus</td>
</tr>
<tr>
<td>CDP-ME</td>
<td>4-diphosphocytidyl-2-C-methyl-D-erythritol</td>
</tr>
<tr>
<td>CDP-ME2P</td>
<td>4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate</td>
</tr>
<tr>
<td>CoV</td>
<td>coronavirus</td>
</tr>
<tr>
<td>DENV</td>
<td>Dengue virus</td>
</tr>
<tr>
<td>DMAPP</td>
<td>dimethylallyl pyrophosphate</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>E. coli</td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>Ec</td>
<td><em>Escherichia coli</em></td>
</tr>
<tr>
<td>ESRF</td>
<td>European Synchrotron Radiation Facility</td>
</tr>
<tr>
<td>GMP</td>
<td>guanosine 5’-monophosphate</td>
</tr>
<tr>
<td>GTase</td>
<td>guanylyltransferase</td>
</tr>
<tr>
<td>GTP</td>
<td>guanosine triphosphate</td>
</tr>
<tr>
<td>HCoV</td>
<td>human coronavirus</td>
</tr>
<tr>
<td>HIV</td>
<td>human immunodeficiency virus</td>
</tr>
<tr>
<td>IBV</td>
<td>infectious bronchitis virus</td>
</tr>
<tr>
<td>IPP</td>
<td>isopentenyl pyrophosphate</td>
</tr>
<tr>
<td>IRES</td>
<td>internal ribosome entry site</td>
</tr>
<tr>
<td>MDR-TB</td>
<td>multidrug-resistant tuberculosis</td>
</tr>
<tr>
<td>MEcPP</td>
<td>2-C-methyl-D-erythritol 2,4-cyclopyrophosphate</td>
</tr>
<tr>
<td>MEP</td>
<td>2-C-methyl-D-erythritol 4-phosphate</td>
</tr>
<tr>
<td>MHV</td>
<td>Mouse hepatitis virus</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>Mt</td>
<td><em>Mycobacterium tuberculosis</em></td>
</tr>
<tr>
<td>MTase</td>
<td>methyltransferase</td>
</tr>
<tr>
<td>MVA</td>
<td>mevalonate</td>
</tr>
<tr>
<td>NADPH</td>
<td>nicotinamide adenine dinucleotide phosphate</td>
</tr>
<tr>
<td>NCR</td>
<td>non-coding regions</td>
</tr>
<tr>
<td>NKV</td>
<td>no known vector</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>NS</td>
<td>non-structural protein</td>
</tr>
<tr>
<td>nsp</td>
<td>non-structural protein</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>Pf</td>
<td><em>Plasmodium falciparum</em></td>
</tr>
<tr>
<td>RdRp</td>
<td>RNA-dependent RNA polymerase</td>
</tr>
<tr>
<td>RRL</td>
<td>rabbit reticulocyte lysate</td>
</tr>
<tr>
<td>RTPase</td>
<td>RNA 5’-triphosphatase</td>
</tr>
<tr>
<td>SAH</td>
<td>S-adenosyl-L-homocysteine</td>
</tr>
<tr>
<td>SAM</td>
<td>S-adenosyl-L-methionine</td>
</tr>
<tr>
<td>SARS</td>
<td>severe acute respiratory syndrome</td>
</tr>
<tr>
<td>SARS-CoV</td>
<td>SARS coronavirus</td>
</tr>
<tr>
<td>SeMet</td>
<td>selenomethionine</td>
</tr>
<tr>
<td>sg-mRNA</td>
<td>subgenomic messenger RNA</td>
</tr>
<tr>
<td>SIRAS</td>
<td>single isomorphous replacement with anomalous scattering</td>
</tr>
<tr>
<td>TB</td>
<td>tuberculosis</td>
</tr>
<tr>
<td>TGEV</td>
<td>Transmissible gastroenteritis virus</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
</tr>
<tr>
<td>XDR-TB</td>
<td>extensively drug-resistant tuberculosis</td>
</tr>
<tr>
<td>WNV</td>
<td>West Nile virus</td>
</tr>
</tbody>
</table>
Introduction

In 2002, a virus causing severe acute respiratory syndrome (SARS) spread rapidly around the world. Almost 8000 people were infected and more than 800 people died before the SARS epidemic was contained in July 2003. The etiological agent of this life-threatening disease was soon identified to be a previously unknown coronavirus, SARS-coronavirus (SARS-CoV). Before the SARS outbreak, coronaviruses were only known to cause mild symptoms in humans, and it was an eye-opener that coronaviruses could cause such severe disease. The World Health Organization fears that this type of outbreak is likely to happen again, but that it is difficult to predict when and where.

SARS is a new threat, but even diseases that we associate with the past are re-emerging, often in more virulent and resistant forms. Tuberculosis is one such disease. For a long time tuberculosis has been a disease mainly prevalent in developing countries, and many people have lived with the belief that it had been eradicated. However, tuberculosis is an increasing threat, although a successful program initiated by the World Health Organization has decreased the number of new cases in the past couple of years. Despite this, drug resistant forms of the bacteria are emerging. By the end of 2011, 77 countries had reported at least one case of extensively drug-resistant tuberculosis.

To be able to deal with infectious diseases like these, there is an urgent need for new drugs as well as research to better understand the pathogenesis and mechanisms by which these organisms function.

Aim of the thesis

Two of the papers in this thesis describe the structural features of two different non-structural proteins from the RNA viruses Transmissible gastroenteritis virus (TGEV) and Modoc virus. The aim of the first paper was to shed light on the function and evolutionary relationship of the coronavirus protein non-structural protein 1 by determining the structure of TGEV nsp1, the first nsp1 structure from this genus. The second paper describes the structure of Modoc virus methyltransferase. Methyltransferases are attractive drug targets against flaviviruses. To assess the value of the Modoc virus as a model
system for flaviviruses, the structure of the Modoc virus methyltransferase was determined.

The third paper describes structural and functional studies of 1-deoxy-D-xylulose 5-phosphate reductoisomerase from *Mycobacterium tuberculosis* (*MtDXR*). DXR catalyses the second step in the methylerthritol phosphate pathway, a pathway that produces isoprenoid precursors. Isoprenoids are compounds essential to all living organisms, and the isoprenoid precursors can be produced by two distinct pathways. In humans, isoprenoid precursors are synthesized by the mevalonate pathway. In contrast, mycobacteria and protozoa use the methylerthritol phosphate pathway, which make the enzymes involved in this pathway attractive as drug targets. DXRs from several other pathogenic organisms have been shown to have similar active sites to that of *MtDXR*, indicating that inhibitory compounds developed against *MtDXR* could also target e.g. DXR from *Plasmodium falciparum*. Activity measurements in combination with determined protein-inhibitor complex structures were used to guide the design of inhibitory compounds.
Coronaviruses

Coronaviruses (CoVs) are enveloped viruses containing capped and polyadenylated positive-stranded RNA genomes. The genome, which has a high level of complexity, is between 26-32kb in length. This is the largest genome of any RNA virus identified to date (Groot et al., 2011). The first coronavirus was isolated in the 1930s - the Infectious bronchitis virus (IBV). In the 1940s Mouse hepatitis virus (MHV) and Transmissible gastroenteritis virus (TGEV) were discovered. On the basis of the characteristic spikes on the capsid surface, visible by electron microscopy, these viruses were grouped and named after their crown-like appearance – coronavirus. Since the SARS outbreak in 2003, new interest has been directed at coronavirus emergence and pathogenesis to be able to find treatment and prevention strategies (Weiss and Navas-Martin, 2005).

Genome organization

The coronavirus genome contains seven to fourteen open reading frames (ORFs), six of which are common to all coronaviruses (Figure 1). The first two, 1a and 1b, constitute the replicase gene. The other four, S, E, M and N, encode structural proteins that make up the capsid or associate with the genomic RNA. The genome also contains a variable number of accessory proteins (Gorbalenya et al., 2006). The replicase gene is translated directly from the genomic RNA as either ORF1a, or ORF1ab via a ribosomal frameshift located between the two ORFs (Brierley et al., 1989). These large polyproteins are cleaved by ORF1a-encoded proteases into 16 non-structural proteins (nsp) involved in replication and evasion of host defence (Ziebuhr et al., 2000; Wathelet et al., 2007; Narayanan et al., 2008). The replicase gene products mediate both replication of the genome as well as transcription of the structural and accessory proteins, which are translated from nested subgenomic mRNA (sg-mRNA). The sg-mRNA templates are generated via minus-strand RNAs, through a combination of sophisticated regulatory mechanisms, unique for coronaviruses (Sawicki et al., 2007). Each resulting piece of mRNA carries a common untranslated region (UTR) leader sequence in the 5'-end, as well as a common sequence in the 3'-end.
Taxonomy

Coronaviruses belong to the order Nidovirales and the family Coronaviridae. This family is further divided into two subfamilies, Coronavirinae, constituting the “original” coronaviruses that were classified based on their crown shaped capsids, and Torovirinae. The Coronavirinae subfamily branches out into three genera: alpha-, beta-, and gammacoronaviruses (α-, β-, and γ-CoVs) (Groot et al., 2011). The first two infect mammals whereas the third mainly infects birds. The coronaviruses discussed in this thesis belong to the alpha- and betacoronavirus genera. In the α-CoV group we find, for example, TGEV and the two human CoVs HCoV-229E and HCoV-NL63. TGEV is the prototype species for the α-CoV genus. Mouse hepatitis virus is the prototype species in the β-CoV group, belonging to lineage A, β-CoVA. SARS-CoV, the etiological agent of SARS, is also a member of the β-CoV-genus, but in lineage B, β-CoVB.

CORONAVIRUS GENOME ORGANIZATION

Figure 1. Coronavirus genome organization. The non-structural proteins are in dark gray, the structural proteins in light gray. White boxes represent accessory proteins that vary between the genera. ORF1a and ORF1b are translated as polyproteins and are cleaved into 16 different non-structural proteins by viral proteases.

Coronavirus diseases

Coronaviruses mostly replicate in the epithelial cells of the respiratory or enteric tracts where they cause enteric and respiratory diseases. The viruses are connected to large economic losses when cattle, poultry and pigs are infected. In humans, CoVs have been known to cause mainly mild disease, and they are the second most common cause of the common cold after rhinovirus (Lai et al., 2007). However, the SARS outbreak demonstrated that coronaviruses do have the potential to cause severe disease also in humans. Coronaviruses generally infect only one animal species, but SARS-CoV proved to be an exception to this rule, where the virus came from bats, ad-
adapted to palm civets, and finally to humans (Guan et al., 2003; Yip et al., 2009). TGEV, in the α-CoV genus, replicates in the intestinal enterocytes and causes diarrhea in pigs. For piglets, this acute disease is almost always fatal. However, not much is known about TGEV in comparison to other coronaviruses, despite its economical importance (Laude et al., 1990; Kim et al., 2000).

**Coronavirus nsp1**

Nsp1 is the most N-terminal gene product in the CoV genome. CoV nsp1 has for a long time been associated with suppression of host gene expression, including that of type I interferon (Kamitani et al., 2006). For SARS-CoV nsp1, this is accomplished both through degradation of host mRNA and suppression of host translation. SARS-CoV nsp1 has also been shown to bind to the 40S subunit of the ribosome (Kamitani et al., 2009; Huang et al., 2011). Recent data indicate that TGEV nsp1 also can suppress host mRNA translation, but does not seem to bind to the 40S subunit nor does it have the ability to degrade mRNA (Huang et al., 2011).

There have been indications that CoV nsp1 also suppresses the expression of the CoV genes (Brockway and Denison, 2005; Huang et al., 2011). However, recent experiments suggest that the virus has a mechanism for protecting its own mRNA. Experiments performed on SARS-CoV nsp1 show that the interaction between the nsp1 protein and the first stem loop in the 5′-UTR effectively protects the CoV mRNA from the nsp1-mediated suppression of mRNA translation (Tanaka et al., 2012). As mentioned previously, coronaviruses have a mechanism to produce mRNA that share 5′- and 3′-sequences, suggesting that all viral mRNAs are protected via this 5′ leader element.

The replicase gene is generally well conserved between the different coronavirus genera. The greatest difference is in the N-terminus, and the nsp1 protein is considered a genus specific marker (Snijder et al., 2003; Connor and Roper, 2007). The nsp1s vary substantially in size in the different CoV clusters, where the α-CoV, β-CoV and β-CoV-B proteins are about 110, 250 and 180 amino acids long, respectively. γ-CoVs do not have an nsp1. The nsp1 sequence similarity between these clusters is very low, and the fact that homology based on sequence could not be inferred between the different nsp1s, nor with any host protein, raised the question whether or not these proteins share the same structure and function (Connor and Roper, 2007).
The TGEV nsp1 protein was recombinantly expressed in *E. coli* BL21-AI cells. Initial work was done with a clone expressing the full length TGEV nsp1 with an N-terminal His-tag. The expressed protein was soluble and produced in good yield. The protein was purified by trapping to a Ni-sepharose column, followed by size exclusion chromatography, and then concentrated. For crystallization, various commercial screens were set up at different protein concentrations. Crystals appeared in several conditions but were difficult to reproduce. A third purification step, an anion exchange column, was introduced after the gel filtration. The crystallization results were still batch-dependent, although every batch was purified in the same way. Some batches of protein failed to crystallize at all.

The first crystals were taken to MAXlab, Lund, Sweden, where cryo conditions were optimized and initial data were collected. However, these data were of poor quality and low resolution. In an attempt to improve crystal packing, the sequence was analyzed with secondary structure prediction software to detect unstructured ends in the N- or C-terminal that could be removed. After analysis with Phyre (Kelley and Sternberg, 2009) and I-Tasser (Zhang, 2008), a new construct was designed where the C-terminal end of the protein was truncated with five amino acids. This truncated version of the protein could be expressed, purified and crystallized under the same conditions as the full length construct. Crystals were brought to the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. Native data were collected to 1.5Å and the space group was determined to be P1. Since the TGEV nsp1 sequence does not show any significant sequence similarity to any other protein, SeMet TGEV nsp1 was produced. Crystals appeared in the same conditions as the native protein. The SeMet incorporation was verified with mass spectrometry. However, the anomalous signal of these crystals was weak and did not give enough phasing power to solve the structure. The lack of crystallographic symmetry also increased the difficulty collecting redundant anomalous data.

New native crystals were produced, this time soaked with different heavy atom salts to obtain a phasing signal. Derivative crystals were brought to ESRF. Crystals soaked with K\(_2\)PtCl\(_4\) diffracted to about 2.4Å and produced a sharp peak close to the Pt-absorption edge despite a 30' back soak in mother liquor. Two 360°-sweeps were collected at the peak using two different kappa angles. The derivative and the native data were scaled together and four
four platinum sites were identified by single isomorphous replacement with anomalous scattering (SIRAS) using ShelxD. The sites were further refined in SHARP (de La Fortelle and Bricogne, 1997). Subsequent solvent flattening and histogram matching using DM (Cowtan and Main, 1998) significantly improved the electron density and extended phases beyond 2Å. Buccaneer (Cowtan, 2006) was used to create a first trace of the polypeptide backbone. The initial model was further improved by alternate cycles of model rebuilding in the program O (Jones et al., 1991) and refinement in autoBuster (Bricogne et al., 2011).

The structure

The TGEV nsp1 structure is characterized by a six stranded β-barrel with an α-helix on the side of the opening of the barrel. The barrel is flanked by a small β-sheet, with a short $3_{10}$-helix preceding one of the strands. Strands $\beta_3$, $\beta_7$, $\beta_5$ and $\beta_6$ make an anti-parallel sheet forming one side of the barrel, where $\beta_3$ and $\beta_6$ are loosely connected to $\beta_1$ and $\beta_8$ that make up the other side of the barrel. The structure is compact, lacking long flexible loops (Figure 2).

The only other coronavirus nsp1 structure available is an NMR structure of the N-terminal domain of SARS-CoV nsp1. However, SARS-CoV belongs to lineage B in the genus β-CoV, and its nsp1 protein is different in size and shows no significant sequence similarity to TGEV nsp1. The SARS-CoV structure contains residues 13-128 from the 180 amino acid long protein, where the C-terminal part is assumed to be disordered (Almeida et al., 2007).
Despite the lack of sequence similarity, the TGEV nsp1 β-barrel fold was found to exhibit significant similarity to the SARS-CoV nsp1\textsuperscript{13-128} structure (Figure 3). However, the structures also display clear differences. The small β-sheet flanking the barrel in the TGEV nsp1 structure is absent in the SARS-CoV nsp1\textsuperscript{13-128} structure. In addition, the strands in the barrel have different angles and the position of the long helix is shifted. The loop between β5 and β6 is also longer in the SARS-CoV nsp1\textsuperscript{13-128} structure. Although these might seem like small differences, the difference between solvent accessible surface in the two structures is larger than first expected. The shape of the structures is different, and the SARS-CoV nsp1\textsuperscript{13-128} structure is significantly bulkier than the TGEV nsp1.

Figure 3. Stereo view of TGEV nsp1 (green) and SARS-CoV nsp1\textsuperscript{13-128} (orange).

**Conservation within alphacoronavirus nsp1**

In order to find clues on areas and residues that could be important for the function of the α-CoV nsp1 proteins, sequences from other nsp1s in the α-CoV genus were gathered by iterative searches using the most distant group member in the next search. Eight sequences were subsequently selected on the basis of diversity and aligned. The alignment featured a number of conserved residues. The majority of the conserved residues were found in the hydrophobic core of the nsp1 structure and can be assumed to be involved in the stability of the protein and not the function. However, when plotting the conserved residues on the surface of the protein, two loosely connected patches emerge (Figure 4). Both patches are located along a ridge on one side of the molecule. A few highly conserved hydrophilic residues stand out as potential candidates to be involved in nsp1 activity or interaction with a partner or target molecule. The first patch include Asp13, Gln15, Asn92 and Asn94, located in the loops between β1-β2 and β7-β8. The second patch is
mainly made up of side chains from strand β8. This group also includes a couple of conserved hydrophobic residues together with Glu98 and Asp99, located in the N-terminal portion of β8.

Figure 4. The surface of TGEV nsp1 displays two areas with high sequence conservation. Two loops make up the first area (A), where Asp13, Gln15, Asn92, and Asn94 form a conserved circle. The second area (B), centered on strand β8, displays both exposed hydrophobic residues and charged residues that potentially could interact with a partner molecule or partner protein.

Nsp1 conservation between the genera

To compare the conservation patterns in the two genera, four representative nsp1 sequences from viruses in β-CoV_B were aligned. Based on the superposition of the two nsp1 structures, the separate α-CoV and β-CoV_B alignments were merged. Interestingly, only a few areas of the combined alignment overlap. The residues that are conserved between the two genera are mainly hydrophobic residues found in the β-barrel core. When mapping the conserved residues in group β-CoV_B on the surface of the SARS-CoV nsp113-128 structure, the pattern is very different from the one seen in TGEV nsp1, and there is very little overlap in the conservation. There is also poor correlation between conserved patches and the electrostatic potential of the surfaces (Figure 5).
Figure 5. Superposed structures of TGEV nsp1 and SARS-CoV nsp1$^{13-128}$ are presented separately in two different rotations: rotation 1 (A to F) and rotation 2 (G to L). Conserved residues within each genus are shown in green, where darker green corresponds to a higher level of conservation. In each rotation, TGEV nsp1 and SARS-CoV nsp1$^{13-128}$ are shown as cartoons and surface representation displaying the conserved areas. Subfigures C, F, I, and L show the electrostatic surface potentials of the two structures. The area within the dotted oval in subfigure H corresponds to the two conserved areas shown in Figure 4.
The function of alphacoronavirus nsp1

The conserved areas within the α-CoV genus does not give any immediate indication on the function or molecular mechanism of the nsp1 proteins. Two conserved patches emerge as being of particular interest, harboring residues that could potentially be important for nsp1 function. The lack of cavities in the structure and the protruding shape of the ridge, suggests that the α-CoV nsp1 might be involved in protein-protein interactions. This is supported by experiments showing that TGEV nsp1 suppresses protein translation in cell-free HeLa extracts but not in rabbit reticulocyte lysate (RRL). This suggests that a host factor that exists in the HeLa extracts but not in RRL is needed for TGEV nsp1 function (Huang et al., 2011).

The majority of the conserved residues overlapping between α-CoV and β-CoV\textsubscript{B} seem to be involved in the stability of the β-barrel fold and there is little overlap between the conserved patches on the surface. Interestingly, biochemical studies of the coronavirus nsp1s have highlighted both differences and similarities between the different genera. Nsp1s from α-CoV and both the A and B lineages of β-CoV have been shown to suppress host mRNA translation (Narayanan et al., 2008; Kamitani et al., 2009; Huang et al., 2011). However, in equivalent experiments, where SARS-CoV nsp1 has been shown to render host mRNA translationally inactive, bind to the 40S-subunit of the ribosome and promote mRNA degradation, TGEV nsp1 does suppress translation of host mRNA, but without binding to the 40S subunit or by promoting mRNA degradation (Huang et al., 2011).

Contrasting to these results, two other α-CoV nsp1 proteins from HCoV-229E and HCoV-NL63, were shown to bind the ribosomal 6S protein which is a part of the 40S subunit (Wang et al., 2010). Both these proteins share the conserved areas within α-CoV nsp1s. In the same study, three-dimensional structure predictions of HCoV-229E and HCoV-NL63 were presented, showing a much higher structure similarity to SARS-CoV nsp1\textsubscript{13-128} than is seen between TGEV nsp1 and SARS-CoV nsp1\textsubscript{13-128} (Wang et al., 2010). If the nsp1 structures from HCoV-229E/HCoV-NL63 and SARS-CoV are more closely related remains to be elucidated by experimentally determined structures. Until then the possibility that the prediction was biased by the SARS-CoV nsp1\textsubscript{13-128} structure cannot be excluded. However, the HCoV binding data do suggest that the closely related TGEV nsp1 might also have the ability to bind the 40S subunit via the 6S protein.

The limited amount of degraded mRNA in the presence of SARS-CoV nsp1 compared to the significant suppression of reporter gene expression suggests that there are two mechanisms for suppression of host mRNA translation, where SARS-CoV nsp1 induced mRNA cleavage only accounts for part of the suppression (Kamitani et al., 2009). It might seem reasonable to hypothesize that the β-barrel domain of SARS-CoV nsp1\textsubscript{13-128} and TGEV
nsp1 share a similar mechanism for the 40S/mRNA cleavage independent suppression of host mRNA translation. However, this is not supported by experimental data. Experiments done with a SARS-CoV nsp1 carrying K164A and H165A mutations in the C-terminal domain showed that the mutant was completely inactive. Compared to wild type SARS-CoV nsp1, the mutant form was unable to bind the 40S-subunit and hence could not degrade mRNA. The mutant also failed to suppress translation, indicating that the N-terminal domain alone is not responsible for this additional suppression.

The shared fold of TGEV nsp1 and SARS-nsp1 suggests that nsp1s from the different coronavirus genera share common ancestry despite the large variance in sequence. However, the conservation pattern and differences in surface electrostatics and shape between the α-CoV and β-CoV nsp1 structures, together with the biochemical data, led us to speculate that the mechanism for 40S-independent nsp1-induced suppression of host mRNA translation is different in the two genera.
Flaviviruses

The genus flavivirus is one out of three genera in the family *Flaviviridae*. The other two genera are pestivirus and hepacivirus. In addition, a fourth genus is currently under evaluation (Simmonds *et al.*, 2011). Viruses in the *Flaviviridae* family are enveloped and carry a positive stranded RNA genome with a size between 9.6 and 12.3 kb. Common to all members of this family is the lack of a poly-A tail in the 3′-end (Wengler, 1981). A feature specific for the flavivirus genus is a specific type 1 cap structure in the 5′-end of the genome, needed for transcription and replication (Chambers *et al.*, 1990; Choi *et al.*, 2004).

There are currently more than 50 different species in the flavivirus genus, divided into three groups: mosquito-borne, tick-borne and a third group with no known vector (NKV). Many flaviviruses are severe human pathogens like the Dengue virus, the West Nile virus and the Yellow fever virus. An incredible 40% of the world population is living in areas where they are at risk of being infected with the Dengue virus. The World Health Organization estimates that 50-100 million people are infected each year and about 20 000 people die from the disease (World Health Organization, 2012a). For Yellow fever virus, about 200 000 people are infected and 30 000 die each year (World Health Organization, 2011). Flaviviruses are spread all around the globe and cause extensive disease and mortality (Gubler, 2002).

Genome organization

The flavivirus genome consists of one ORF, flanked on both sides by two non-coding regions (NCR) (Chambers *et al.*, 1990; Markoff, 2003). A conserved di-nucleotide AG-sequence is found in the 5′-end, after the type 1 cap (Cleaves and Dubin, 1979). The viral polyprotein is co- and posttranslationally processed into three structural proteins and seven non-structural proteins (NS). The structural proteins are called C, prM and E, where C is the capsid protein and the other two are envelope proteins. The seven non-structural proteins are called NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5 (Figure 6). NS2B, together with the N-terminal one-third of NS3, form the viral serine protease involved in processing of the viral polyprotein together with host proteases. NS5 is the largest of the non-structural proteins and shows the highest level of conservation (Mandl *et al.*, 1989; Chambers *et al.*, 1990).
The N-terminal part of NS5 harbours the methyltransferase domain, and the C-terminal part is the RNA-dependent RNA polymerase (RdRp). The methyltransferase domain has recently been shown to also have guanylyltransferase activity, and has therefore been suggested to be called the NS5 capping enzyme (CE) (Geiss et al., 2011).

Figure 6. Flavivirus genome organization.

mRNA capping

The mRNA cap is a unique structure found at the 5′-end of mRNA (Figure 7). The cap promotes stability and protects the RNA from 5′-exonucleases and is important for the recognition of the mRNA by the ribosomal protein eIF4E. The cap also enhances initiation of mRNA translation (Furuichi and Shatkin, 2000; Shuman, 2001; Gu and Lima, 2005).

The nascent mRNA has a triphosphate in the 5′-end (pppN-RNA). In the conventional capping process, one of these phosphates is first hydrolyzed by an RNA 5′-triphosphatase (RTPase). Next, a guanosine 5′-monophosphate (GMP) is linked to the last phosphate in a 5′-5′ orientation by a guanylyltransferase (GTase), converting it into GpppN-RNA. This is followed by one or several methylation events. The N7 of the guanine is methylated by a guanine-N7-methyltransferase, making a cap 0 structure (m7GpppN-RNA). This is also called a minimal cap, and has the lowest level of methylation. The next reaction is carried out by a nucleoside-2′-O-methyltransferase that methylates the 2′-OH group of the first nucleotide, making a cap 1 (m7GpppNm-RNA). This is the cap found in flaviviruses. In some cases, the 2′-OH is methylated on the next nucleotide as well, making it a cap 2 structure (m7GpppNmNm-RNA). The methylation reactions use S-adenosyl-L-methionine (SAM) as a cofactor and produces S-adenosyl-L-homocysteine (SAH) as a by-product.

The mRNAs cap of the host is added directly in the nucleus when the mRNA exits the RNA polymerase (Shatkin, 1976a; Shatkin and Manley, 2000). Since most viruses are transcribed and replicate in the cytoplasm, they cannot use the capping machinery of the host. Viruses have solved this in a wide range of ways. Most viruses express their own capping machinery and use the conventional route for the acquisition of the cap. However, some viruses have alternative features to the cap structure. Other viruses use an internal ribosome entry site (IRES) in their 5′-UTR. With the help of the IRES, the virus is able to directly recruit the ribosomal subunits (Cullen,
Other viruses that have uncapped RNA, e.g. caliciviruses, covalently attach their RNA to a protein called VPg. This VPg protein can directly interact with the cap-binding protein eIF4E and initiate translation (Goodfellow et al., 2005). However, the most common way for viruses to avoid mRNA degradation and ensure efficient translation is to acquire a 5′-end cap. Not all viruses use the conventional route for acquisition of the cap, but use completely different pathways. For example, Semliki forest virus, methylates GTP, and the m7GMP is subsequently transferred to the 5′-end of the mRNA (Ahola and Kaariainen, 1995). Others, like the influenza virus, use ‘cap-snatching’, meaning that they steal the cap of a capped host mRNA and attach it to viral mRNA (Krug et al., 1981; Plotch et al., 1981). In several viruses, the order of the methylation events is switched. However, the Modoc virus, belonging to the genus flavivirus, employs the conventional mode of capping.

Figure 7. Cap 1 structure found at the 5′-end of flaviviral mRNA. (a) N7 methylation results in a cap 0 structure, referred to as m7GpppN-RNA. (b) When the 2′-O position on the first ribose also is methylated, a cap 1 structure is formed. This is referred to as m7GpppNm-RNA. (c) In some cases, the 2′-O position on the second ribose is also methylated, making a cap 2 structure, m7GpppNmNm-RNA.
Diversity of methyltransferases

Methyltransferases are a diverse group of proteins with a wide range of targets. There are those that methylate DNA, histones, RNA and small molecule secondary metabolites. Most of these reactions involve SAM as the methyl donor (Figure 8). Two types of specific methyltransferases are involved in methylation of the cap structure: guanine-N7-methyltransferase and nucleoside-2’-O-methyltransferase. In cellular systems, these reactions are performed by different enzymes that are assigned either methylation activity.

The NS5 protein

NS5 is the largest cleavage product in the flaviviral genome. It is located at the very 3’-end of the genome, and consists of two separate domains. The C-terminal one has been identified as an RNA-dependent RNA polymerase (RdRp), but the nature of the N-terminal domain was unclear for a long time (Rice et al., 1985; Poch et al., 1989; Koonin, 1991). Eventually, it was predicted to be a methyltransferase (Koonin, 1993). When the domain was found it was speculated that it might be responsible for both methylation events in flaviviruses (Koonin, 1993). When the structure of the predicted methyltransferase domain from Dengue virus (DENV) was determined, it was also verified that this protein indeed exhibited 2’-O-methyltransferase activity (Egloff et al., 2002). However, no N7-methyltransferase activity could be observed under the experimental conditions used. Later, the methyltransferase from West Nile virus (WNV) was shown to exhibit both 2’-O and N7-methyltransferase activity, thereby confirming the initial hypothesis (Koonin, 1993; Ray et al., 2006). At the time, it was known that the flaviviral NS3 protein harbourd RNA-triphosphatase activity, and the two methylation events needed for completion of the 5’-type 1 cap were attributed to the NS5 methyltransferase. However, the whereabouts of the guanylyltransferase remained unclear. Several studies had shown that the methyltransferase domain could bind GTP, leading to speculation that this domain could also exhibit guanylyltransferase activity (Egloff et al., 2002; Egloff et al., 2007; Bollati et al., 2009; Geiss et al., 2009). In 2009, it was finally demonstrated that the N-terminal domain of the flaviviral NS5 protein indeed has guanylyltransferase activity (Issur et al., 2009).

Development of antivirals

The diversity of the viral capping machinery, along with differences from the cellular capping process in humans, make these proteins potential targets
for antiviral drugs. During the last decade, viral RNA cap methyltransferases have become popular drug targets (Dong et al., 2008; Liu et al., 2010; Issur et al., 2011; Ferron et al., 2012). The lack of N7-methylation blocks viral RNA translation, and is therefore essential for virus replication (Ray et al., 2006; Dong et al., 2008). For this reason, N7-methyltransferase activity was considered a better target than 2’-O-methylation. However, 2’-O methylation was discovered to be connected with antiviral response (Zust et al., 2011), and now both methyltransferase activities are the focus for development of antivirals.

Figure 8. The cofactor S-adenosyl-L-methionine (SAM), and the two antiviral drugs ribavirin and sinefungin. The circle in the SAM molecule and sinefungin highlights the difference between the two molecules.

Targeting the flaviviral NS5 methyltransferase

The essentiality of the flavivirus methyltransferase (MTase) for replication is known for several species. This protein has two clear binding pockets, where GTP and SAM bind. Both these binding sites have been explored in terms of enzyme inhibition.

The GTP analogue ribavirin was one of the first compounds to be tested against the MTase$_{DENV}$. Ribavirin is a broad spectrum antiviral (Sidwell et al., 1972; Declercq, 1993). The mechanism of this compound was unknown although interference with capping had been proposed (Shatkin, 1976b; Goswami et al., 1979; Scheidel et al., 1987; Scheidel and Stollar, 1991). The first proof of action in the capping machinery came when ribavirin was
shown to inhibit the guanylyltransferase activity of the vaccinia virus capping enzyme (Bougie and Bisaillon, 2004). Shortly thereafter, ribavirin 5′-triphosphate was shown to inhibit the 2′-O-methylation activity of the MTase\textsubscript{DENV}. A crystal structure of MTase\textsubscript{DENV} in complex with SAM and ribavirin 5′-triphosphate also revealed the mode of binding (Benarroch \textit{et al.}, 2004).

Several studies have reported inhibition of the NS5 MTase with the SAM analogue sinefungin. Structures in complex with the analogue and inhibition data are available for MTases from WNV (Dong \textit{et al.}, 2008; Dong \textit{et al.}, 2010), Yellow Fever virus (Dong \textit{et al.}, 2010), DENV (Milani \textit{et al.}, 2009; Dong \textit{et al.}, 2010; Selisko \textit{et al.}, 2010) and Wesselsbron virus (Bollati \textit{et al.}, 2009; Milani \textit{et al.}, 2009). The inhibition data, in combination with results from viral titer reduction assays and inhibition in cell culture, suggest that sinefungin analogues potentially can be used in antiviral development (Liu \textit{et al.}, 2010).

In addition, virtual screens against both GTP- and SAM-binding sites have been performed (Luzhkov \textit{et al.}, 2007; Milani \textit{et al.}, 2009; Podvinec \textit{et al.}, 2010; Lim \textit{et al.}, 2011). High-throughput screening against the GTP binding site has also been developed, and in combination with an assay for guanylyltransferase activity these have identified a number of potential hits (Geiss \textit{et al.}, 2011; Stahla-Beek \textit{et al.}, 2012).

A conserved hydrophobic pocket has been identified adjacent to the SAM binding site. This pocket is critical for replication and cap methylation, and opens up the possibility to expand SAM analogues in order to increase drug specificity (Dong \textit{et al.}, 2010).
Methyltransferase domain from Modoc virus (Paper II)

Flaviviruses belonging to the no-known vector (NKV) group are not human pathogens and they are believed to have evolved separately from those in the mosquito-borne and tick-borne groups. The Modoc virus belongs to the NKV group, and infects rodents. Due to its similarity to human pathogens, but its inability to infect humans, the Modoc virus has been proposed as a model virus for studying flaviviral infection and drug discovery. To bring a piece to the puzzle of the question of the evolutionary origin of the NKV group, and to further evaluate the Modoc virus as a model virus, the structure of the Modoc virus methyltransferase was determined.

Modoc virus MTase

The original construct included the full length version of the MTase domain of Modoc virus NS5 protein (MTase\textsubscript{MODV}), residues 1-293 with an N-terminal His-tag. The protein was recombinantly expressed in \textit{E. coli} Rosetta cells, which produced soluble protein in a good yield. The protein was purified using affinity chromatography on a Ni-sepharose column followed by size exclusion chromatography. Crystallization screens were set up, but no crystals could be obtained. Based on the available structure of the Dengue virus methyltransferase (MTase\textsubscript{DENV}) together with sequences of other flaviviral MTases, a new construct was designed, with a truncation of 25 amino acids in the C-terminal. This new construct, containing residues 1-268, was expressed and purified in the same way as the full length construct. After crystallization screening and optimization, well-diffracting apo and SAM-cofactor complex crystals were obtained. Later on, crystallization screening with a cap-analogue was also set up, but none of these attempts yielded any crystals.

Data for both the apo crystals and the SAM co-crystallizations were collected at the ESRF. Although they were crystallized in the same conditions, the apo enzyme crystallized in space group I4 and the SAM co-crystals in space group \textit{P}2\textsubscript{1}2\textsubscript{1}2\textsubscript{1}. The structure of MTase\textsubscript{MODV} in complex with SAM was determined by molecular replacement using the first structure of MTase from DENV (MTase\textsubscript{DENV}) as a search model (Egloff \textit{et al.}, 2002). Electron
density was found in the expected SAM binding site and the complete molecule could be unambiguously modelled. The apo structure was subsequently determined with molecular replacement using the SAM complex structure as a search model.

The MTase\textsubscript{MODV} structure

Overall, the structure of the MTase\textsubscript{MODV} shares the same fold as the previously determined structures of flaviviral MTases (Figure 9). The fold is characterized by a core with a central seven-stranded \(\beta\)-sheet, flanked by two helices. The core domain is complemented with N- and C-terminal extensions. The N-terminal part consists of a helix-turn-helix followed by a \(\beta\)-strand and another helix. The C-terminal part consists of a helix and a strand making up a small \(\beta\)-sheet together with the strand from the N-terminal extension. A central cleft, located between strands \(\beta1\) and \(\beta4\), form the SAM binding site.

The MTase\textsubscript{MODV} structure and the previously described structures of flaviviral methyltransferases (Egloff \textit{et al.}, 2002; Assenberg \textit{et al.}, 2007; Mastrangelo \textit{et al.}, 2007; Zhou \textit{et al.}, 2007b) are similar to the \(\alpha\)-\(\beta\)-\(\alpha\) consensus fold described for SAM-dependent methyltransferases (Fauman \textit{et al.}, 1999), consisting of a central seven-stranded \(\beta\)-sheet flanked by three helices on each side. However, in contrast to the consensus fold, the flaviviral MTase structures lack one helix on each side of the central sheet.

Binding sites

Three binding sites have been identified in the flaviviral MTase structure. The methyl donor SAM binding site, a GTP/cap binding site, and a surface where presumably RNA binds. The flaviviral MTase has been associated with both methylation at the guanine-N7 position and the ribose 2\(^{\prime}\)-O. Nevertheless, only one binding site for SAM has been found, even though it is assumed to be the methyl donor in both reactions. Mutational studies in the SAM-binding site, that impair both N7 and 2\(^{\prime}\)-O-methylation also supports this idea (Ray \textit{et al.}, 2006; Zhou \textit{et al.}, 2007a; Dong \textit{et al.}, 2008). However, studies with sinefungin on the MTase\textsubscript{WNV} show a significant difference in inhibition between the two methylation events, suggesting that there are different modes of SAM binding (Dong \textit{et al.}, 2010).

The location of the identified GTP/cap-binding site only allows ribose 2\(^{\prime}\)-O methylation. To allow guanine-N7 methylation, the RNA is required to bind differently to bring the guanine closer to the methyl donor.
Figure 9. (A) MTase^MODV^ in complex with the cofactor SAM. The core domain in gray, the N-terminal extension in blue and the C-terminal extension in light red. (B) SAM binding site including the side chains interacting with the cofactor molecule.

**MTase^MODV^ SAM-binding site**

The SAM binding site is located in a central cleft between strands β1 and β4, where the adenine moiety stretches towards the loop between β3 and β4, and the tail points towards helix αA (Figure 9). In the MTase^MODV^ structure, the SAM adenine is accommodated in a hydrophobic pocket made up from Leu106, Val133 and Ile148. It is further coordinated via hydrogen bonding with the main chain nitrogen of Val133, and the side chains of Asp132 and His111. The ribose is coordinated by His111 and the amine on the tail interacts with Asp147, Ser57 and Ser87. The SAM binding site is well-conserved in all three groups of flaviviruses - the majority of the residues are conserved or exchanged for a residue with similar characteristics. All 2′-O MTases share a conserved KDKE-tetrad, responsible for catalytic activity. In the MTase^MODV^ structure, this tetrad consists of Lys62, Asp147, Lys184 and Glu220. Mutational studies of the MTase^WNV^ indicate that the entire KDKE-tetrad is essential for 2′-O MTase activity, whereas N7 MTase activity only requires the central aspartic acid (Debasish et al., 2006). Glu112 is conserved in all flaviviral MTase-SAM complexes. However, it coordinates the ribose in some of the structures, whereas it is flipped away from the ribose in others, as is the case in MTase^MODV^.

Mutagenesis studies in several flaviviral MTases have also identified residues in the binding pocket or near the binding pocket that are important for methylation activity (Ray et al., 2006; Zhou et al., 2007a; Dong et al., 2008; Kroschewski et al., 2008). The corresponding residues in MTase^MODV^ are Ser57, Trp88, His111, Asp132 and Ile148.
Figure 10. (A) Mosaic model of the MTase\textsubscript{MODV}-SAM structure together with the guanosine from the crystal structure of MTase\textsubscript{DENV}-m7GpppA (pdb id 2P41, yellow) and five bases from the structure of vaccinia virus VP39 MTase in complex with hexameric RNA (pdb id 1AV6, green). Gray dots mark the position of the triphosphate linker. (B) Surface representation of MTase\textsubscript{MODV} in the same orientation as in (A). The green patch corresponds to residues that have been shown to be important for cap binding in MTase\textsubscript{WNV}.

GTP and RNA binding sites

As soon as the first flaviviral MTase structure was determined, it was hypothesized that the N-terminal extension where a GTP analogue bound, was in fact the cap-binding site (Egloff \textit{et al.}, 2002). Several flaviviral MTases do indeed exhibit strong GTP affinity (Egloff \textit{et al.}, 2002; Bollati \textit{et al.}, 2009; Geiss \textit{et al.}, 2009). The guanine is positioned between helix A1 and A2 in the N-terminal extension from the central core. The only completely conserved residue is a phenylalanine involved in stacking interactions with the guanine, corresponding to Phe25 in the MTase\textsubscript{MODV}. Biochemical experiments show that this is a major interaction point between the cap structure and the protein (Geiss \textit{et al.}, 2009). The GTP binding site mainly seems to be involved in the 2′-O methylation event, which is supported by experiments showing that GTP inhibits 2′-O methylation. In contrast, methylation is not inhibited by ATP (Egloff \textit{et al.}, 2007).

Several structures of flaviviral MTases in complex with different cap analogues have been determined, but so far there is no structural explanation as to how specificity in the binding is achieved. The GTP binding pocket can accommodate both N7 methylated and unmethylated cap analogues.

To date, no structure of a flavivirus MTase-SAM-cap-RNA complex has been determined. However, a positively charged surface has been suggested to be involved in RNA binding, supported by extensive mutagenesis studies on the MTase\textsubscript{WNV} (Dong \textit{et al.}, 2008). By combining the structure of the
remotely related vaccinia VP39 MTase in complex with a RNA hexamer (Hodel et al., 1998) and the MTase$_{DENV}$ m7GpppGm complex (Egloff et al., 2007), a model of a MTase$_{MODV}$-SAM-m7G-RNA complex can be created (Figure 10).

The combined model indicates two positively charged residues on the edge of the elongated MTase$_{MODV}$ cavity, Arg29 and Arg215, that are likely to be involved in coordination of the cap triphosphate. The importance of the latter residue is supported by mutation studies of the corresponding Arg213 residue in the MTase$_{WNV}$ (Dong et al., 2008). MTase$_{MODV}$ Lys62, part of the KDKE tetrad, is positioned to coordinate the RNA backbone phosphate in between the flavivirus-specific AG nucleotides. Arg58 is positioned to coordinate the next phosphate. The importance of Arg58 is also supported by mutation studies of the corresponding residue in the MTase$_{WNV}$ (Dong et al., 2008). Since the MTase$_{MODV}$ Arg38-Lys47 helix has no counterpart in the vaccinia structure, the geometry of the two 3’ nucleotides is hard to predict. However, mutation of the residue corresponding to Arg38 also abolishes activity, indicating that the RNA might be bent past the MTase$_{MODV}$ N-terminal helix.

The observed pathological effect of Modoc virus in immunocompromised mice is highly similar to that of flavivirus encephalitis in humans (Leyssen et al., 2001; Leyssen et al., 2003). The structural conservation between flaviviral MTases, including MTase$_{MODV}$ from the NKV group, support the use of Modoc virus as a model virus for studying flaviviral pathology in general and the flaviviral MTase as a drug target.
Targeting isoprenoid synthesis in *Mycobacterium tuberculosis*

**Tuberculosis**

According to the World Health Organization, about one-third of the world’s population is infected with tuberculosis (TB). Although the number of new TB cases has decreased for several years, 8.7 million people were infected and 1.4 million people died from TB in 2011 (World Health Organization, 2012b).

5-10% of the people infected with TB will develop the active disease. In the rest of the cases the disease is latent and asymptomatic. However, people that are co-infected with HIV run a greater risk of developing the active disease. Tuberculosis is also the leading cause of death in this patient population. Most common is a pulmonary infection, but TB can also spread to other parts of the body. This extrapulmonary infection is also more frequent in immunocompromised patients.

Current TB treatment is a therapy over six months, combining four different drugs, called the first-line drugs. Non-compliance with this demanding therapy leads to patient relapse and increases the risk of developing drug-resistant TB strains. In recent years there has been an increasing problem with drug resistance, either through the patient being infected with a resistant strain, or emergence of bacterial resistance during the course of treatment. About 6% of the TB cases are multidrug-resistant (MDR-TB), meaning that the bacteria are resistant to at least two of the first-line drugs. About 10% of the MDR-TB cases are also resistant to two other classes of drugs apart from the first-line drugs. These cases are classified as extensively drug-resistant TB (XDR-TB).

The combination of resistant strains and a long, expensive treatment, severe side effects and problems with drug-drug interference when co-treating TB and HIV, stress that there is an urgent need for new and more efficient therapies. The currently TB drugs are 50 years old and little progress has been made in TB-treatment during this time. In 2012, the new drug Bedaquiline was granted accelerated approval by the US Food and Drug Administration for treatment of adults with MDR-TB (Jefferys, 2012). However, additional research and clinical trials are needed to validate the safety of the drug and its efficacy in children.
The only vaccine developed so far is the bacillus Calmette-Guérin (BCG), based on the *Mycobacterium bovis* strain. The BCG vaccine is not extensively used due to its limited efficiency, but it does demonstrate that it is possible to develop a TB vaccine (Brewer, 2000). Extensive efforts have been made to develop an improved vaccine, and in 2012 there were 12 new vaccine candidates in human trials (Jefferys, 2012; Tameris *et al.*, 2013).

*Mycobacterium tuberculosis*

The main causative agent for TB, *Mycobacterium tuberculosis* (*Mt*), was discovered in 1882 by Robert Koch. In 1905 he was awarded the Nobel Prize for this discovery. *Mt* is a rod shaped bacterium. It is not classified as either Gram negative or Gram positive, but rather as an acid-fast bacillus due to the ability to retain stain even after acid treatment. The bacterium is characterized by its extremely slow growth rate and complex cell wall. *Mt* divides every 16-20 hrs. By comparison, *Escherichia coli* divides every 20 minutes. The outer membrane of *Mt* is made up from mycolic acids which makes it difficult for antibiotics to permeate the cell wall.

Isoprenoid biosynthesis *via* the MEP pathway

The isoprenoids form a large and diverse group of natural compounds including more than 35,000 different molecules found in bacteria, archaea and eukaryotes. This class of molecules is essential for cell survival and the molecules are integral in a wide variety of biological functions. All isoprenoids are derived from the precursor isopentenyl pyrophosphate (IPP) or its isomer, dimethylallyl pyrophosphate (DMAPP). Examples of isoprenoids are ubiquinones, involved in electron transport, and sterols such as cholesterol that act as membrane stabilizers and precursors for vitamin D among other molecules. A number of essential isoprenoid compounds have been characterized in *Mt*. One of these is polyprenyl phosphate that is essential for the synthesis of the cell-wall components arabinogalactan and lipoarabinomannan (Wolucka *et al.*, 1994).

The isoprenoid precursors, IPP and DMAPP, can be synthesized *via* two independent pathways, the mevalonate (MVA) pathway and the methylenylthritol phosphate (MEP) pathway, also known as the non-mevalonate pathway. The MVA pathway was discovered in yeast and mammals in the 1960s (Katsuki and Bloch, 1967; Lynen, 1967). In this pathway, IPP and DMAPP are produced from acetyl-CoA *via* a mevalonate intermediate. For a long time, this was the only known pathway for synthesis of these precursors. However, in the 1990s, a new route was discovered, the methylenylthritol phosphate (MEP) pathway (Rohmer *et al.*, 1993; Rohmer *et al.*, 1996).
The MEP pathway produces IPP in a series of seven steps, starting from pyruvate and D-glyceraldehyde 3-phosphate (G3P) (Figure 11). The second step is the first committed step in the pathway. Isomerisation between IPP and DMAPP is carried out by the enzyme isopentenyl diphosphate isomerase in both pathways, and is the only enzyme shared by both the MVA and MEP routes.

The MEP pathway exists in many bacteria and protozoa and it is used by both *Mycobacterium tuberculosis* and the malaria parasite *Plasmodium falciparum*. The fact that the MEP pathway does not share any of the intermediates with the MVA pathway, nor have any homologous proteins in humans, make the MEP enzymes attractive targets for drug design.

**Figure 11.** The methylerythritol phosphate pathway for biosynthesis of the isoprenoid precursors IPP and DMAPP. DXR catalyses the second reaction, an NADPH-dependent reduction and isomerisation of DXP to MEP.
1-deoxy-D-xylulose 5-phosphate reductoisomerase

The enzyme 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR/IspC) catalyses the second step in the MEP pathway. 1-deoxy-D-xylulose 5-phosphate (DXP) is converted into 2-C-methyl-D-erythritol 4-phosphate (MEP) in an NADPH-dependent isomerisation and reduction. That this reaction was performed by DXR was first described by Kuzuyama et al. (Kuzuyama et al., 1998). The reaction also requires a divalent cation, like Mn$^{2+}$, Mg$^{2+}$ or Co$^{2+}$ (Takahashi et al., 1998). Although the biologically relevant ion is not known, it is likely to be Mg$^{2+}$ given the abundance of this ion in vivo compared to Mn$^{2+}$ and Co$^{2+}$. In both Escherichia coli (Ec) and Bacillus subtilis, DXR is essential for cell survival (Rodriguez-Concepcion et al., 2000; Kobayashi et al., 2003). DXR is also essential for M. tuberculosis growth in vitro (Brown and Parish, 2008).

The first DXR crystal structure, that of EcDXR, was solved in 2002 (Reuter et al., 2002). This structure reveals the three-dimensional fold of the active homodimer where each monomer is composed of three domains: the NADPH-binding domain, the central catalytic domain, and the C-terminal domain. The dimer interface consists of two β-sheets made up of strands from both monomers (Figure 12). To date, other DXR structures include those from Zymomonas mobilis (Ricagno et al., 2004), Mycobacterium tuberculosis (Henriksson et al., 2006), Yersinia pestis (Osipiuk et al., 2008), Thermotoga maritima (Takenoya et al., 2010) and Plasmodium falciparum (Umeda et al., 2011). These structures all share similar cofactor and substrate binding sites as well as a flexible active site flap involved in substrate binding.

Figure 12. Structure of the functional MtDXR homodimer with NADPH and fosmidomycin bound in the A-chain. The NADPH-binding domain is colored in blue, C-terminal domain in red and catalytic domain in green with the active site flap in black. The B-chain is colored in gray.
Previous work on \textit{MtDXR}

The first \textit{MtDXR} structure was solved in 2006 in our laboratory (Henriksson \textit{et al.}, 2006). This was an apo structure with a sulphate ion bound in the position where the phosphonate portion of the substrate was expected to bind, as seen in a previously determined \textit{EcDXR} structure (Mac Sweeney \textit{et al.}, 2005). Subsequently, structures were determined of \textit{MtDXR} in complex with the natural product antibiotic fosmidomycin, both with and without NADPH (Figure 12) (Henriksson \textit{et al.}, 2007). The structure of \textit{MtDXR} in complex with fosmidomycin and NADPH exhibited a well-defined active site flap, with the indole ring of Trp203 forming close interactions with the fosmidomycin backbone. Both fosmidomycin and its analogue FR-900098 have antibacterial activity against Gram-negative and some Gram-positive bacteria (Okuhara \textit{et al.}, 1980). Fosmidomycin is also a potent inhibitor of \textit{MtDXR} in enzymatic assays (Dhiman \textit{et al.}, 2005). When we tested fosmidomycin and FR-900098 against \textit{MtDXR}, the IC$_{50}$ values were determined to 0.08µM and 0.16µM, respectively (Henriksson \textit{et al.}, 2007; Andaloussi \textit{et al.}, 2011). Although fosmidomycin has been shown to inhibit both \textit{EcDXR} and \textit{MtDXR} in enzymatic assays, it had no effect on the \textit{Mycobacterium} itself, due to poor uptake (Brown and Parish, 2008). In a study by Haemers \textit{et al.}, various α-substituted fosmidomycin and FR-900098 analogues were evaluated in an \textit{EcDXR} inhibition assay and against parasite growth in two different strains of \textit{Plasmodium falciparum} (Haemers \textit{et al.}, 2006). Several of the compounds had elevated activity against the malaria parasite compared to fosmidomycin and FR-900098. The most potent compounds had a 3,4-dichlorophenyl substitution at the α-carbon (Figure 13). We evaluated these inhibitors, together with several other compounds, against \textit{MtDXR}. The 3,4-dichlorophenyl substituted compounds were also the most efficient against \textit{MtDXR}, although they were not more efficient than fosmidomycin. Structures of \textit{MtDXR} in complex with 3 and 4 (Figure 13) were determined in order to study the mode of binding (Andaloussi \textit{et al.}, 2011). These complexes were subsequently used as a starting point for design of new inhibitors.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure13.png}
\caption{Inhibitory compounds fosmidomycin, FR-900098, 3,4-dichlorophenyl-fosmidomycin and 3,4-dichlorophenyl-FR-900098 were tested against \textit{MtDXR} and their IC$_{50}$ values were determined to be 0.08µM, 0.16µM, 0.15µM and 0.7µM, respectively.}
\end{figure}
Structural and functional studies of fosmidomycin analogues as MtDXR inhibitors (Paper III)

To further develop inhibitors against MtDXR, two series of compounds analogous to fosmidomycin were synthesized by our collaborators at the Department for Medicinal Chemistry, Uppsala University, in the research group of Professors Anders Karlén and Mats Larhed. In both these series, the hydroxamate of fosmidomycin was substituted. These new compounds were tested against MtDXR in an enzymatic assay and the most potent compounds were put into crystallization trials with MtDXR to obtain crystal structures of the complexes. The compounds were also sent to collaborators at AstraZeneca, Bangalore, to be tested against Mycobacterium tuberculosis (Mt) cells and in a Plasmodium falciparum (Pf) blood cell assay. Recent structural data on PfDXR, revealing a highly conserved active site between MtDXR and PfDXR, raises the hope that inhibitors potent against MtDXR also will prove to be efficient against the Plasmodium enzyme (Umeda et al., 2011).

Structural details of fosmidomycin analogues

The first series of compounds was based on a fosmidomycin scaffold where the hydroxamate was substituted with various groups (Table 1). In the second series, the previously successful scaffold with a 3,4-dichlorophenyl substitution at the α-carbon, was expanded with a phenyl substitution at the hydroxamate. The phenyl group at the hydroxamate was further substituted at the ortho-, meta- and para-positions, although compounds with substitutions at the ortho-position were primarily explored (Table 2). By adding another ring structure to the fosmidomycin backbone, we hoped to form intramolecular stacking interactions that would stabilize the compound and improve its ability to compete with the substrate.
Table 1. Inhibition of *MtDXR* activity for mono-substituted fosmidomycin analogues. Literature values for *EcDXR* and *PfDXR* are included for comparative purposes.

<table>
<thead>
<tr>
<th>R</th>
<th><em>MtDXR</em> IC₅₀ (µM)</th>
<th><em>EcDXR</em> IC₅₀ (µM)</th>
<th><em>PfDXR</em> IC₅₀ (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.08 ± 0.02</td>
<td>0.050</td>
<td>0.032</td>
</tr>
<tr>
<td>2</td>
<td>methyl 0.16 ± 0.03</td>
<td>0.051</td>
<td>0.018</td>
</tr>
<tr>
<td>8a</td>
<td>ethyl 27.2 ± 7.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8b</td>
<td>2-furyl 48.7 ± 6.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8c</td>
<td>cyclopropyl 156 ± 68</td>
<td>10-4.44</td>
<td></td>
</tr>
<tr>
<td>8d</td>
<td>phenyl 2.0 ± 0.6</td>
<td>0.13</td>
<td>0.061</td>
</tr>
<tr>
<td>8e</td>
<td>3-pyridyl 1.6 ± 0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8f</td>
<td></td>
<td>3.6 ± 0.5</td>
<td></td>
</tr>
<tr>
<td>8g</td>
<td>2-naphthyl 2.0 ± 0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8h</td>
<td>benzyl 67.7 ± 23.3</td>
<td>13</td>
<td>n.d</td>
</tr>
<tr>
<td>8i</td>
<td>&gt;200</td>
<td>7.1</td>
<td>3.3</td>
</tr>
<tr>
<td>8j</td>
<td>&gt;200</td>
<td>1.0</td>
<td>0.4</td>
</tr>
</tbody>
</table>

*a* Henriksson *et al.*, 2007; *b* Andaloussi *et al.*, 2011; *c* Giessmann *et al.*, 2008; *d* Silber *et al.*, 2005
Table 2. Inhibition of MtDXR activity and *Plasmodium falciparum* growth by di-substituted fosmidomycin analogues.

<table>
<thead>
<tr>
<th>R&lt;sub&gt;1&lt;/sub&gt;</th>
<th>R&lt;sub&gt;2&lt;/sub&gt;</th>
<th>R&lt;sub&gt;3&lt;/sub&gt;</th>
<th>% inh at 100 μM</th>
<th>MtDXR IC&lt;sub&gt;50&lt;/sub&gt; (μM)</th>
<th>Pf growth IC&lt;sub&gt;50&lt;/sub&gt; (μM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14a</td>
<td>H</td>
<td>H</td>
<td>H</td>
<td>100</td>
<td>0.32 ± 0.05</td>
</tr>
<tr>
<td>14b</td>
<td>CH&lt;sub&gt;3&lt;/sub&gt;</td>
<td>H</td>
<td>H</td>
<td>95</td>
<td>0.83 ± 0.08</td>
</tr>
<tr>
<td>14c</td>
<td>CF&lt;sub&gt;3&lt;/sub&gt;</td>
<td>H</td>
<td>H</td>
<td>70</td>
<td>9 ± 8</td>
</tr>
<tr>
<td>14d</td>
<td>OCF&lt;sub&gt;3&lt;/sub&gt;</td>
<td>H</td>
<td>H</td>
<td>90</td>
<td>8 ± 2</td>
</tr>
<tr>
<td>14e</td>
<td></td>
<td>H</td>
<td>H</td>
<td>90</td>
<td>7 ± 3</td>
</tr>
<tr>
<td>14f</td>
<td></td>
<td>H</td>
<td>H</td>
<td>80</td>
<td>19 ± 3</td>
</tr>
<tr>
<td>14g</td>
<td></td>
<td>H</td>
<td>H</td>
<td>60</td>
<td>21 ± 4.1</td>
</tr>
<tr>
<td>14h</td>
<td>H</td>
<td></td>
<td>H</td>
<td>30</td>
<td>&gt; 100</td>
</tr>
<tr>
<td>14i</td>
<td></td>
<td>H</td>
<td>H</td>
<td>70</td>
<td>13 ± 3</td>
</tr>
<tr>
<td>14j</td>
<td>H</td>
<td></td>
<td>H</td>
<td>90</td>
<td>0.14 ± 0.04</td>
</tr>
<tr>
<td>14k</td>
<td>H</td>
<td>H</td>
<td></td>
<td>80</td>
<td>1.2 ± 0.2</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>0.08 ± 0.02</td>
<td>0.2-0.4</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>0.16 ± 0.03</td>
<td>0.32</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>0.15 ± 0.02</td>
<td>0.09</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>0.7 ± 0.1</td>
<td>0.25</td>
</tr>
</tbody>
</table>

<sup>a</sup> Henriksson *et al.*, 2007; <sup>b</sup> Andaloussi *et al.*, 2011; <sup>x</sup> Haemers *et al.*, 2006
MtDXR enzymatic assay

The inhibitory capacity of the two series of mono- and di-substituted fosmidomycin analogues was evaluated in a spectrophotometric assay in which the MtDXR-catalyzed NADPH-dependent rearrangement and reduction of DXP to form MEP was monitored at 340nm. None of the compounds in the first series improved on the IC$_{50}$ values for the lead compounds fosmidomycin and FR-900098. A few of the compounds in this series did produce relatively low IC$_{50}$ values although they were not able to reach the levels of fosmidomycin and FR-900098. These were compounds with the 3-pyridyl (8e), 3-nicotinamide (8f), phenyl (8d) and 2-naphthyl (8g) substitutions, with values in the low µM-range. This indicated that the aromatic group directly connected to the acyl might be beneficial for the binding. A few of the compounds showed a significant increase in the IC$_{50}$ value compared to fosmidomycin and FR-900098, more specifically those with ethyl (8a), 2-furyl (8b), benzyl (8h) and cyclopropyl (8c) substitutions. This can potentially be attributed to clashes with the indole ring of Trp203 in the active site flap. Compounds 8i and 8j are reasonably potent on EcDXR and PfDXR (Giessmann et al., 2008), and they were therefore resynthesized and evaluated on MtDXR. Unfortunately they had very poor IC$_{50}$ values.

The second series of compounds was based on the previously tested 3,4-dichlorophenyl substitution at the phosphonate α-carbon of fosmidomycin (Andaloussi et al., 2011). The previously determined protein-inhibitor complexes of inhibitors 3 and 4 with MtDXR showed approximate co-planarity between the hydroxamate and the dichlorophenyl substitution (Andaloussi et al., 2011). The space at the hydroxamate towards the solvent, suggested that we could combine the Cα-dichlorophenyl-substitution with the phenyl-substitution at the hydroxamate to produce a di-substituted compound, 14a. The phenyl was in turn substituted at the ortho-, meta- and para-positions. In this series several compounds produced IC$_{50}$ values very close to or below µM-levels. The unsubstituted scaffold, 14a, came up as the second most potent compound in the series with an IC$_{50}$ value of 0.32µM. Introduction of progressively larger substituents at the ortho-position of the phenyl ring resulted in a decline of MtDXR inhibitory activity. The most potent compound turned out to be the meta-substituted triazole with an IC$_{50}$ value of 0.14µM, 14j.

Crystallographic data of MtDXR-inhibitor complexes

To investigate the mode of binding, several of the inhibitors were put into crystallization trials with MtDXR. Crystal structures of four protein-inhibitor complexes were determined: two complexes with compound 14a, the scaffold for the second series, with and without NADPH, and structures of
MtDXR in complex with 14f and 14i. Unfortunately we were not able to obtain structures of the protein in complex with the meta- and para-substituted triazoles, 14j and 14k. The two complexes with compound 14a with and without NADPH, crystallized in different space groups. The former crystallized in a tetragonal space group whereas the three complex structures without NADPH all crystallized in a monoclinic space group. The protein-inhibitor complexes with compounds 14f and 14i did not have a Mn$^{2+}$ bound in the active site.

Figure 14. Comparative view of the previously solved complex of 3,4-dichlorophenyl-fosmidomycin (A) with compound 14a (B). The propylene backbone and the 3,4-dichlorophenyl are accommodated in the same way in the active site in both structures.

In all three complexes were the protein crystallized in a monoclinic space group, the crystal structure revealed inhibitors bound in a state where the two ring formations did indeed form intramolecular ring-ring interactions. (Figure 14). However, in the MtDXR-Mn$^{2+}$-14a complex the inhibitor interacted with a tryptophan from a symmetry related molecule. To verify that this did not effect the mode of binding the MtDXR-Mn$^{2+}$-14a-NADPH complex in the tetragonal space group was obtained. This complex did not exhibit any symmetry related interactions, and a superposition of the two structures revealed a nearly identical inhibitor binding.

The phosphonate group of the three inhibitors is involved in an extensive set of hydrogen bond and salt-link interactions similar to what has been observed in previously determined complexes. The conformation of inhibitor 14a is very similar to those of 3 and 4 in the previously determined protein-inhibitor complexes. The interactions between the 3,4-dichlorophenyl group and the protein, as well as the orientation of the dichlorophenyl, is practically identical. The conformation of the phenyl at the hydroxamate in 14a allow for intramolecular interactions between the two aromatic groups where
the ring planes are inclined to one another at an angle of ~50°. The other face of the ring interacts with the side chain of Met267. Interaction of the aromatic group on the hydroxamate and Met267 is also seen in the complex structure with 14f. Complexes with 14f and 14i show ring-ring interactions with a considerably smaller angle between the planes of the rings, compared to the 14a complex. The angles were ~25° and ~35° for 14f and 14i respectively.

![Figure 15. Accommodation of inhibitors 14a (A), 14f (B) and 14i (C) in the MtDXR substrate binding site. The position of the propylene backbone is slightly shifted in the three complexes as well as the relative position of the two aryl rings. However, the same amino acid residues are involved in the binding where a change in the position of the ligand is compensated by a change in the neighboring side chains. In the complexes with inhibitors 14f and 14i, where there is no Mn²⁺ in the active site: Asp151 is turned away from the metal binding site. (D) Superposition of ligands 14a, 14f and 14i reveals a difference in the relative position of the rings, where ligand 14f is closest to co-planarity and compound 14a shows the largest deviation.](image)
A superposition of the protein main chain of the protein-inhibitor structures revealed that all three compounds exhibit strikingly different binding modes, despite their chemical similarity (Figure 15). However, in each of them, a shift in the inhibitor is compensated by a shift in the interacting amino acids, meaning that a similar set of interactions is observed but their relative positions are shifted. A superposition of the propylene backbone of the inhibitors clearly shows that all three conformations are similar to each other. The difference in position can be explained in part by the lack of a metal ion in the MtDXR-14f and MtDXR-14i complexes, thereby allowing the hydroxamate to shift closer to the metal-binding site, as seen in the MtDXR-14f complex. Despite similar binding and backbone conformation of the inhibitors, the relative positions of the two ring structures to each other are different.

In the MtDXR-14f complex structure, we can see the trifluoro-methyl-anilino substituent in the ortho-position on the scaffold, stretching towards the NADPH binding site. When superimposing NADPH from the MtDXR-Mn$^{2+}$-14a-NADPH complex onto the MtDXR-14f complex structure, it becomes clear that the 14f inhibitor enters the cofactor binding site occupying the space of the nicotinamide ring, thus competing with NADPH.

**In vitro assays: Mt cells and Pf infected blood cells**

Our effort to find new inhibitory compounds against *Mycobacterium tuberculosis* was done in collaboration with AstraZeneca, Bangalore, India. The compounds were sent to AstraZeneca for testing in two different cell based assays. In the first one the compounds were tested for their ability to inhibit growth of Mt cells. The second assay, including only compounds from the second series, evaluated their potential as anti-malarial inhibitors with a Pf blood cell assay. Since the active sites of MtDXR and PfDXR are very similar to each other and their mode of fosmidomycin binding has also been confirmed to be similar, we were optimistic that these compounds against MtDXR might also be potent against PfDXR.

Unfortunately, none of the compounds had any effect in the Mt cell growth assay. However, several of the inhibitors were potent in the Pf blood cell assay, where compound 14a showed the highest level of inhibition of the compounds tested, with an IC$_{50}$ of 0.04µM (Table 2).

**Inhibition of MtDXR and antimalarial activity**

Although none of the compounds tested in the MtDXR enzymatic assay was more potent than fosmidomycin or FR-900098, a few compounds produce IC$_{50}$ values very close to or below µM-level. These include compounds 8e, 8f, 8d, 8g from the first series, and 14a, 14j and 14k from the other. The
MtDXR inhibition of 14a is poorer than that of the previously tested formyl derivative 3 but better than the acetyl derivative 4, also described earlier (Table 2). The complexes with compounds 3, 4 and 14a all show a disordered active site flap and share a great deal of similarity. More specifically, the conformation and binding of the propylene backbone, the dichlorophenyl substitution and the metal-ion coordination are highly conserved between these structures. The improved enzyme inhibition of 14a compared to 4 is attributed to the energetically more favourable interactions between the two aromatic rings seen in 14a, compared to the interaction between an aromatic group and a methyl substitution seen in 4. The sub-µM IC\textsubscript{50} values for 14i and 14k suggest that the triazole in the meta- and para-positions make specific interactions with the enzyme.

None of the compounds was effective in the Mt cell growth assay. Since the Mt cell wall is very hydrophobic due to its high content of mycolic acid, the compounds might not be able to penetrate the cell wall, as reported previously for fosmidomycin (Brown and Parish, 2008). However, the compounds with the lowest IC\textsubscript{50} values from the MtDXR enzymatic assay were also potent in the Pf blood cell assay, although not in the same relative order. The most potent compound in the Pf blood cell assay was compound 14a without any additional groups added to the aryl on the hydroxamate. In contrast, the meta substituted triazole was most potent in the MtDXR enzymatic assay. This might be attributable to differences in the active sites between MtDXR and PfDXR. It can also be due to how the compounds interact with other molecules in the infected blood cells, or how easily the compounds cross the cell membrane.

As can be seen in all three MtDXR-inhibitor crystal structures, the two aryl groups in the inhibitors form stacking interactions. The idea behind these aryl-aryl substituted compounds was that they would be stabilized by the ring-ring interaction: to this end, the design aim was met. In the complex structure of MtDXR-14f, which had a third ring structure substituted with a trifluoro group, the trifluoro moiety reaches into the NADPH binding pocket. The steric clash between the NADPH molecule and the inhibitor can potentially explain the relatively high IC\textsubscript{50} for this compound, since the assay is performed in presence of excess NADPH. This is supported by another, fluorescence based, assay indicating that 14f and 14i have comparable dissociation constants to fosmidomycin both in the presence and absence of metal-ion, and that these values are lower than the dissociation constant for 14a.

Together, these data give new insights into the binding of a promising new series of inhibitors and provide a stepping-stone for further optimization of their properties.
På senare tid har ett antal nya virussjukdomar brutit ut och bland bakteriesjukdomarna är antibiotikaresistens ett allt vanligare problem. Ett exempel på detta är tuberkulos som huvudsakligen orsakas av bakterien Mycobacterium tuberculosis (Mt). De senaste åren har man upptäckt flera Mt-varianter som är motståndskraftiga mot de läkemedel som används i det första behandlingssteget. Nu finns även varianter som är resistenta mot den antibiotika som används i flera av de senare behandlingsstegen. Under 2011 hade denna extremt resistenta variant av tuberkulosbakterien identifierats i 77 länder.

I början av 2003 spred sig virussjukdomen SARS snabbt över världen och infekterade över 8000 människor och av dessa avled nästan 800. Inom kort stod det klart att SARS-viruset var ett tidigare okänt virus tillhörande coronavirusfamiljen. Detta var första gången som ett liknande virus hade visat sig orsaka allvarlig sjukdom hos människor. SARS-viruset visade sig dessutom vara väldigt smittsam och spreds till ett stort antal länder inom bara några veckor.


För att studera den tredimensionella strukturen hos de här proteinerna har vi använt oss av s.k. röntgenkristallografi. Metoden innebär att man först renar fram det protein som man är intresserad av att studera och sedan hittar betingelser under vilka proteinet kan bilda välordnade kristaller, liknande tex saltkristaller. Genom att utsätta kristallerna för oerhört kraftig och fokuserad röntgenstrålning kan man registrera information som gör det möjligt att beräkna proteinets tredimensionella struktur. Eftersom ett proteins struktur är nära förknippat med dess funktion, kan informationen sedan användas för att ge detaljerade insikter i vad proteinet gör och hur det görs. Strukturen kan också ge oss information om vilka små kemiska föreningar som proteinet interagerar med och på vilket sätt dessa föreningar binder. Den här informa-
tionen kan sedan användas för att hitta liknande små molekyler, så kallade inhibitorer, som kan hindra proteinet från att utföra dess funktion.

I det ena projektet som presenteras i den här avhandlingen har vi utvecklat föreningar som inhiberar proteinet 1-deoxy-D-xylulos 5-fosfat reduktoisomeras (DXR) från *Mycobacterium tuberculosis*. DXR är involverat i biosyntesen av s.k. isoprenoider som är nödvändiga små byggestenar i alla livsförmåger. Genom att blockera aktiviteten hos DXR kan vi alltså stoppa tillverkningen av de här byggestenarna. Isoprenoider syntetiseras på helt olika sätt i människa och i mikroorganismer, som tex tuberkulosbakterien och parasiten som orsakar malaria. Detta gör DXR attraktivt för utveckling av läkemedel. Vi har utvärderat en mängd små molekyler med avseende på deras förmåga att inhibera funktionen hos DXR. Några av dessa föreningar har också studerats med röntgenkristallografi för att ta reda på vilket sätt de binder till proteinet. Den här kunskapen kan sedan komma att användas för att designa nya molekyler som förhoppningsvis har ännu bättre inhibitoriska egenskaper.


Acknowledgements

An amazing journey has come to its end. Along the way, I have had the pleasure to meet a lot of people to whom I would like to give thanks.

My greatest thanks to Alwyn Jones for so many things: for caring support during these years, for the confidence to give me free hands on the virus work, and for that one squash game that nearly killed me. I have finally recovered enough for a rematch... Torsten Unge, for the protein engineering course where it all started, enthusiasm beyond compare and for shared conviction that behind every corner lures a new breakthrough. Sherry Mow-bray, for constructive discussions, especially on PjDXR.

I would like to express my great appreciation to all the people that in many different ways contributed to this thesis. There is no work like teamwork! A special thanks to Christofer Björkelid for joint efforts on the DXR project; assay work, ligand building, data collection and for always taking the time – it has been a true pleasure. To Terese Bergfors, for outstanding crystallization expertise, brutal honesty feedback and for truly caring. Sanjeeewani Sooriyaarachichi for binding experiments on DXR. Thanks to our collaborators at the Department for Medicinal Chemistry, Uppsala University in the group of Anders Karlén and Mats Larhed: Anna Więckowska, Suresh Surisetti, Samir Yahiaoui, Martin Lindh, Shyamraj Dharavath, Surisetti Suresh, Mounir Andaloussi, Matthieu Desroses, Johan Gising and Anneli Nordqvist. And to our collaborators at AstraZeneca in Bangalore, India. Thanks to Rautela Nikhil, Sharma Sreevalli and Bachally Srinivasa. Thanks also to Eric Snijder at Leiden University Medical Center, The Netherlands, for taking an interest in the TGEV nsp1 project and for generously sharing your knowledge on coronaviruses.

Being a grad student is not only about your own science, but also about passing on knowledge. I want to thank Maria Selmer and Lars Liljas for making my teaching experiences great ones. Maria especially for shared teaching ambitions and the pedagogical development project and Lars for appreciated co-teaching and many discussions about life and science.
Thanks to Johan Unge for beamline support at MAXlab and for generously giving your time. Thanks to everyone in the administration: Sigrid Pettersson, Lena Jansson, Erika Anundsson, Sofia Sundbaum, Anders Eriksson, Åsa Klefbom and Ali Elfadil, for fixing with payments, papers and signatures with smiles on your faces, and for taking the ICM Christmas party to a new level.

This experience would not have been the same without great colleagues and friends at the department. My greatest thanks to Emma Jakobsson for being my initial mentor at xray. A brief collaboration that turned into a lasting friendship. Evalena Andersson, for true friendship and nerdy discussions dealing equally with new gym exercises, data processing and things to try in the lab. Henrik Keränen, for many mind-clearing runs, talks in the sauna and for being family. Kitos. Christofer Björkelid, for the snoring that brought us together and for making sure that I have a beer every now and then. Yang “sötnos” Chen, for Chinese-Swedish cooking exchange, company at Monday yoga and for coming with me to Jægtvolden. Alina Castell for doing the right thing at the right time and for that first trip to the ESRF. Annette Roos for top notch office sharing and for climbing. Mikael “Nisse” Nilsson for appreciation of nerdy sports stuff and the conviction that Excel is good for everything. Lena Henriksson for story-telling and for generously sharing your knowledge. Daniel Ericsson for making me (finally) watch Star Wars. Magnus Persson for quote of the day and for not being upset about the night I cleaned your lab bench. Nina Bäckbro for keeping order in the lab and for rescuing my knitting. Lotta Avesson, for sharing the boyz. Glareh Askarieh for the biggest smile and the biggest laugh no matter what. Avinash Punekar and Wojciech Krajewski for engaged scientific discussions. Christoffer Lind and Micke Gudmundsson for fikas and lunch company.

And to so many people, past and present, for making this a fantastic place to work: Martin Andér, Inger Andersson, Lars Boukhatra, Jens Carlsson, Adrian Suarez Covarrubias, Anna Suarez Larsson, Sara Dodd, Hans Eklund, Xiaohu Guo, Hugo Gutierrez de Teran, Janos Hajdu, Henrik Hansson, Mark Harris, Andrea Hinas, Yafei Huang, Gösta Hult, Henrik Ingvarsson, Wangshu Jiang, Yasmin Kahn, Masoud Kazemi, Gerard Kleywegt, Stefan Knight, Cha San Koh, Lu Lu, Seved Löwgren, Jenny Mackenzie, Al Mackenzie, Jie Nan, Agata Naworyta, Maria Pränting, Mats Sandgren, AnaLaura Stern, Johan Sund, Jerry Stålberg, Fredrik Söderbom, Wimal Ubhayasekera, Ulla Uhlin, Erling Wikman, Miao Wu, Shiyang Wu, Li Zhang and Johan Åqvist.
Although it might sometimes feel like it, life isn’t only about work and science. Luckily, I have fantastic friends and family to remind me. For that I am truly grateful.

Niklas Johansson, for being a wonderful friend – you are my platonic special someone. Sofia Nilsson, for so many good times at home and on the road – Ulla-resan, backcountry skiing, sessions with ‘Janzzon & Nilzzzon’ and creative ideas for every situation. Anna Nordström, for making me appreciate 5.30-morning runs, for breakfast afterwards and the amazing Låktatjäkka trail tour. Veera Kainiemi, for climbing, listening and for introducing me to Ressu Redford. My ultra-running friends, Wanja Strid and Louise Bäckelin, for inspiration and for showing that long distance-running can be combined with having a family, demanding full-time job and a long commute.

The extended Johansson family, for making a Stockholm girl feel welcome in the heart of the north. For fabulous dinners, fishing, berry picking and perfect preparation of “åkytan”.

Thanks that really cannot be put into words to my family. To mamma och pappa for always being there and for limitless support in whatever endeavours I embark on. To Magnus, best brother ever. For being my complete opposite, for perspective and for being the counter balance I need. For hilarious text messages that I save for rainy days. Karine, for coming to our family. Now there will finally be time for us to get to know each other better. Mormor, jag vet att du tänker på mig. Tack för att du lär mig sånt som verkligen spelar roll, som hur man bäst staplar ved och bakom vilken knix i bäcken som fisken kommer att nappa.

Patrik. For endless patience and for waiting. For always being right next to me no matter how far away. For loving support, it has been truly invaluable.
References


Henriksson, L. M., Björkeld, C., Mowbray, S. L. and Unge, T. (2006) The 1.9 A resolution structure of Mycobacterium tuberculosis 1-deoxy-D-xylulose 5-


55


56


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.