

**Structural-functional studies
of chorismate synthase and
glutamyl-tRNA synthetase
from *M. tuberculosis***

Vom Promotionsausschuss der
Technischen Universität Hamburg-Harburg
zur Erlangung des akademischen Grades
Doktor-Ingenieur
genehmigte Dissertation

von
Marc Bruning
aus
Lohne
2008

1. Gutachter: Prof. Dr. rer. nat. A. Liese
2. Gutachter: Prof. Dr. rer. nat. A. Torda

Prüfungsausschussvorsitzender: Prof. Dr. rer.nat. Dr. h.c. G. Antranikian

Tag der mündlichen Prüfung: 02.11.2007

Acknowledgement

This work presents the results obtained during the time of my PhD from January 2004 to July 2007 in the research group of Dr. H. D. Bartunik at the Max-Planck Unit for Structural Molecular Biology and supervised by Prof. Dr. A. Liese at the Hamburg-Harburg University of Technology. I would like to sincerely thank both Prof. Liese and Dr. Bartunik for accepting me in their research and PhD programs. I am more than grateful for their guidance, support, encouragement and for the proposal of the interesting and important topic of this thesis. Both Dr. Bartunik and Prof. Liese provided me the best possible framework for successfully conducting my research.

I thank Prof. Dr. Torda for taking part in the evaluation of my thesis and for additional reporting. I would also like to express my gratitude towards Prof. Dr. Antranikian for chairing the doctoral committee.

I would like to thank all my colleagues in the research group of Dr. Bartunik at the Max-Planck Unit for Structural Molecular Biology. In this respect I would especially like to thank Dr. Gleb Bourenkov and Dr. Galina Kachalova for their invaluable help, contributions and guidance in the field of structural biology during all stages of my thesis. Furthermore, I would like to thank Dr. Attila Oberschall, Dr. Nicolai Strizhov, Mrs. Susanne Meier and Ms. Birgitta Fried for their help, contributions and guidance in the wet-lab. Finally, I would like to thank all the administrative staff of the Max-Planck Unit of Structural Molecular Biology and here especially Mrs. Elke Spader and Mrs. Katja Schröder for providing the perfect administration that allowed me to concentrate on my research. Similarly, I would like to thank the secretary of Prof. Dr. Liese Mrs. Ulrike Zimmermann for her outstanding help and care in all administrative issues at the Hamburg University of Technology. I would also like to thank all students and co-workers in the research department of Prof. Dr. Liese for the nice atmosphere, interesting and helpful discussions and their generally good humoured spirit.

Last but not least my gratitude and thankfulness extends to my entire family and especially my parents for their help and support. This work is dedicated to them.

List of abbreviations

%	Percent
α	Alpha
β	Beta
γ	Gamma
σ	Sigma
μ l	Microlitre
$^{\circ}$	Degree
$^{\circ}$ C	Degree celsius
$2mF_{\text{O}}-DF_{\text{C}}$	Weighted electron density map
3D	Three dimensional
A	Adenine
Å	Angström
AA	Amino acid
AaCS	<i>Aquifex aeolicus</i> chorismate synthase
AARS	Aminoacyl-tRNA synthetase
AB	Anticodon binding
AcO	Acetoxy
ADSC	Area Detector Systems Corporation
AdT	Amidotransferase
AM1	Austin Model1
AMBER	Assisted Model Building and Energy Refinement
AMP	Adenosine monophosphate
Arg	Arginine
ATP	Adenosine triphosphate
BCC	Bond charge correction
Br	Bromine
BW6	Wiggler Beamline 6
C	Carbon
C	Cytosine
C α	Alpha carbon
Ca	Calcium
CCD	Charge-coupled device
CCP4	Collaborative Computational Project Number 4
CDC	Centres for Disease Control and Prevention
CHARMM	Chemistry at Harvard Macromolecular Mechanics
CjCS	<i>Campylobacter jejuni</i> chorismate synthase
Cl	Chloride
CP	Connective peptide
CS	Chorismate synthase
CS2	Dimeric MD simulations of SpCS
CS2 _a	Dimeric MD simulation of SpCS with liganded monomer A
CS2 _b	Dimeric MD simulation of SpCS with liganded monomer B
CS4 _{ox}	Tetrameric MD simulation of SpCS with FMN _{ox}
CS4 _{red}	Tetrameric MD simulation of SpCS with FMN _{red}
DESY	Deutsches Elektronensynchrotron
D-GluRS	Discriminating GluRS
DHQ	Dehydroquinate
DHQase	Dehydroquinase synthase
DNA	Desoxyribonucleic acid
EBI	European Bioinformatics Institute
EcCS	<i>Escherichia coli</i> chorismate synthase
EcGluRS	<i>Escherichia coli</i> glutamyl-tRNA synthetase
EM	Electron microscopy
EPSP	5-enoylpyruvyl-shikimate phosphate
ESP	Electrostatic potential
ESRF	European Synchrotron Radiation Facility
F	Fluor
F _C	Calculated structure factor
F _{hkl}	Structure factor

Fig	Figure
FMN	Flavin mononucleotide
FMN _{ox}	Oxidized flavin mononucleotide
FMN _{red}	Reduced flavin mononucleotide
F _O	Observed structure factor
FPLC	Fast protein liquid chromatography
fs	Femtosecond
G	Guanine
GAFF	General Amber Force Field
GlnRS	Glutamyl-tRNA synthetase
GlxRS	Glutamyl- or glutamyl-tRNA synthetase
GROMACS	Groningen Machine for Chemical Simulations
GROMOS	Groningen Molecular Simulation
H	Hydrogen
HIV	Human immunodeficiency virus
HpCS	<i>Helicobacter pylori</i> chorismate synthase
I	Iodine
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K	Kelvin
K _d	Dissociation constant
kJ	Kilojoule
L	Litre
L1	Loop 1
L4	Loop 4
L8	Loop 8
L20	Loop 20
L22	Loop 22
L25	Loop 25
L27	Loop 27
LB	Lysogeny broth
M	Molarity
Mar	MarResearch
MD	Molecular dynamics
MDR-TB	Multi-drug resistant tuberculosis
mF _O -DF _C	Weighted difference electron density map
Mg	Magnesium
mg	Milligram
ml	Millilitre
MM	Molecular mechanics
MPG	Max-Planck Gesellschaft
MR	Molecular replacement
mRNA	Messenger ribonucleic acid
MTB	<i>Mycobacterium tuberculosis</i>
MtCS	<i>M. tuberculosis</i> chorismate synthase
MtGluRS	<i>M. tuberculosis</i> glutamyl-tRNA synthetase
N	Nitrogen
N	Nitrogen
Na	Sodium
NADPH	Nicotinamide adenine dinucleotide phosphate
NCA	Nicotinamide
NcCS	<i>Neurospora crassa</i> chorismate synthase
ND-GluRS	Non-discriminating GluRS
NH ₂	Amino group
NH ₄ ⁺	Ammonium
Ni	Nickel
nl	Nanolitre
ns	Nanosecond
NTA	Nitrilotriacetic acid
O	Oxygen
OD	Optical density
OPLS	Optimized Potential for Liquid Simulations
P	Phosphor

PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PO ₄ ³⁻	Phosphate
POA	Pyrazinoic acid
PP _i	Pyrophosphate
ps	Picosecond
PZA	Pyrazinamide
PZAse	Pyrazinamidase/nicotinamidase
QM	Quantum mechanics
RF	Rossmann fold
RMS	Root mean square
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RNA	Ribonucleic acid
S	Sulfur
S3P	Shikimate-3-phosphate
SC	Stem contact
ScCS	<i>Saccharomyces cerevisiae</i> chorismate synthase
SDS	Sodium dodecyl sulphate
SfCS	<i>Shigella flexneri</i> chorismate synthase
SO ₄ ²⁻	Sulfate
SpCS	<i>S. pneumoniae</i> chorismate synthase
TB	Tuberculosis
TDR	Programme for Research and Training in Tropical Diseases
TeGluRS	<i>Thermosynechococcus elongatus</i> glutamyl-tRNA synthetase
Tris	2-amino-2-hydroxymethyl-1,3-propanediol
tRNA	Transfer ribonucleic acid
tRNA ^{Asn}	Asparaginyl-tRNA
tRNA ^{Asp}	Aspartyl-tRNA
tRNA ^{Gln}	Glutaminyl-tRNA
tRNA ^{Glu}	Glutamyl-tRNA
tRNA ^{Glx}	Glutaminyl- or glutamyl-tRNA
TtGluRS	<i>T. thermophilus</i> glutamyl-tRNA synthetase
U	Uridine
WHO	World Health Organisation
XDR-TB	Extremely drug resistant tuberculosis
XMTB	The German <i>Mycobacterium tuberculosis</i> Structural Genomics Consortium

Amino acid codes

- * G - Glycine (Gly)
- * P - Proline (Pro)
- * A - Alanine (Ala)
- * V - Valine (Val)
- * L - Leucine (Leu)
- * I - Isoleucine (Ile)
- * M - Methionine (Met)
- * C - Cysteine (Cys)
- * F - Phenylalanine (Phe)
- * Y - Tyrosine (Tyr)
- * W - Tryptophan (Trp)
- * H - Histidine (His)
- * K - Lysine (Lys)
- * R - Arginine (Arg)
- * Q - Glutamine (Gln)
- * N - Asparagine (Asn)
- * E - Glutamic Acid (Glu)
- * D - Aspartic Acid (Asp)
- * S - Serine (Ser)
- * T - Threonine (Thr)

Table of contents

List of abbreviations.....	4
Amino acid codes	6
1. Motivation.....	8
1.1. Introduction	8
1.2. Aims of thesis	11
2. Theoretical background.....	13
2.1. Protein crystallization	13
2.2. X-ray diffraction and electron density	16
2.3. Phasing, Model building and refinement	18
2.4. Molecular mechanics and molecular dynamics	22
3. Materials and methods	27
3.1. Cloning, expression and purification.....	27
3.2. Crystallization	29
3.3. X-ray data collection and processing.....	31
3.4. Phasing, Model building and refinement	33
3.5. Structural analysis.....	35
4. Chorismate synthase from <i>M. tuberculosis</i>	36
4.1. The shikimate pathway.....	36
4.2. Structural analysis	40
4.2.1. Model quality	40
4.2.2. Overall structure.....	41
4.2.3. FMN binding in the binary MtCS complex.....	44
4.2.4. Binding of nicotinamide to MtCS	47
4.2.5. Interim summary	49
4.3. Implications for the enzymatic mechanism.....	50
4.3.1. Active site switches during sequential ligand binding.....	50
4.3.2. Molecular dynamics simulations of CS.....	55
4.3.3. Interim summary	69
4.4. Discussion.....	69
5. Glytanyl-tRNA synthetase from <i>M. tuberculosis</i>	74
5.1. Role of AARSs in protein synthesis.....	74
5.2. Relaxed specificity and non-discriminating GluRS	79
5.3. Structural analysis	82
5.3.1. Model quality	82
5.3.2. Overall structure.....	83
5.3.3. Interim summary	86
5.4. Enzyme-tRNA interactions	87
5.4.1. Docking models of GluRS and tRNA.....	87
5.4.2. Anticodon recognition.....	90
5.4.3. MD simulations of the modelled GluRS-tRNA complexes	93
5.4.4. Interim summary	103
5.5. Discussion.....	104
6. Conclusions and outlook	110
References	111

1. Motivation

1.1. Introduction

Tuberculosis (TB) is a deadly infectious disease caused by the rod shaped, gram-positive aerobic or facultative anaerobic bacterium *Mycobacterium tuberculosis* (Fig. 1-1). Although TB can occur in any anatomical site, such as the central nervous system, lymphatic system, circulatory system etc, the most common form is pulmonary TB, where the lungs are affected by the tubercle bacillus. In this region the bacilli are ingested by alveolar macrophages. Usually the immune system response prevents the development of TB and contains the disease causing agent. However, about 10% of all infected persons eventually develop TB with a considerably higher risk for persons with immunosuppression, especially from an HIV infection [CDC 2003]. Progression from TB infection to TB disease usually occurs by multiplication of bacilli overcoming the defence of the immune system. If untreated, the death rate for these active TB cases is more than 50% [Onyebujoh and Rook, 2004]. TB cases are present in the entire world (Fig. 1-2) and the largest incidence is observed in the regions of south-east Asia and sub-Saharan Africa. In total about one third of the earth's population is infected with TB according to estimates by the World Health Organisation (WHO)¹ and each year, there are about eight million new cases, and more than two million people die from TB. Due to the synergy with HIV, the highest mortality has been reported for the African region.

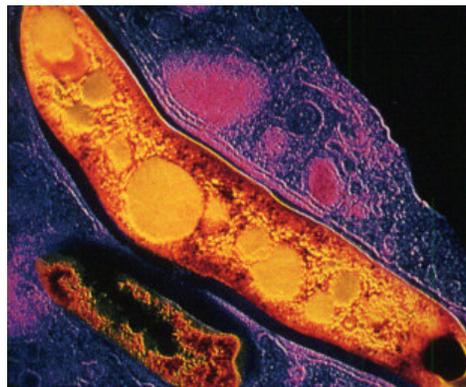


Figure 1-1: The rod shaped gram-positive aerobic or facultative anaerobic bacterium *M. tuberculosis* (taken from <http://microbewiki.kenyon.edu>).

¹ <http://www.who.int/tb/en/>

In every country a resistance of *M. tuberculosis* strains towards a single drug has been observed. Furthermore multi-drug resistant TB (MDR-TB) has been reported on the basis of a resistance towards at least the two most powerful TB drugs isoniazid and rifampicin. Resistance towards all major TB drugs has emerged [TDR 2004], ultimately resulting in the description of the so-called extremely drug resistant TB (XDR-TB) in early 2006. XDR-TB leaves patients virtually untreatable with all currently available anti-TB drugs. This situation and the fact that only one new drug (Rifapentine®) has been developed since 1972 and introduced on the market in 1998 has led to a renewed interest into MTB research. Therefore the WHO and other organisations are supporting several initiatives, aiming to find new avenues of drug development by the pharmaceutical industry, such as the Global Alliance for TB Drug Development².

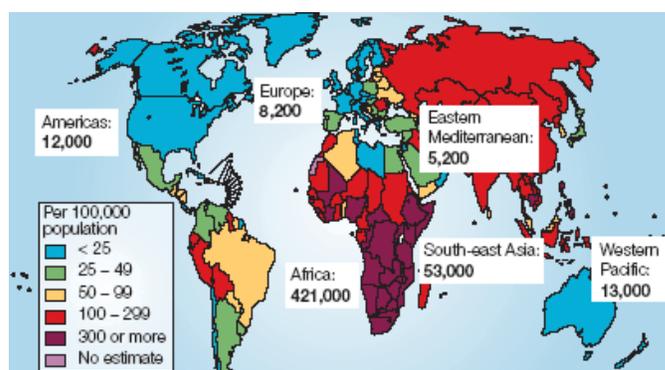


Figure 1-2: Estimated worldwide TB incidence rates in 2000 estimated by the WHO. Furthermore, estimated incidences of HIV in TB cases are shown for each region. (Taken from <http://www.who.int>)

Due to the great need for effective, new drugs for TB, structural proteomics approaches are of growing importance for drug development in this area. With the complete genome of the best-characterized strain of *M. tuberculosis* H37Rv becoming available [Cole ST et al., 1998; Camus JT et al., 2002], hopes were raised for a better understanding of the pathogen and the development of new therapeutic interventions for TB. The German Mycobacterium Tuberculosis Structural Proteomics Project (XMTB)³ uses the knowledge of the complete genome of *M. tuberculosis* and integrates a range of proteomics-oriented key technologies. This project integrates

² <http://www.tballiance.org/>

³ <http://xmtb.org/start.html>

both academic and industrial partners with the aim to find new drugs against TB based on the structural knowledge of the protein machinery that drives the mycobacterium. For this approach target proteins were selected in view of potential relevance for possible new therapies. The three dimensional structures of enzymes selected as targets will be solved by X-ray crystallography in order to gain further insight into mechanisms employed by *M. tuberculosis*.

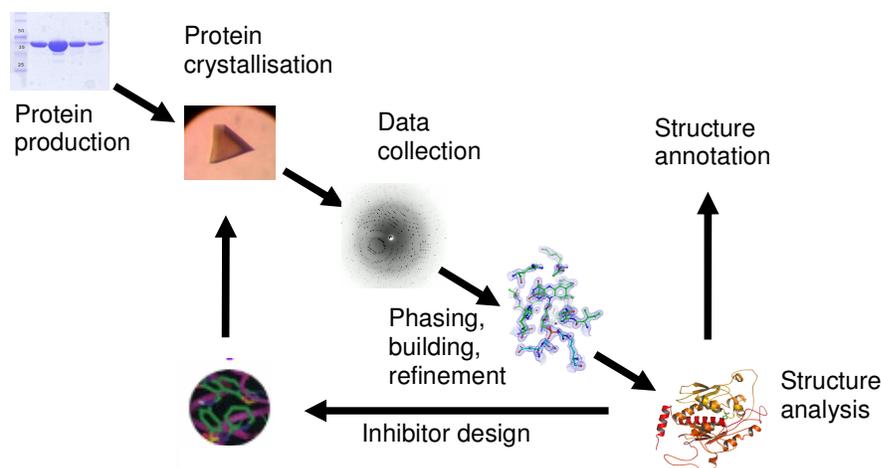


Figure 1-3: The structure solution cycle for a target protein encompasses the stages of protein production; crystallization; data collection; phasing, model building, refinement; and structure analysis. Structural data are made available to the scientific community and can be used for inhibitor design. Results from inhibitor design studies may be fed back in the structure solution cycle

This work was carried out in the framework of the XMTB project and presents the structural analysis of two important metabolic systems of *M. tuberculosis*. For each system a representative enzyme was chosen and its structure solved by X-ray crystallography, using methods as illustrated in Figure 1-3. To complement the structural analysis molecular dynamics (MD) simulations were employed. In the following chapters an introduction into the field is followed by application studies on chorismate synthase and glytanyl-tRNA synthetase from *M. tuberculosis* that are highly relevant for structure based development of novel therapeutic agents against TB. Chorismate synthase belongs to the shikimate pathway that leads to the production of chorismate, which is required for the biosynthesis of aromatic α -amino acids and a number of further aromatic compounds [Herrmann KM and Weaver LM, 2000]. The shikimate pathway is essential in mycobacteria [Stoker and Parish, 2002] and its absence from mammals makes it a potential target for the design of new

therapeutic agents against tuberculosis (TB). Glutamyl-tRNA synthetase belongs to the aminoacyl-tRNA synthetases (AARSs) that are essential components of protein translation. They provide the correctly charged tRNAs that are directly used in protein biosynthesis at the ribosomes. Inhibition of the AARSs causes a halting of the protein chain elongation that in turn results in the attenuation of bacterial growth under both in vitro and infectious conditions [Tao J et al. 2000]. Hence, these enzymes are interesting antimycobacterial drug targets.

Structures of the two enzymes were solved at high resolution and the solved structures were functionally analysed. Both structures allowed identifying important structural aspects that lead to novel mechanistic implications for the two enzymes. These new aspects were further validated using molecular dynamics (MD) methods leading to a more complete picture of the function of the enzymes. Finally, for chorismate synthase a small molecule lead is presented for the future use in rational drug design efforts. While both enzymes show a distinctly different enzymatic mechanism their joint presentation in this work is of high interest in a structural context. The structural studies on both enzymes show how rather small structural features such as single amino acids strongly influence the enzymatic mechanism on a macroscopic level. Therefore the data presented in this work is of high interest for the guidance of rational drug design studies aiming at novel anti-mycobacterial compounds.

1.2. Aims of thesis

Subsuming the central points of the motivation for this work, the essential aspects presented in this thesis are:

- The two enzymes from *Mycobacterium tuberculosis* chorismate synthase (MtCS) and glutamyl-tRNA synthetase (MtGluRS) are interesting drug targets for the design of novel drugs against the Tuberculosis disease.
- The employment of molecular biology methods for successful expression of MtCS and MtGluRS and the subsequent purification of highly pure protein at high concentration in a homogeneous state for the use in crystallization trials.

- The generation of protein crystals of MtCS and MtGluRS plus the application of protein X-ray crystallographic methods for the collection of high resolution data sets of the two enzymes in their apo form and, if possible, in complex with their natural substrates.
- Determination of the three dimensional structure of the enzymes MtCS and MtGluRS alone and, if applicable, in complex with their natural substrates using the molecular replacement method.
- Functional analysis of the different structures for novel insights into the enzymatic mechanism of the two proteins, based on the comparison of the differently liganded structures from *M. tuberculosis* and/or comparison with homologue enzymes from other organisms.
- Application of molecular modelling methods including molecular dynamics for a profound validation of the functional analysis made with the help of the three dimensional models from MtCS and MtGluRS and their comparison among each other and with their homologue counterparts.
- A comprehensive presentation of the structural-functional analysis for MtCS and MtGluRS plus the public availability of the newly determined enzyme structures should provide the basis for future drug design efforts in the important area of Tuberculosis treatment.

2. Theoretical background

2.1. Protein crystallization

X-ray crystallography is a technique with a necessity for highly pure protein at high concentration in a homogeneous state. This is important for the growth of crystals that are used to determine protein structures. While it is possible to generate protein samples from the host organism directly, i.e. from plants, animals or microbes, the most promising and frequently used approach utilizes gene recombination techniques (Fig. 2-1). In genetic recombination the gene coding for the protein of interest is amplified from genomic DNA of the host organism by use of the polymerase chain reaction (PCR) with the help of sequence specific oligo-nucleotide primers. This guarantees high amounts of the sought gene product for the subsequent incorporation into a proper vector-host combination. A typical expression system using *E. coli* and pET24b as host and vector, respectively, allows the expression of high amounts of protein in a short time.

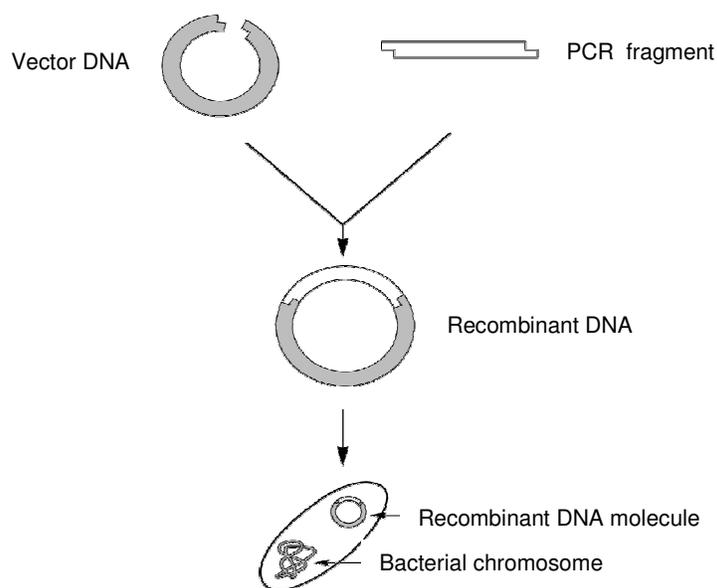


Figure 2-1: Schematic representation of the cloning procedure used in this work. A PCR fragment amplified from genomic DNA is inserted in a suitable vector and the resulting recombinant DNA is transformed into the bacterial host (adapted from <http://www.ornl.gov/>).

The produced recombinant proteins are usually purified using chromatography techniques. For easy separation the sought protein is expressed with an affinity tag. The use of a Hexahistidine- or simply His-tag, consisting of six successive either N- or C-terminally attached histidine residues, allows the convenient use of Ni-NTA-agarose columns for purification (Fig. 2-2). The histidine-tags bind with micromolar affinity to the metal ion and non-specifically bound proteins can easily be separated by washing of the used column. Subsequently the specifically bound protein of interest is eluted from the column by the use of high concentrations of imidazole. This affinity purification is typically followed by gel filtration to get rid of imidazole and other unnecessary buffer compounds used during purification and to finally generate salt free protein samples for crystallization. The highly efficient production and purification of large amounts of soluble protein is a corner stone of the success of structural genomics projects. While protein expression and purification steps are mostly routine steps in the work flow of molecular biology today, the amount of time and care that goes into preparation of purified protein samples of sufficient concentration (e.g. 10 – 50 mg/ml) can not be overestimated.

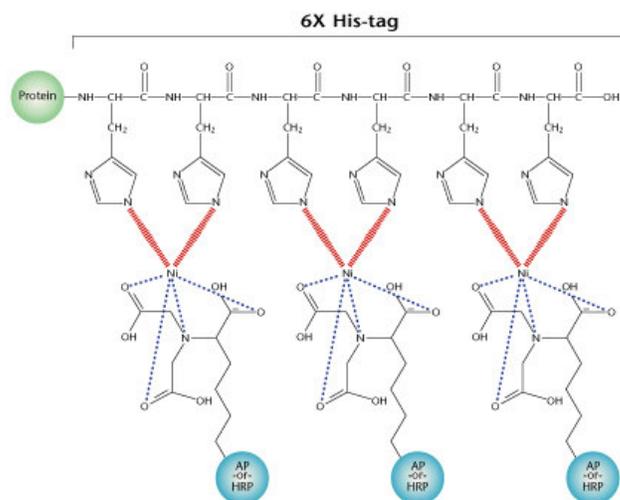


Figure 2-2: Affinity His-tag purification. The N- or C-terminally attached six consecutive histidine residues bind with micromolar affinity to the Ni-NTA resin of an affinity column (taken from <http://www.kpl.com>)

Another time consuming step prior to X-ray crystallographic studies of a given protein is the generation of suitable crystals. It is one of the most crucial steps, as not all proteins that can be obtained in amounts sufficient for use in crystallisation trials, are

amenable to this task. Crystallisation is a multifactor process that is affected by protein purity, concentration, temperature, pH, precipitants, and additives etc. This process is difficult to control and generally consists of the laborious screening of up to thousand or even more different crystallisation conditions.

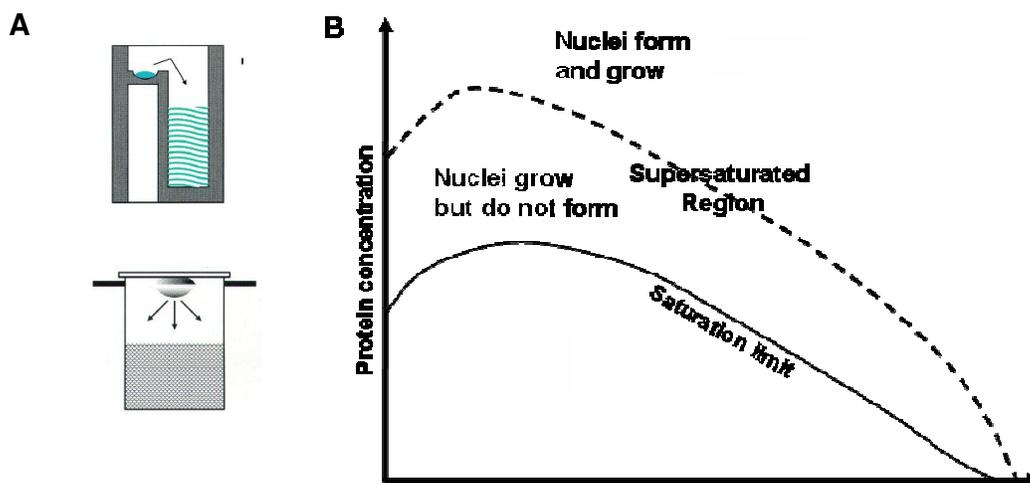


Figure 2-3: Vapour diffusion crystallization: A. Sitting drop (top) and hanging drop (bottom) techniques. B. Crystal formation in the supersaturated region is dependent on protein and precipitant concentration. Crystals grow best if they are brought as slowly as possible only slightly above the protein's saturation limit.

Crystallization occurs when the protein exceeds its limit of solubility and therefore is in a supersaturated state. The commonly used crystallization technique to achieve this state is vapour diffusion in a sitting or hanging drop approach (Fig. 2-3a). In either case a drop of protein solution is mixed in a sealed container with one of the crystallization conditions under trial and put next to a reservoir further containing the undiluted crystallization condition. Due to the concentration differences between the protein and reservoir solutions, vapour diffusion occurs and the protein concentration is driven towards its solubility limit (Fig. 2-3b). While purity of the protein sample is the most important prerequisite, the crystallization process always needs the presence of precipitants, as well. Water transfer from protein to reservoir solution raises the concentration of the precipitant to a value optimal for protein crystallization. If this process proceeds in a manner slow enough to allow the protein molecules with the help of the precipitant to accept a well ordered three-dimensional arrangement, eventually a protein crystal might occur (Fig. 2-3b).

Sometimes crystals are too small to be used directly for X-ray measurements. In this case it is possible to use them as seeds to grow larger crystals. While the setup is as described above additionally the seeds are added to the protein solution. Here the protein in solution can go directly into the small crystals and the crystal may grow to a size suitable for X-ray measurements.

2.2. X-ray diffraction and electron density

Structural information of proteins on the three-dimensional (3D) level can be gained by different experimental techniques such as nuclear magnetic resonance (NMR) spectroscopy, cryo-electron microscopy (cryo-EM) and X-ray crystallography. X-ray crystallography is currently the most powerful tool for 3D-structure determination of proteins in atomic detail.

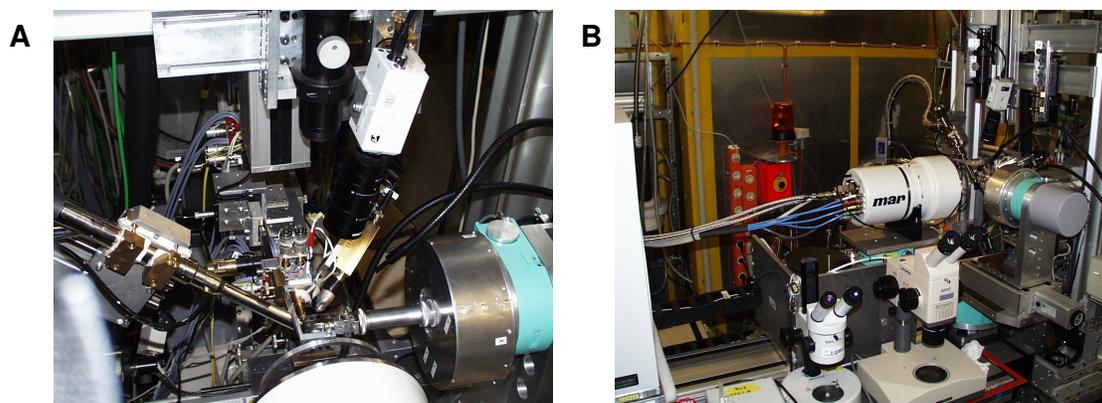


Figure 2-4: BW6 at the DESY in Hamburg. A. Detail view from the top on the goniometer head. B. Detail view of the BW6 hut at DESY. The MarCCD detector is seen in the centre.

X-rays are electromagnetic waves with photon energies in the range of 100 eV to 100 keV. In this work, the wiggler beamline BW6 of the Max-Planck Gesellschaft (MPG) located on the premises of the Deutsche Elektronensynchrotron (DESY) in Hamburg was used as an X-ray source (Fig. 2-4). BW6 is dedicated to protein crystallography and generates an X-ray beam in the energy range between 6 and 21 keV that is tunable to wave-lengths between 0.6 and 3.1 Å. Protein X-ray crystallography utilizes elastic scattering from a protein crystal placed in the path of an incident X-ray beam. Emitted X-rays from the crystal will have the same frequency and wave-length as the incident X-ray beam and will produce a diffraction pattern. Due to the periodic

arrangement of the atoms in a crystal, a distinct pattern of constructive and destructive interference emerges from the scattered waves (Fig. 2-5). This pattern can be recorded by the help various detectors, including X-ray sensitive films, image plates or CCD cameras (see also Fig. 2-4).

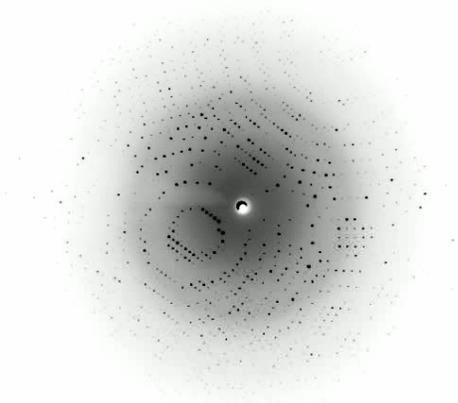


Figure 2-5: Diffraction from a protein crystal

Constructive interference of scattered X-ray photons results in the appearance of reflections in certain directions determined by the Laue conditions of diffraction or the Bragg equation. A reflection — identified by its Miller indices h , k and l — can be characterised by the structure factor (F_{hkl}):

$$F_{hkl} = \sum_{j=1}^n f_j e^{2\pi i(hx_j + ky_j + lz_j)}$$

In this equation f_j is the scattering factor of the atom j with the fractional coordinates x , y and z . F_{hkl} can be calculated from the electron density distribution (ρ) of the unit cell of the crystal by Fourier transformation. The Fourier transform describes precisely the relation-ship between the object (i.e. the protein in the crystal) and its diffraction pattern. Hence, it is possible to convert a Fourier-series description of the measured reflections to one that gives a description of the electron-density of the molecules in the crystal, allowing the formulation of an alternative description of F_{hkl} :

$$F_{hkl} = \iiint_{xyz} \rho(x, y, z) e^{2\pi i(hx + ky + lz)} dx dy dz$$

Since the Fourier transform operation is reversible, the electron density $\rho(x, y, z)$ can be calculated as a Fourier transform of the structure factors F_{hkl} , too:

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

2.3. Phasing, Model building and refinement

The electron density $\rho(x,y,z)$ can be obtained by constructing a Fourier series from the structure factors F_{hkl} that can be written as a complex vector:

$$F_{hkl} = |F_{hkl}| e^{i\alpha_{hkl}}$$

The modulus $|F_{hkl}|$ can be determined from the measured intensities:

$$F_{hkl}^2 \sim I_{hkl}$$

Missing from the measurement are the phases α_{hkl} of the diffracted X-rays. These can not be recorded in a diffraction image.

In this work initial estimates of the phases for the structure solution process were successfully obtained by molecular replacement (MR). The MR technique is used to determine the orientation and position of the target protein in the unit cell using a known atomic search model. In order for the MR procedure to be successful the search and target molecules must have reasonable sequence identity (i.e. > 25 %). As the number of already determined protein structures with different folds rapidly increased in the last couple of years, MR has become one of the most effective techniques for macromolecular structure determination. This method has the advantage over other techniques that it requires the acquisition of only one native dataset with an average redundancy and resolution. In MR, obtaining a preliminary solution of the target structure involves a correct orientation and positioning of the

search molecule in the target unit cell. For this purpose a rotation matrix [R] and a translation vector T are sought and applied to the search model (A) to get a solution for the target protein (A') (see also Fig. 2-6):

$$A' = A[R] + T$$

The necessary six-dimensional search for MR is simplified by doing the rotation and translation searches separately. First the rotational search is performed by obtaining the rotation matrix [R] from the rotation function:

$$R = \int P_T(u) P_S([R]u) du$$

Here $P_T(u)$ is the target Patterson map and $P_S([R]u)$ is the search models Patterson map rotated by the matrix [R]. Patterson maps can be calculated solely from the amplitudes and self-vectors are shifted to the origin. Maximum overlap between the Pattersons is obtained when the rotation applied to the search Patterson corresponds to the difference in orientation between the molecules. The rotation function should be dominated by self-vectors and therefore the integral is calculated for a shell in Patterson space that excludes the Patterson origin peak and cross-vectors from neighbouring molecules. For the translation search the translation function measures overlap of the cross-vectors:

$$T(t) = \int P_C(u, t) P_O(u) du$$

In this equation $P_C(u, t)$ is the cross-Patterson of the re-oriented search molecule translated along the unit cell and $P_O(u)$ is the observed Patterson function of the target structure.

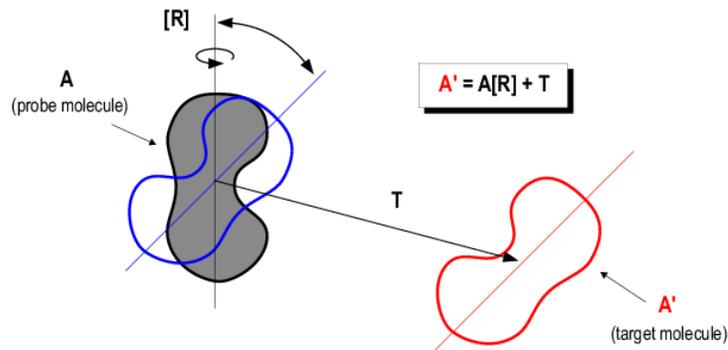


Figure 2-6: Molecular replacement technique: The search molecule A is positioned in the unit cell of the target molecule A' by application of a rotation matrix [R] and a translation vector T. (Taken from <http://www-cryst.bioc.cam.ac.uk/>)

After successful MR electron density maps are calculated by the help of Fourier synthesis:

$$\rho(x, y, z) = \frac{1}{V} \sum \sum \sum |n|F_o| - |F_c| e^{-2\pi i(hx+ky+lz-\alpha'_{calc})}$$

While $|F_o|$ are the observed structure factor amplitudes from the diffraction data, $|F_c|$ and α_{calc} are the structure factor amplitudes and phases from the MR solution, respectively. In this work manual correction of the model with help of electron density maps is interspersed by automated refinement based on maximum likelihood. Maximum likelihood measures the quality of the model based on its consistency with the observations. This is done with respect to the probability that the measurements would be made if the model were correct. For this the likelihood of the model is calculated as the product of all probabilities of the individual observations, based on the assumption that all observations are independent:

$$L = \prod_i P(F_{i,o}; F_{i,c})$$

$P(F_{i,o}; F_{i,c})$ is the probability distribution of the structure factor $F_{i,o}$ given the model structure factor $F_{i,c}$. Hence, if the model is improved the probability goes up. In

maximum likelihood refinement, both errors in the model and the measured observations are taken into account. This is especially useful in the early stages of refinement, where the model is incomplete. Generally atomic coordinates and temperature factors are refined together. A big problem in structure determination is that protein crystals diffract poorly and therefore exhibit a low ratio of observation to parameters. Considering further experimental errors the model is not well determined by the reflection data. To avoid overfitting of the data, prior knowledge of the model geometry is imposed on the refinement process in form of a restraint and/or constraint. Typical restraints are bond lengths, bond angles and van der Waals distances, as these are well known from small molecule crystallography and/or theoretical calculations and are believed to be true in protein structures as well. Constraints are used to reduce the number of parameters to be refined. For example only torsion angles can be refined instead of all x, y, z coordinates.

The progress of refinement is measured by the R-factor. After each round of refinement the new F_{calc} s are compared to the F_{obs} by calculating the R-factor:

$$R = \frac{\sum \left| |F_{obs}| - |F_{calc}| \right|}{\sum |F_{obs}|}$$

The R-factor converges as F_{calc} s should get closer to F_{obs} in the course of the refinement. Protein models with a resolution around 2 Å for example should reach an R-factor of 0.2 in the final stage of the refinement. In contrast R-factors at the beginning can be around 0.5 or higher indicating very poor agreement between F_{calc} s and F_{obs} . To avoid an overfitting of the data, 5-10% of all observed reflections are put into a test set. Reflections in this R_{free} -set are not used during the refinement but are thought for cross-validation. Thus, calculating the R_{free} -factor controls how well the current model is able to predict observations that are not used in the refinement; the deviation of R_{free} from R gives thus a measure of overfitting. It has been shown that R_{free} correlates very well with phase accuracy and therefore it is believed that it gives a better measure of the overall model. Usually R_{free} -factors are higher than R-factors, especially at the beginning of the refinement process.

The entire model building and refinement process involves repeated rounds of manual intervention and automatic model improvement. This is repeated in an excessive effort to finally give the best model that can be built. There is no clear end to this procedure and as it has been put: “90% of the effort goes into establishing the last 10% of the structure” [Read RJ, 2005].

2.4. Molecular mechanics and molecular dynamics

While proteins are generally highly flexible molecules, crystal structures only represent the average of all conformations of the protein present in the crystal. Here the protein molecules reside in an aqueous environment and exhibit reduced flexibility due to the tight crystalline lattice. This is displayed by the atomic B-factors that display how much atoms of a crystal structure fluctuate about their middle position. Further the flexibility of the protein molecules in a crystal is displayed by conformational changes upon ligand binding in soaking experiments [reviewed e.g. by Hassel AM et al., 2007]. To get further insight into the dynamical behaviour of proteins based on their X-ray structure theoretical methods may be employed. A common theoretical approach to study the behaviour of a protein in solution, based on an experimentally determined crystal structure, is the usage of molecular dynamics (MD) simulations. The most exact theoretical description of a molecular system is given by quantum mechanical (QM) methods, as these include electronic interactions between all particles of the studied system. MD studies of proteins instead usually rely on molecular mechanics (MM) descriptions as generally biological macromolecules are not feasible to QM methods due to their size. However QM methods are used as well for small systems [Gogonea V et al., 2001] and for bigger system the option exists to mix MM and QM descriptions in the so-called QMMM methods [Friesner RA and Guallar V., 2005].

MM methods only incorporate nuclear positions in their energy calculations ignoring any electronic aspect of the studied system. This description is based on the Born-Oppenheimer approximation that assumes that the electronic motions can be separated from the nuclear motions. MM methods only need a fraction of the computer time of QM methods, while the exactness of the MM results can rival higher

level of theory calculations. Force fields in MM describe the potential energy (V) of a system based on four different terms:

$$E_{Total} = \sum_{\text{bonds}} K_r (r - r_{eq})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} (1 + \cos(n\varphi - \gamma)) + \sum_{i < j} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right)$$

This equation presents the general model of the original Amber forcefield [Cornell WD et al., 1996] which is the basis for the molecular dynamics calculations in this work. The total potential energy (E_{Total}) is dependent on the positions of the atoms in the system. The four main contributions to the potential are from atomic interactions due to bonds (orange), angles (green), dihedrals (blue) and from non-bonded atomic interactions (red) (Fig. 2-7) in the system. The total potential energy of the system is the total sum of these different interactions from all interacting atoms. The first three terms (orange, green and blue) in the potential energy sum describe the bonded interactions. Here K_r and K_θ are the force constants for the bond and bond angles, respectively; r and θ are the bond lengths and angles; r_{eq} and θ_{eq} are the equilibrium bond length and angle; φ is the dihedral angle (in a 1-4 atomic interaction) and V_n is the corresponding force constant; γ is the phase angle with either 0° or 180° [Duan Y et al., 2003]. The fourth term in the force field equation (red) describes the non-bonded interactions between two atoms i and j at an interatomic distance R_{ij} by van der Waals (A_{ij}) and London (B_{ij}) dispersion terms plus interactions between partial atomic charges (q_i and q_j); ϵ is the dielectric constant of the medium (usually 1 with explicitly represented solvent) [Duan Y et al., 2003]. Extra force field terms might be included accounting for improper torsions (i.e. torsion angles where the four participating atoms are not bonded in the sequence A-B-C-D) or out-of-plane bending motions (i.e. keeping sp^2 hybridized atoms from going out of plane). Further cross terms might be included where the above named four main force field contributions are mixed, like in stretch-stretch, stretch-bend or bend-torsion motions. Before a potential energy function becomes a force field, it must be supplied with parameters for the molecule of interest. The parameterisation of a force field is often based on QM calculations of representative molecules that are similar to the studied system.

For example, the alanine dipeptide is the standard model system for the generation of protein backbone parameters (supplemented with data from glycine- and proline-dipeptides) [Mackerell AD, 2004]. Further, experimental data e.g. from crystal structures is used in the parameterization process as well. Usually partial atomic charges for a force field are determined based on electrostatic potentials (ESP) obtained also by QM methods [e.g. Singh UC and Kollman PA 1984]. The most commonly used force fields for protein and/or nucleic acid simulations are: CHARMM [Olafson BD et al., 1983], AMBER [Cornell WD et al., 1995], GROMOS [van Gunsteren WF, 1987] and OPLS [Jorgensen WL and Tirado-Rives J, 1988]. Often the choice of one of these force fields relies more on personal taste than on superiority over the others. The MM description in this work uses the AMBER94 [Cornell WD et al., 1995] and the AMBER03 [Duan Y et al., 2003] forcefields. The potential energy model of the two force fields corresponds to the equation given above.

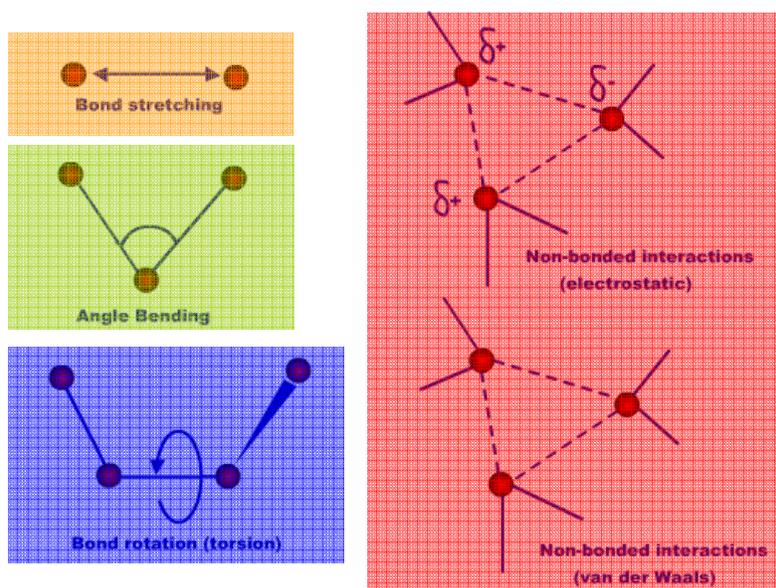


Figure 2-7: Molecular mechanics and atomic interactions contributing to the potential energy. Bonded and non-bonded interactions are shown on the left and right side, respectively. Further cross terms might exist, like in stretch-stretch, stretch-bend or bend-torsion motions (taken from <http://folding.stanford.edu>).

In this work the GROMACS package [Lindahl E et al., 2001] is used for all MD simulations and therefore the following description of MM based calculation methods refers to their implementation in the GROMACS package. Before conducting MD simulations with the chosen potential, the studied system is usually first energy

minimized. This is done in order to avoid a crash of the simulation due to inaccuracies in the starting model. Such inaccuracies might be atoms that are closer than their van der Waals distances. These would cause an unfavourably high potential energy that might dominate the system and hence compromise the entire simulation. Therefore the system is first brought to an energetic minimum in order to have a reasonable starting conformation. The first derivative of the potential energy function with respect to the coordinates is the gradient and in this work the steepest descend method is used for minimization to find the nearest local minimum by systematically moving down the steepest local gradient. The method is iterative and stops based on a user specified criterion, e.g. the number of iterations. The steepest descend approach is a very robust minimization method.

After generating a good starting configuration, the dynamical behaviour of the system under study can be followed over time in an MD simulation by solving Newton's equations of motion for all N atoms of the system:

$$m_i \frac{\partial^2 r_i}{\partial t^2} = F_i, i = 1 \dots N$$

The force F is the negative gradient of the potential V :

$$F_i = - \frac{\partial V}{\partial r_i}$$

These equations are solved in small time steps and coordinates are stored as a function of time in a trajectory that represents the dynamical behaviour of the system. In GROMACS the equations of motion are integrated by the so-called leap-frog algorithm. The leap-frog algorithm uses positions r at time t and velocities v (the first derivative of the coordinates with respect to time) at a time $t - \Delta t/2$; it updates positions and velocities with the forces $F(t)$ calculated at time t :

$$v\left(t + \frac{\Delta t}{2}\right) = v\left(t - \frac{\Delta t}{2}\right) + \frac{F(t)}{m} \Delta t$$

$$r(t + \Delta t) = r(t) + v\left(t + \frac{\Delta t}{2}\right) \Delta t$$

Hence the name leap-frog, as this calculation is somehow reminiscent of velocities and positions leaping over each others backs (Fig. 2-8). As at the beginning only coordinates are known, initial velocities are usually chosen from a Maxwell distribution and forces are supplied by the force field. The time step is chosen very small in all-atom MD simulations of proteins, as it should be an order of magnitude smaller than the highest frequency motion in the simulated system. These motions are generally bonds to hydrogen atoms (e.g. the C-H bond vibrates with a periodicity of 10 fs). To be able to chose a time step larger than 1 fs modifications are made to the equations of motion. These account for constraints imposed on the system that remove the high frequency motions. Further modifications to the equations of motion are made to keep pressure and temperature in the system constant. This work uses the NPT ensemble for MD simulations, where the number of particles N, the pressure P and the temperature T are kept constant throughout the simulation. Using the described methods, the system is simulated for a given number of steps specified in the simulation setup. Coordinates and energies are saved at specified time intervals and can later be used to follow the dynamics of the system and calculate several macroscopic properties of it that may complement the rather static crystallographic analysis.

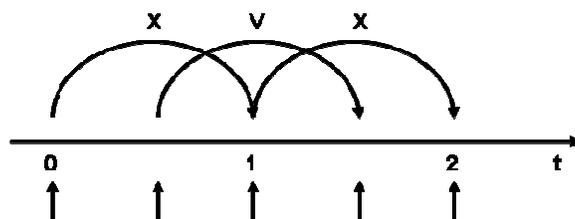


Figure 2-8: The leap-frog algorithm as implemented in the GROMACS package. The positions x and velocities v are leaping over each others backs along the time coordinate

3. Materials and methods

3.1. Cloning, expression and purification

Both the gene for *M. tuberculosis* chorismate synthase (MtCS) and *M. tuberculosis* glutamyl-tRNA synthetase (MtGluRS) were amplified by PCR from genomic DNA of *M. tuberculosis* H37Rv using primer pairs specifically designed for the 5'- and 3'-ends of the respective gene. The gene products for MtGluRS and MtCS from PCR were separately inserted between the NdeI and XhoI restriction sites of a pET-24b vector (Novagene) and thus contained a C-terminal 6xHis-extension.

Table 3-1: Composition of purification buffers. All buffers were used at pH 7.0

Buffer	A1	B1	A3	A4
Tris	50mM	50mM	50mM	50mM
NaCl	500mM	1M	500mM	-----
Imidazol	50mM	50mM	500mM	-----

The cloned genes for MtGluRS and MtCS were used for transformation of competent *E. coli* cells. The transformation procedure was similar for both clones. Frozen aliquots of *E. coli* cells were gently thawed on ice and 1 μ l of plasmid DNA (~100 ng/ μ l) containing the cloned gene was added. Cells were incubated for 30 min on ice and afterwards for 45 seconds at 42°C. Following this heat shock the cells were incubated another 2 minutes on ice after which 250 μ l of SOC medium were added to the cells. For recovery and expression of the antibiotic marker the cells were then incubated at 37°C for 1 hour. Finally the transformed cells were transferred on Agar plates containing 50 μ g/ml of Kanamycin and plates were stored at 37°C for 12 to 16 hours to allow growth of antibiotic resistant transformed colonies. Transformation was carried out twice. First DH5 α cells were transformed for selective production of high amounts of plasmid DNA. For plasmid extraction one colony was picked from an Agar plates and was transferred to a small shaker flask with 100-200 ml LB medium for overnight culture at 37°C. Cells were spun down by centrifugation and the plasmid DNA was extracted from the pellet using the MidiPrep protocol and kit from Qiagen. Subsequently the extracted plasmid DNA was used for transformation of

Rosetta2(DE3) cells. Glycerol stocks were produced both of transformed DH5 α and Rosetta cells and stored at -75°C for later use. All cells were supplied by Novagen.

Expression of both MtGluRS and MtCS was carried out in *E. coli* Rosetta2(DE3) cells at 37°C. The same protocol was used for expression and purification of both proteins. For cell culture 8 L LB medium were prepared and sterilized in shaker flasks and the medium was inoculated at 37°C. The inoculum was cultivated overnight at 32°C on a rotary shaker. The cells in the pre-induction phase were grown up to an OD₆₀₀ of 0.6. Subsequently expression was induced with 0.7 mM 1-thio- β -D-galactopyranoside (IPTG) and the cells are grown overnight at 20°C on shakers set at 180 – 200 rpm. Cells were collected by centrifugation and stored at -75 °C until further use. A full culture of 8 L usually yielded about 20 - 30 g of wet cell mass. For protein purification the harvested cells were thawed and resuspended in lysis buffer (1:2 w/v) containing 50 mM Tris (pH 7.0) and 500 mM NaCl (i.e. buffer A1, see Table 3-1). Cells were lysed via sonication by 0.35s bursts for 5 minutes on ice (Branson Digital sonifier Model 250 – D). The supernatant was centrifuged at 15,000g for 60 min. Samples were taken after every sonication step before and after centrifugation. Later the samples were checked on SDS gels (Fig. 3-1a: MtCS lanes 5-9; Fig3-1b: MtGluRS lanes 12-16). These gels clearly showed that most of the sought protein was present in the supernatant of a sonication step after the centrifugation. The filtered, clear cell lysate containing His-tagged protein from three cycles of sonication was then purified using affinity chromatography on an FPLC ÄktaXpress device from Amersham. The ÄktaXpress system was used for automation of the purification process, thereby effectively reducing manual intervention steps. The protocol developed for this purpose is described in the following text. The composition of the mentioned buffers is found in Table 3-1. First the lysate was loaded on a 1 ml Ni-Agarose HisTrap HP affinity column from Amersham Biosciences. Subsequently the column was washed with buffer A1, followed by buffer B1. Buffer B1 contained 1M NaCl for elution of unspecifically bound protein. All buffers except for A4 already contained at least 50 mM imidazol to prevent unspecific binding events to the Ni-column during the purification process. Efficiency of the washing step was checked by SDS-page (Fig. 3-1a: MtCS lanes 10-13; Fig3-1b: MtGluRS lanes 17-19). After the washing procedure affinity elution of His-tagged protein was carried out with buffer A3. Protein was eluted by application of a step gradient. This was done in order to store the peak

fractions in a loop system incorporated in the ÄktaXpress. Subsequently, the peak fractions were concentrated with buffer A4 on a Hiloal Superdex 75 gel filtration column (Amersham Biosciences). Eluted fractions were collected in a 96 deep-well plate and checked by SDS-PAGE for efficiency of the purification process and integrity of the protein (Fig. 3-1a: MtCS lanes 1-4; Fig3-1b: MtGluRS lanes 1-11). All pure MtCS or MtGluRS fractions were finally concentrated by ultrafiltration against a membrane with a 10 kDa cutoff. Final concentrations of the proteins were 8 - 20 mg/ml and 200 μ l aliquots were frozen at -75°C until their use in crystallization.

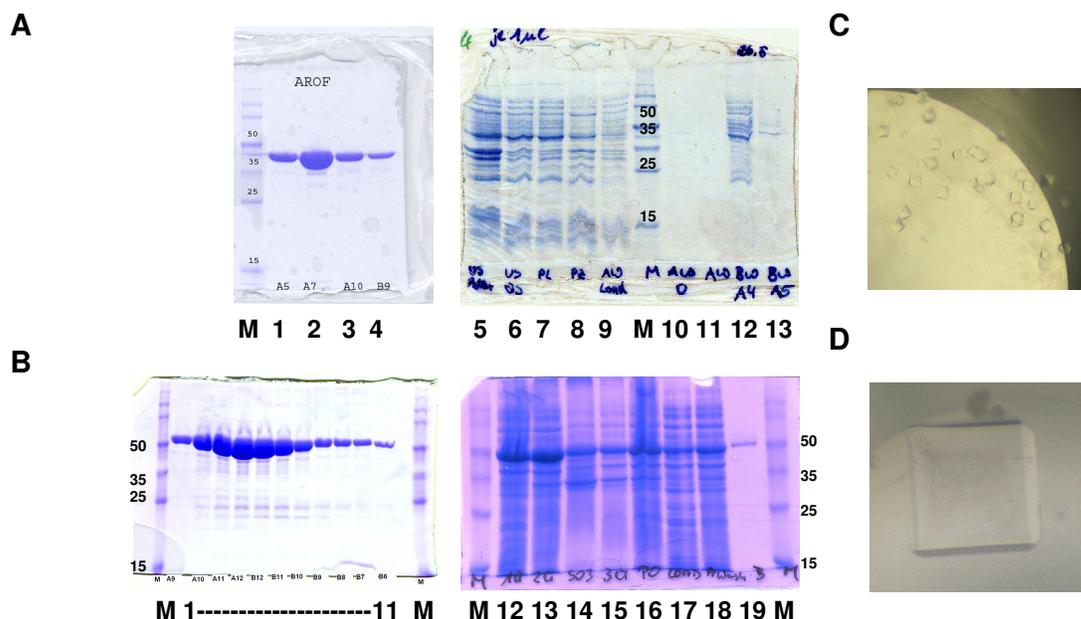


Figure 3-1: Purification and crystallisation results for MtCS and MtGluRS: A. MtCS gel filtration (Lanes 1-4); lysis (Lanes 5-9); and wash fractions (Lanes 10-13). B. MtGluRS gel filtration (Lanes 1-11); lysis (Lanes 12-16); and wash fractions (Lanes 17-19). C. MtCS crystals grown in 0.1 M Na-Acetate (pH 4.0) and 3.9 M NH₄-Acetate at 18°C. D. MtGluRS crystals grown in 10% (w/v) PEG 8000, 200 mM Ca(AcO)₂, and 100mM imidazole (pH 7.0) at 4°C. M indicates the molecular weight marker.

3.2. Crystallization

Crystallization was carried out for both MtGluRS and MtCS using the protocol described in the following text. Initial crystal screening was carried out using a sparse matrix grid consisting of 960 conditions from the following Nextal kits marketed by Qiagen: AmSO₄ Suite, Anions Suite, Classics Suite, Classics Lite Suite, Cryos Suite, MPDs Suite, PEGs Suite, pHClear Suite, and pHClear II Suite. Protein concentrations of 6 - 8 mg/ml for MtCS and 13 mg/ml for MtGluRS were used for all crystallization trials in 96-micro-well plates (Fig. 3-2, red). Crystallization was carried

out with the 96-channel Hydra-PP system (Fig. 3-2) with an additionally attached NanoFill microsolenoid dispenser (Fig. 3-2, blue). The robotic system was developed in-house by integrating commercial equipment and special hardware and software tailored by our group. The developed system uses an automatic routine for setting up one 96-microwell plate as follows:

1.) A volume of 20 μl of a single crystallization condition is applied to the reservoir container of the micro-well plate by transferring 96 conditions simultaneously from a manually prepared deep-well plate using the 96-channel head of the Hydra-system (Fig. 3-2, orange).

2.) In a subsequent step 400 nl of each conditions is transferred as described above from the reservoir well into the smaller protein well.

3.) Finally the NanoFill dispenser is used to apply 400 nl of the protein solution to the protein well, where it mixes with the previously applied crystallization condition solution.

Microplates prepared in the described way were sealed using of each set of 96 conditions one plate was stored at 4°C and one at 18°C, resulting in a total amount of 20 plates per used protein concentration.

Plates were regularly checked for the occurrence of a crystal using a light microscope. Crystal growth for MtCS occurred within one month at 18°C in a condition containing 0.1 M Na-Acetate pH 4.6 and 4.5 M NH_4^+ -Acetate. Crystals grown by the sparse matrix method diffracted to 2.8 Å and to get better diffracting crystals a grid was constructed around the initially positive screening condition. The manually constructed regular grid contained 0.1 M Na-Acetate in a pH range from 4.0 - 5.4 and NH_4^+ -Acetate in concentrations varying from 3.9-5.0 M. In total 96 conditions were constructed in a 12 x 8 matrix, where the molarity of NH_4^+ -Acetate was changed in steps of 0.1 units and the pH value in steps of 0.2 units. New crystallization trials with the customized grid were carried out with the Hydra robot as described above. Within several days crystals grew in several conditions of the grid (Fig. 3-1c). Crystals from the grid diffracted between 1.5 and 1.7 Å resolution and

complete datasets could be collected from a single crystal. Crystals grown at pH 4.0 contained nicotinamide (NCA) in the active site, while crystals grown at pH 5.4 contained apo MtCS. For soaking experiments with FMN, apo and NCA-containing crystals were harvested from a protein well of a crystallization plate using a cryo-loop and transferred to a container holding 10 μl of the respective crystallization condition with a concentration of 100 mM FMN. Crystals were soaked for up to 5 hours.

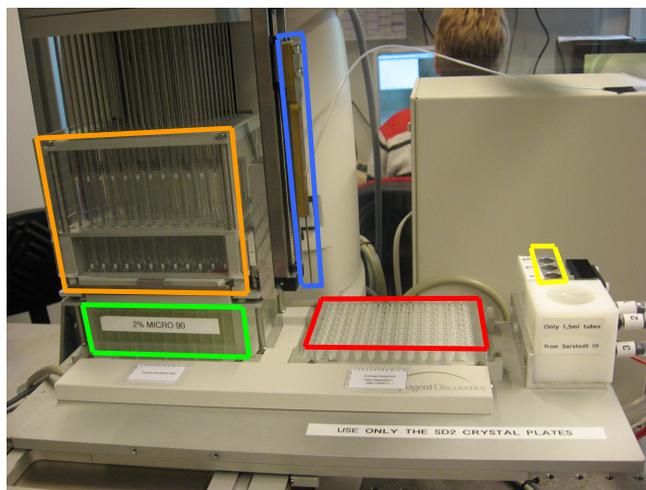


Figure 3-2: Hydra crystallisation robotics: 96-channel head (orange); 96-deep-well plate (green); 96-micro-well crystallization plate (red); NanoFill microsolenoid protein drop dispenser; and protein solution holder (yellow). For performance details see text.

In contrast, no refinement of the crystallization conditions was necessary in the case of MtGluRS. Crystallization trials were carried out with the sparse screen as described above and crystals were obtained at 4°C in a condition containing 10% (w/v) PEG 8000, 200 mM $\text{Ca}(\text{AcO})_2$, and 100mM imidazole (pH 7.0). Crystals grew over a period of a few weeks (Fig. 3-1d). These crystals diffracted up to 1.6 Å and no further refinement of the crystallization conditions was carried out.

3.3. X-ray data collection and processing

Complete datasets were collected for both MtCS and MtGluRS from single crystals. No extra cryo-protectant was necessary for data collection of MtCS as the crystallisation conditions generally contained high amounts of salt and no ice rings

were observed in test measurements. In contrast, prior to diffraction data collection, MtGluRS crystals were transferred into a solution containing 20% PEG400 as a cryo-protectant. Crystals were mounted with a cryo-loop onto the goniometer head for subsequent X-ray measurements. The crystals were cooled in a nitrogen stream at 100K to reduce radiation damage during the experiment.

Data sets from crystals containing MtCS in the apo, NCA- or FMN-bound state were collected at 100K and an X-ray wavelength of 1.05 Å on a MarCCD detector on the beamline BW6 of Deutsches Elektronen Synchrotron (DESY) in Hamburg. Additionally a data set of MtCS containing NCA in the active site was collected at 100K and an X-ray wavelength of 0.9756 Å using an ADSC-Q4 detector on the beamline ID14-2 at the ESRF in Grenoble. Due to the higher resolution, the dataset from the ESRF was used for refinement of the MtCS-NCA complex, as it made possible the unequivocal identification of the bound NCA due to very well defined electron density. Similarly diffraction data for MtGluRS were measured at 100 K and an X-ray wavelength 0.9756 Å using an ADSC-Q4 detector on the beamline ID14-2 at the ESRF in Grenoble.

The MtGluRS data were indexed, integrated, and scaled with the HKL program suite [Otwinowski Z and Minor W, 1996]. A complete data set was obtained at 1.65 Å resolution. The crystals belong to the space group $P2_12_12_1$ with unit-cell dimensions $a = 50.1$ Å, $b = 61.5$ Å, $c = 169.5$ Å. They contain one GluRS molecule in the asymmetric unit. The MtCS data sets were indexed and integrated with DENZO [Otwinowski Z and Minor W, 1996] and scaled with SCALEPACK [Otwinowski Z and Minor W, 1996] using standard protocols. Complete data sets were collected for the apo and FMN or NCA complexed states of CS between 1.5 Å and 1.7 Å. The crystals of MtCS belong to the space group $P6_422$ with the unit cell dimensions $a = 132.5$ Å, $b = 132.5$ Å, $c = 159.5$ Å. They contain two CS molecules in the asymmetric unit. After scaling the averaged intensities were used to calculate structure factor amplitudes using TRUNCATE from the CCP4 suite of programmes [CCP4. 1994] for MtGluRS and MtCS. Finally a unique set of reflections was generated and 5% were flagged for the calculation of the free R-factor for cross validation in the refinement process.

3.4. Phasing, Model building and refinement

Approximate initial phases for electron density calculation were obtained by using the molecular replacement (MR) program MOLREP [Vagin A and Teplyakow A, 1997] for phasing of both MtGluRS and MtCS.

MR for MtCS was carried out using the crystal structure of CS from *S. pneumoniae* (SpCS) complexed with EPSP and FMN as a search model. Only one chain of 1QXO corresponding to the residues 1 to 388 of the SpCS sequence was used for and all non-protein atoms were removed from the model. SpCS displays a very high homology towards MtCS with a sequence identity and similarity of 43% and 60%, respectively. The MR procedure was carried out in MOLREP using the entire protein chain of SpCS. Prior to MR the number of molecules in the asymmetric unit was estimated using the Matthews coefficient. For an assumed solvent content of approximately 50% a number of two molecules per asymmetric unit were estimated. Therefore MR was performed in order to place two molecules in the asymmetric unit. Rotation and translation searches were performed with MOLREP and the correctness of the solution was checked by submitting the output model to several rounds of refinement using REFMAC [Murshudov GN et al., 1997] (for details see below). This showed that the initial model converged with the structure factors used for the MR procedure. The convergence of the refinement was checked using both R- and R_{free} -factors and the R_{free} factor improved from 0.5 to 0.45 after several cycles of refinement. Additionally, correct formation of the tetramer upon application of symmetry operations was checked visually.

For MtGluRS the structure of TtGluRS complexed with Mg-ATP and glutamine (PDB 1J09) was used as a search model. Structure solution involved Dr. Gleb Bourenkov. The sequence of the search model was adjusted on the basis of an alignment in MOLREP. For MR the input model from 1J09 was split in two parts comprising the N-terminal part ranging from residues 6-320 (40% sequence identity between TtGluRS and MtGluRS) and the C-terminal part from 321-468 (22%), respectively. Rotation and translation searches for the N-terminal part revealed a clear solution

characterized by a correlation coefficient 0.40, as compared to 0.34 for the highest spurious solution. Similarly as in the MtCS case the solution for the N-terminal part was submitted to refinement with the structure factors for MtGluRS. After multiple cycles of refinement with REFMAC the R_{free} -factor of 0.52 calculated with the initial model improved to 0.49. Subsequently, the C-terminal part was located applying MOLREP, keeping the N-terminal part of the structure fixed.

Initial electron density maps from the molecular replacement solutions proved already to be well interpretable for both MtGluRS and MtCS and density for the better part of the sidechains was visible. Due to the good quality of the initial estimates of the phases an attempt for autotracing of the peptide chain was made. The automatic building of the correct sequence of the model was done using ARP/WARP [Perrakis A et al., 1999]. Afterwards incorrectly modelled parts of the sequence were identified in COOT with the newly obtained model and initial electron density maps and corrected manually using the modelling functions of COOT. Correctness of the model was improved by repeated cycles of model refinement. Typically one cycle followed the procedure outlined below:

- 1) The current model was submitted to 10 cycles of maximum-likelihood refinement in REFMAC using the experimental structure factors as input.
- 2) Manual checking and remodelling of the refined model was done by using COOT with likelihood weighted $2mF_{\text{O}}-DF_{\text{C}}$ and $mF_{\text{O}}-DF_{\text{C}}$ electron density maps obtained from the corresponding amplitudes and phases calculated by REFMAC5 and the experimental amplitudes. $2mF_{\text{O}}-DF_{\text{C}}$ maps were contoured at 1σ and $mF_{\text{O}}-DF_{\text{C}}$ maps were contoured at $\pm 3\sigma$

The convergence of the refinement process was checked quantitatively using both R- and R_{free} -factors of REFMAC and qualitatively the goodness of the electron density was checked. At the beginning of the refinement process a higher weight was put on the geometry of the model by using tight restraints in REFMAC. Upon convergence the restraints were lowered in favour of the experimental data. This allowed the model to adapt to the increasingly better contoured electron density.

Additionally every cycle crystallographic waters were inserted using COOT. For this purpose the mF_o-DF_c map was masked with the protein chain and peaks above 1σ were used to place waters. Wrongly placed waters were deleted if they had B-factors above 80 \AA^2 , were modelled into peaks with sigma levels below 1 and showed closest contacts with less than 2.3 \AA or more than 3.5 \AA . Alternatively waters were modelled using ARP/WARP. Remaining missing waters were identified and modelled manually. Additionally the presence of ions was identified by difference electron density, B-factors and hydrogen bond geometry. In the case of MtCS the ligands FMN and NCA were only modelled into their respective density upon completion of the modelling process to avoid any bias.

3.5. Structural analysis

The molecular geometry for all MtGluRS and MtCS models was checked with COOT, PROCHECK [Laskowski RA et al., 1993] and WHAT_CHECK [Hoofst RWW et al., 1996]. Different checks implemented in COOT were performed simultaneously to the manual modelling steps in order to keep the model in constant good agreement with the state-of-the-art of molecular geometry. Main chain torsion angles were additionally checked at regular intervals throughout the refinement process and at the final stage with PROCHECK. The final models were again submitted to a detailed check in WHAT_CHECK for finalisation of the models. Secondary structure was assigned with DSSP [Vriend G, 1983]. The final models were deposited together with the structure factors in the Brookhaven Protein Data Bank⁴. Entry codes for the different MtCS structures are 2O11, 2O12 and 2QHF for apo MtCS, MtCS-FMN and MtCS-NCA, respectively. Apo MtGluRS was submitted with the entry code 2JA2. Comparison of different structural conformations was carried out using LSQKAB [CCP4, 1994]. All figures of the structural analysis of MtGluRS and MtCS presented in the following chapters were prepared with PyMOL⁵ [DeLano W., 2002]. Alignments were prepared with CLUSTALW [Chenna R et al., 2003] and ESPript [Gouet P et al., 1999].

⁴ <http://www.pdb.org>

⁵ <http://www.pymol.org>

4. Chorismate synthase from *M. tuberculosis*

4.1. The shikimate pathway

The shikimate pathway is the main route for the biosynthesis of all key aromatic components involved in primary metabolism in prokaryotes, fungi, plants and apicomplexan parasites [Bentley R, 1990 and Haslam E, 1993] and was discovered more than 50 years ago [Srinivasan PR and Shigeura HT, 1956]. The pathway consists of seven metabolic steps beginning with the condensation of phosphoenolpyruvate and erythrose-4-phosphate and ending with the synthesis of chorismate (Fig. 4-1). In bacteria more than 90% of the metabolic energy is used for protein biosynthesis, nearly the entire output of aromatic biosynthesis are the aromatic aminoacids phenylalanine, tyrosine and tryptophane, while in higher plants a wide variety of secondary metabolites with aromatic ring structures such as flavonoids, alkaloids etc. are produced [Hermann KM, 1995]. In contrast the shikimate pathway is not present in mammals, which therefore have to derive aromatic compounds from their diet.

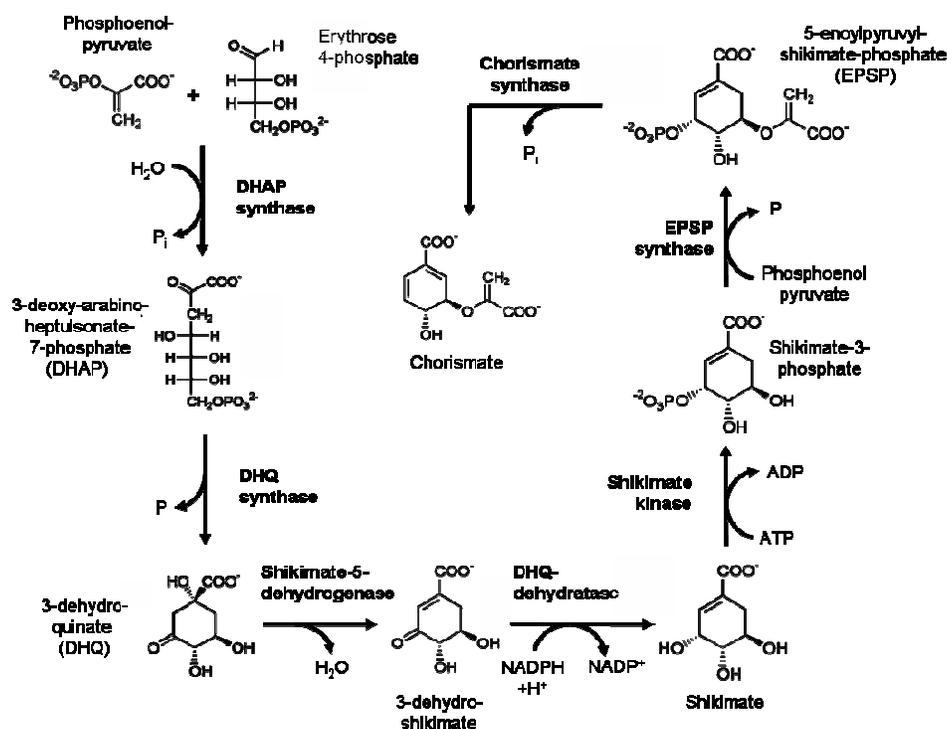


Figure 4-1: The shikimate pathway leads to the formation of chorismate from phosphoenolpyruvate and erythrose 4-phosphate via shikimate

Due to this fact there has been major interest for more than 25 years to use the enzymes from the shikimate pathway in the development of both herbicides and anti-microbial compounds [Coggins JR et al, 2003]. Generally the enzymes of the shikimate pathway are promising targets for anti-mycobacterial drug design as the pathway has been proven to be essential for *M. tuberculosis* [Parish T and Stoker NG, 2002]. In 1972 the first major success was reported in inhibitor design against shikimate pathway enzymes with the discovery of glyphosate (N-phosphonomethylglycine) [Jaworski EG, 1972], marketed by Monsanto as the billion dollar herbicide Roundup. Later, the enzyme 5-enoylpyruvyl-shikimate-phosphate synthase (EPSP synthase) was identified as the target for glyphosate [Steinrucken HC and Amrhein N, 1980]. EPSP synthase catalyzes the penultimate step in the shikimate pathway by formation of 5-enoylpyruvyl-shikimate-phosphate (EPSP) from shikimate-3-phosphate (S3P). Further inhibitor studies targeting shikimate pathway enzymes have been carried out on dehydroquinase (DHQase) [Coggins JR et al, 2003] and on chorismate synthase [Thomas MG et al, 2003] which catalyses the formation of chorismate in the final step of the shikimate pathway. Notably, despite all efforts no other potent inhibitor targeting shikimate pathway enzymes has come forth with a potential comparable to glyphosate.

Chorismate synthase (EC 4.2.3.5, 5-O-(1-carboxyvinyl)-3-phosphoshikimate phosphate-lyase) belongs to the group of flavoenzymes that uses flavin mononucleotide (FMN) for catalysis. FMN is basically riboflavin (Vitamin B₂) with a phosphorylated ribosyl-moiety (Fig. 4-2a). The isoalloxazine-moiety of flavins is involved in the catalysis of either free radical or two-electron chemistry. Here the flavin exists in three states (Fig. 4-2b): The fully oxidized flavin is reduced by one electron to a semiquinone, while further reduction by a second electron yields 1,5-dihydroflavin. Catalysis by flavoenzymes most often involves net redox changes, while sometimes it does not [Bornemann S, 2002]. Chorismate synthase (CS) is an example for the latter case and catalyzes a reaction that does not involve a change in the net redox state of FMN. The catalyzed reaction involves a trans-1,4-elimination of phosphate and the C(6_{proR}) hydrogen from the substrate EPSP (Fig. 4-2c). This reaction is unique in nature as it has an absolute requirement for reduced FMN, while the bond cleavages necessary for the eliminations at the substrate EPSP do not

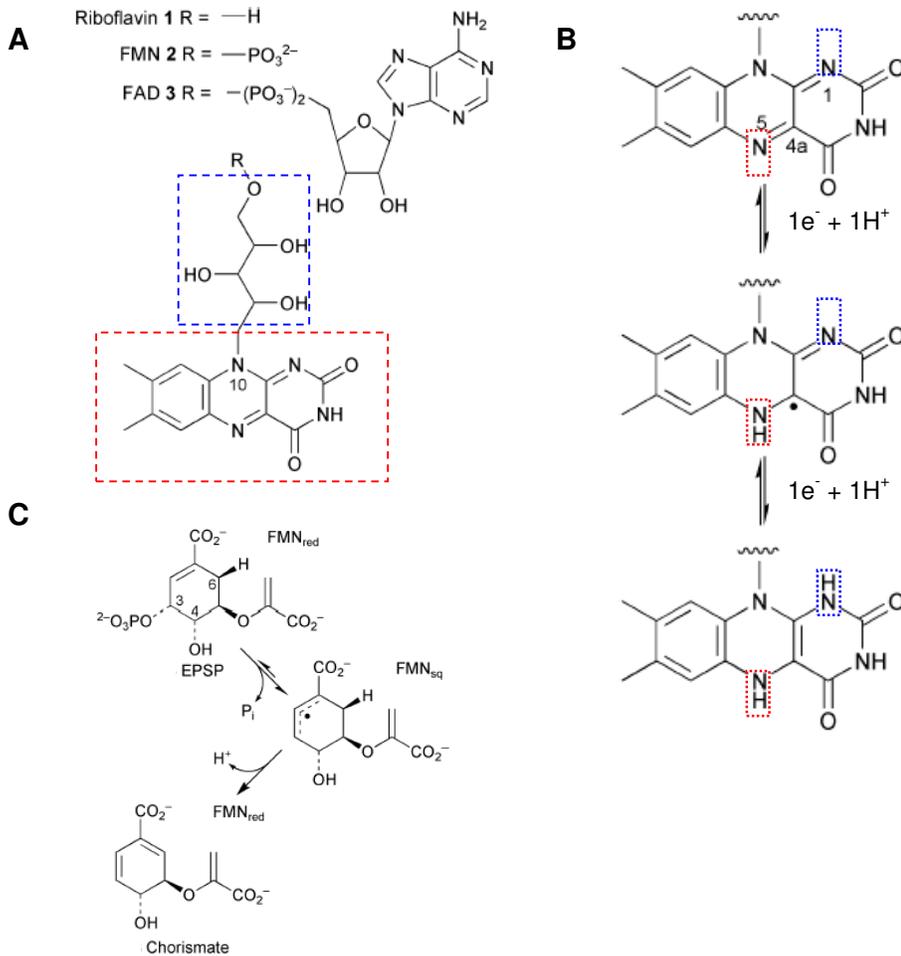


Figure 4-2: Flavins: A. Isoalloxazine- (red) and ribosyl-moieties (blue) of flavins. In FMN the ribosyl is phosphorylated. B. Flavin redox states involve proton transfers at N1 (blue) and N5 (red). The fully oxidized flavin (top) is reduced by one electron to a semiquinone (middle), while further reduction by a second electron yields 1,5-dihydroflavin (bottom). C. The chorismate synthase reaction involves a trans-1,4-elimination of phosphate and the C(6proR) hydrogen from the substrate EPSP without a net redox change in FMN. (Adapted from Bornemann S, 2002).

involve a net redox change [Macheroux et al, 1999]. Due to the special chemistry it catalyzes, CS has caused much attention to illuminate this unique reaction.

From all shikimate pathway enzymes CS is a prime candidate for drug design because it has been identified as one of the rate-limiting enzymes of this pathway [Dell KA and Frost JW, 1993]. CS lies at the node of five distinct biosynthetic pathways and in addition to the compounds mentioned above chorismate is also a precursor for the production of folate in the folate pathway. As in the case of the shikimate pathway the folate pathway is also largely absent in mammals and has

therefore been a target for drug design. An example for this is the inhibition of dihydrofolate reductase by trimethoprim or the inhibition of dihydropteroate synthase by sulphonamides. As inhibition of CS would further affect the folate pathway, the enzyme is also interesting for a dual therapy approach, using both an inhibitor for CS and an antifolate. As CS is highly conserved throughout species it is very interesting for the design of broad-band therapeutics.

To date crystal structures of CS were provided for several organisms. These included *Saccharomyces cerevisiae* (ScCS) [Quevillon-Cheruel S. et al., 2004], *Campylobacter jejuni* (CjCS) [Forouhar F. et al., unpublished data], *Mycobacterium tuberculosis* (MtCS) [Dias M.V.B. et al., 2006], *Aquifex aeolicus* (AaCS) [Viola C.M et al., 2004], *Helicobacter pylori* (HpCS) [Ahn H.J. et al., 2004] and *Streptococcus pneumoniae* (SpCS) [Maclean J. and Ali S., 2003]. While in most of the structures the enzymes were reported to be in an unliganded state, two structures presented chorismate synthase in complex with its natural ligands FMN and/or EPSP. The structure of HpCS was solved as a binary complex containing a bound FMN molecule, while SpCS was solved as a ternary complex having bound both FMN and the substrate EPSP. The SpCS-FMN-EPSP structure was consistent with previous kinetic studies [Macheroux et al., 1998] indicating ordered sequential binding of FMN followed by EPSP [Maclean J. and Ali S., 2003; Bornemann S et al., 2003]. The relative orientation of both ligands was in agreement with a productive enzymatic reaction state belonging either to a reductive free radical mechanism involving donation of an electron by the flavin and abstraction of a hydrogen atom from the substrate, or to an alternative cationic mechanism involving the formation of a cationic substrate intermediate following phosphate elimination [Maclean J. and Ali S., 2003; Bornemann S et al., 2003]. The binary HpCS-FMN structure, however, exhibited a substantially different binding mode of the cofactor [Ahn H.J. et al., 2004], leaving the question open whether initial binding of FMN was followed by re-orientational motions associated with or preceding subsequent binding of EPSP.

This work reports crystal structures of MtCS in the unliganded state and in complex with FMN and NCA, respectively. All structures were solved at high resolution, 1.6-1.7 Å. The binary MtCS-FMN complex shows the cofactor in a closely similar location and orientation as in the ternary SpCS-FMN-EPSP structure 1QX0. This result and

the high degree of structural homology provide a basis for a detailed description of the conformational changes associated with the sequential binding of the ligands to chorismate synthase. This description allowed for the first time the identification of a single, strictly conserved arginine residue (Arg341 in MtCS, see also Fig. 4-4b) that is involved in all key conformational events necessary for the correct enzymatic function of the enzyme. Further, molecular dynamics calculations were employed to validate the proposed role of Arg341 and to gain a more detailed insight on the dynamical behavior of CS and its active site. The NCA-bound structure of MtCS, whereby NCA was not present in the crystallization solution, presents a completely new binding mode in MtCS. This binding mode displays an active site conformation similar to the apo form where Arg341 adopts a conformation that disfavours ligand binding. Therefore the MtCS-NCA complex proposes an interesting lead for the development of compounds that allow a stabilization of the unproductive conformation of the active site in MtCS. The combination of experimental and theoretical results provides comprehensive structural information that might be highly valuable for rational drug design efforts targeting the shikimate pathway of *M. tuberculosis*.

4.2. Structural analysis

4.2.1. Model quality

The crystal structures of MtCS in unliganded form and its binary complexes with FMN and NCA were solved in space group $P6_422$. Final statistics for data processing and refinement are presented in Table 4-1. The crystal structure of the ternary SpCS complex 1QXO was used as a search model for molecular replacement with the program MOLREP [Vagin A and Teplyakov A, 1997]. Based on the solution from molecular replacement MtCS structures were modelled and refined at high resolution of 1.6 Å for MtCS-NCA and apo MtCS and at 1.7 Å for MtCS-FMN (Fig. 4-3a). All three models converged to low R -/ R_{free} -values of approximately 16-18% and 18-20%, respectively. Stereochemical quality was assessed by PROCHECK and showed for the apo MtCS and MtCS-NCA models 100% of all residues in the most favoured or allowed regions of the Ramachandran plot. In MtCS-FMN 99,7% of all residues were in the most favoured or allowed regions, while Ala323 was observed in a disallowed region.

Table 4-1: Data collection and refinement statistics

Data set	Apo	FMN-bound	NCA-bound
Wavelength (Å)	1.05	1.05	1.05
Space group	P6 ₄ 22	P6 ₄ 22	P6 ₄ 22
Unit cell (Å)	a=b=133.5, c=160.0 $\alpha=\beta=90^\circ$, $\gamma=120^\circ$	a=b=133.4, c=158 $\alpha=\beta=90^\circ$, $\gamma=120^\circ$	a=b=132.6, c=159.5 $\alpha=\beta=90^\circ$, $\gamma=120^\circ$
A. Diffraction statistics			
Resolution	20.00 – 1.65 (1.67 – 1.65)	20.00 – 1.72 (1.74 – 1.72)	20 – 1.65 (1.67 – 1.65)
No. of observations	453,503	676,002	305,669
No. of unique reflections	100,281	88,273	98,095
Completeness	99.3 (99.7)	99.9 (99.9)	98.8 (99.3)
I/sigma(I)	20.5 (2.8)	20.6 (2.3)	30.3 (2.3)
Rmerge (%)	0.075 (0.448)	0.086 (0.932)	0.055 (0.521)
B. Refinement statistics			
Resolution range (Å)	19.71 – 1.65	19.74 – 1.72	19.65 – 1.65
No. of reflections (working/test)	93,729/5,015	83,603/4,415	93,091/4,977
R/Rfree (%)	18.8/20.7	16.5/18.8	16.3/18.6
Mean B-factor (Å ²)	23.9	23.0	22.5
Rms bond length deviation	0.014	0.011	0.011
Rms bond angle deviation	1.511	1.317	1.346

4.2.2. Overall structure

All structures contained two MtCS molecules in the asymmetric unit. An analysis with the EBI PISA tool [Krissinel E and Henrick K, 2005] suggested a homotetrameric organization of the biological unit. In MtCS it is build up by application of point group symmetry operations on the single protomer visible in the unit cell (Fig 4-3b). The tetrameric organization of MtCS is in agreement with previous observations of a similar configuration of chorismate synthases in several crystal structures (1R52, ScCS; 1Q1L, AaCS; 1UMF/1UM0, HpCS; 1QXO, SpCS). Further this organization of the enzyme was observed in solution as well [White PJ et al., 1988].

The topology of the MtCS protomer is closely similar as in the other known CS structures. Due to the exhaustive description of the unique CS topology in the recently published literature [Quevillon-Cheruel, S et al., 2004; Dias MVB et al., 2006; Viola CM et al., 2004; Ahn HJ et al., 2004; MacClean J. and Ali S., 2003], this work will only give a comprehensive report on this topic. MtCS is composed of a 4 β -4 α -4 β sandwich like core domain that is surrounded by discrete stretches of β -strands and α -helices (Fig. 4-4). The apo-MtCS structure presented in this work was determined at much higher resolution, 1.6 Å, as compared to the recently described apo-MtCS structure 1ZTB (2.65 Å resolution) [Dias MVB et al., 2006]. Secondary structural counts are slightly different between the here presented MtCS structures (16 β -

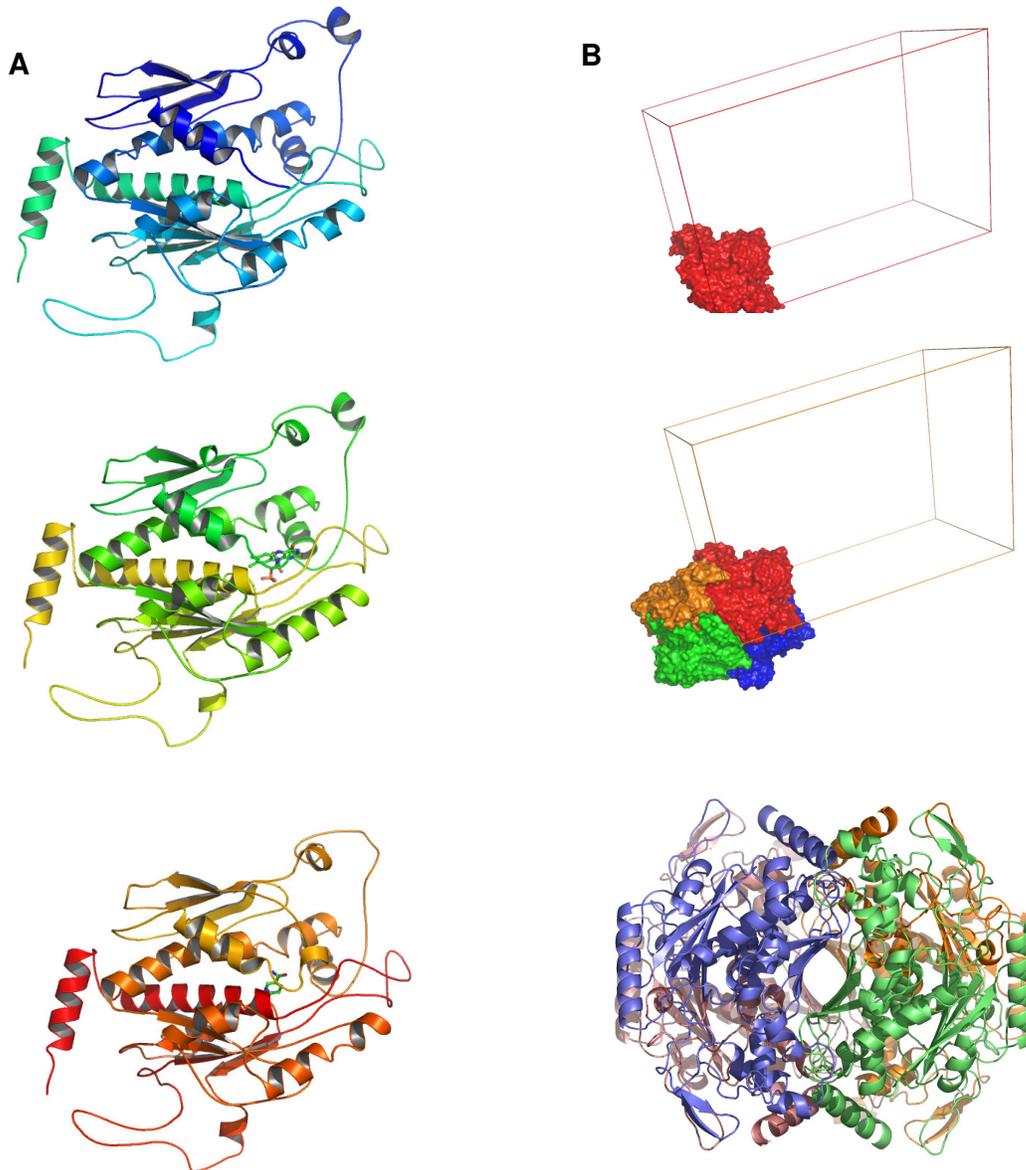


Figure 4-3: MtCS topology and $P6_422$ symmetry. A: MtCS protomers from top to bottom: 1.) apo MtCS; 2.) FMN-bound MtCS; 3.) NCA-bound MtCS. B: Tetramer formation by $P6_422$ point group symmetry: top: MtCS protomer (red) in the unit cell. Middle: Tetramer formation upon application of $P6_422$ point group symmetry. Single protomers are colored red, orange, green, and blue. Bottom: MtCS tetramer in cartoon representation. Single protomers are colored orange, blue, green and wheat.

strands and 14 α -helices) and the one published by Dias et al. (17 β -strands and 13 α -helices). However, both apo-MtCS structural models superimpose with an r.m.s. deviation (RMSD) of 0.6 Å between corresponding C_α atoms, showing close structural similarity. All further references made to MtCS will be concerned with the

A

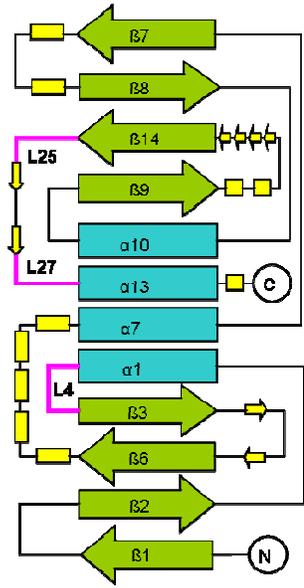
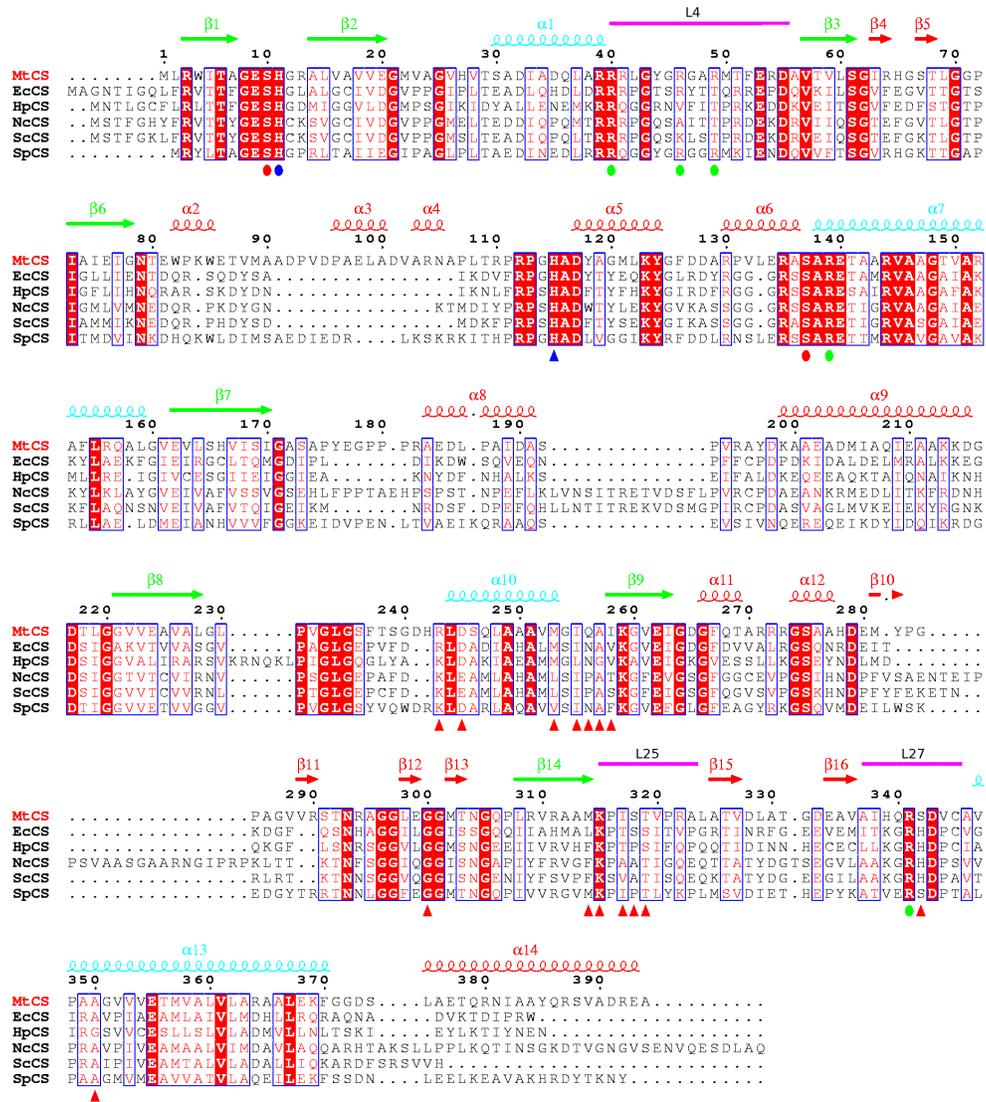


Figure 4-4: MtCS topology and alignment. A. The MtCS topology consists of a 4 α -4 β -4 α sandwich like core domain that is surrounded by discrete stretches of β -strands and α -helices. Magenta colored loops are involved in conformational changes during sequential ligand binding. B. Alignment of CS from *M. tuberculosis* (MtCS), *E. coli* (EcCS), *H. pylori* (HpCS), *N. crassa* (NcCS), *S. cerevisiae* (ScCS), and *S. pneumoniae* (SpCS). Secondary structure is indicated for MtCS as presented in this work (see also Table 4-1) in red, blue and green. Blue and green secondary structure elements indicate the 4 β -4 α -4 β sandwich core. Red circles denote EPSP binding residues from the ternary SpCS complex (PDB-ID 1QXO); green circles denote arginines in contact to EPSP. FMN binding residues from ternary SpCS and binary MtCS complexes are depicted by red triangles. Blue triangles/circles indicate invariant histidines in contact to FMN and EPSP, respectively. Loops involved in conformational changes during ligand binding are shown in magenta.

B



structures presented in Table 4-1, if not otherwise stated. While in the present NCA-liganded structure, all 392 residues of MtCS were located in electron density the surface loop L4 was partly disordered in the structural models of apo-MtCS (residues 46-54) and MtCS-FMN (47-52). All three structural models contain a number of acetate molecules taken up from the crystallization buffer. However, it becomes clear by the description of the protomer conformation and its comparison to other CS structures that these acetates do not compromise the overall fold of the protein.

4.2.3. FMN binding in the binary MtCS complex.

Soaking MtCS crystals with FMN yielded the binary MtCS-FMN complex, which exhibited the cofactor with full occupancy in unambiguous electron density (Fig. 4-5). Both the phosphate- and ribityl-parts of FMN were found to be buried deep inside the enzyme, whereas the isoalloxazine-moiety was oriented towards the protein surface. In total, the cofactor is bound in a pocket formed by residues belonging to two neighboring monomeric subunits (Fig. 4-6a) of the tetramer. The binding site of the isoalloxazine moiety is consisting of two parts. The ortho-xylyl ring of FMN is located in the hydrophobic part of the binding pocket comprising the apolar residues Met314, Ile317 and Ala138. More polar parts of the binding site, including Thr319, Arg341, Asp343, Ser118 and His115, close around the N1-O2-locus of the isoalloxazine moiety. The ribityl-chain is in contact with His115 and Ala136. The phosphate group is coordinated by Arg243, Lys315, Gly300, and Ala257. In addition, it is in contact with three residues (Asp245, Gly301, Leu298) from a symmetry-related

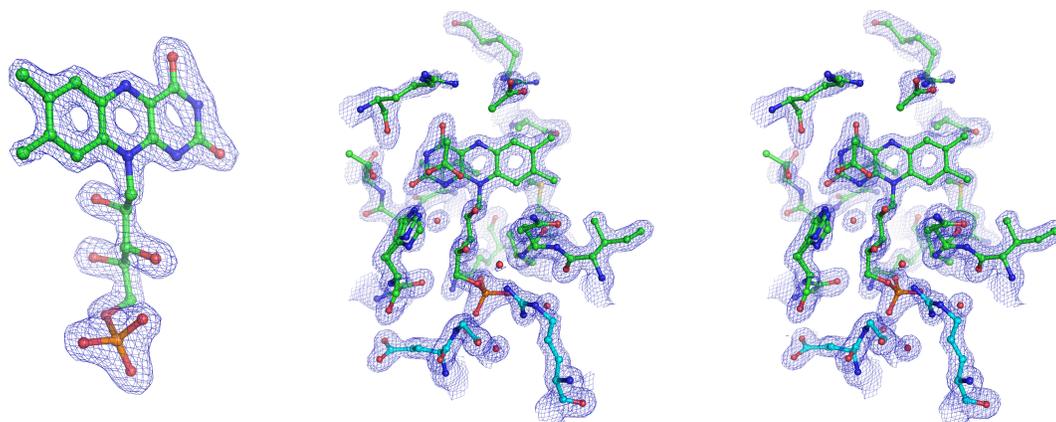


Figure 4-5: FMN binding and electron density. On the left, the FMN mF_0 - DF_C omit map contoured at 8σ unambiguously shows the orientation of the isoalloxazine- and phosphoribosyl-moieties of the co-factor. On the right, a stereo view of the FMN binding site is shown with $2mF_0$ - DF_C density contoured at 2σ .

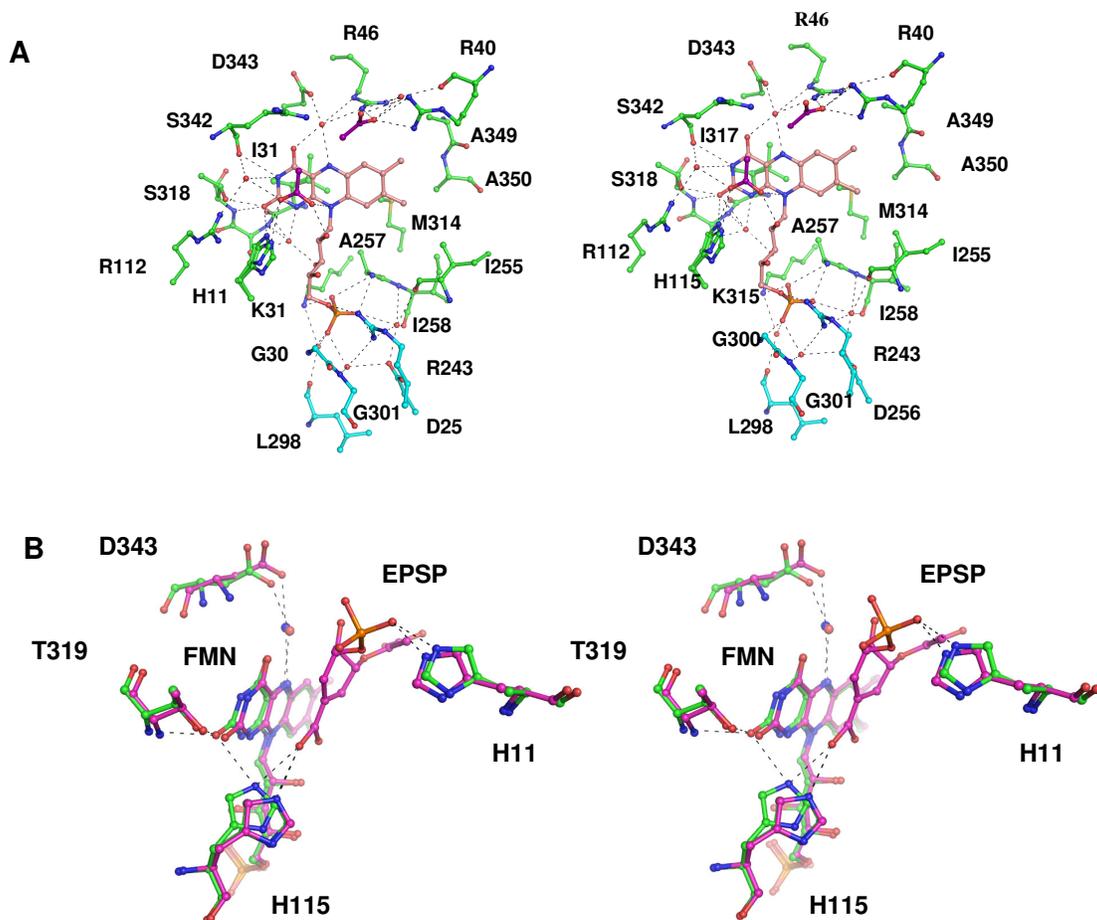


Figure 4-6: FMN binding and active site details. A. Stereo view of the active site in the MtCS-FMN complex. The co-factor FMN (salmon) is bound in a pocket that is built up by two neighbouring protomers (green and cyan). Polar contacts are displayed by broken black lines and include several crystal waters (red) and two acetate molecules (purple). B. Contacts between FMN, Thr319, and the invariant His115 are conserved in MtCS (green) and SpCS (PDB-ID 1QXO, purple). His11 adopts a conformation in the binary complex that is necessary for proper interaction with EPSP in the ternary complex. In contrast, contacts between the invariant aspartate (Asp343 in MtCS) slightly differ in binary and ternary complexes (see text for details).

monomer belonging to the same biological tetramer. Several of these interactions are mediated via water molecules.

The mode of FMN binding in the binary MtCS complex is essentially the same as in the ternary SpCS-FMN-EPSP structure 1QXO. Eleven of the sixteen residues contacting FMN in SpCS are conserved in MtCS, and the remaining non-conserved residues fulfill equivalent functions as in SpCS (Fig. 4-4b). The C_{α} atoms of the sixteen residues superimpose with an RMSD of 0.3 Å, showing that the FMN binding sites have nearly identical conformations in both binary and ternary structures. Generally SpCS and MtCS show close structural similarity, as the overall C_{α} -RMSD

between the monomers of SpCS to MtCS is only 0.8 Å. The FMN binding mode in the MtCS-FMN complex is consistent with the description of catalysis given for SpCS [MacLean J and Ali S, 2003; Kitzing K et al., 2003; Rauch G et al., 2007]. An essential aspect is the insertion of the N1-O2 locus of FMN into a pocket formed by Thr319 and His115 in MtCS (Fig. 4-6b and Fig. 4-11a). Based on kinetic studies of two mutants from *Neurospora Crassa* CS (NcCS), it was suggested that two invariant histidines participate in the protonation-deprotonation events during catalysis [Kitzing K et al., 2003]. The corresponding residues His11 and His115 of MtCS (see also Fig. 4-4b) are in perfect agreement with this mechanistic description. A superposition shows that His11 in MtCS-FMN, similar as the equivalent His10 in SpCS-FMN-EPSP (Fig. 4-6b), occupies a position near the phosphate group of EPSP, where it may act as a proton donor during the catalytic reaction. The side chain orientation of His115 is appropriate for protonating the reduced FMN in its monoanionic state, thereby maximizing either charge or electron transfer from FMN to the ring double bond in EPSP. This is believed to be crucial for the catalysis of phosphate release from EPSP [MacLean J and Ali S, 2003; Kitzing K et al., 2003]. His115 occupies a second, alternate position in MtCS-FMN, where the side chain is in contact to an acetate molecule. This and another acetate are located at essentially the same positions as the two carboxylate groups of EPSP in the ternary complex of SpCS (Fig. 4-7). Both acetates are coordinated by residues that are conserved in SpCS as well as in MtCS and involved in EPSP binding in the ternary complex (see also Fig. 4-4b). The location and binding interactions of the two acetates prove that the active site in the binary complex adopts already a conformation required for the binding of the carboxyl-moieties of EPSP. However, the two surface loops L25 and L27 have to undergo changes from their conformations in the MtCS-FMN complex in order to liberate the binding site for the entire EPSP molecule.

In contrast to the close agreement in the FMN binding modes in the MtCS-FMN and SpCS-FMN-EPSP complexes, the binary HpCS-FMN structure 1UM0 displayed a significantly different binding mode of the cofactor. The orientation of the isoalloxazine ring in 1UM0 is flipped by 180°. As a consequence, the invariant histidine corresponding to His11 of MtCS is not able to contact the FMN as described above. Similarly, contacts to the invariant aspartate are hindered suggesting that the binding mode of FMN in the HpCS-FMN structure is catalytically inactive, in contrast

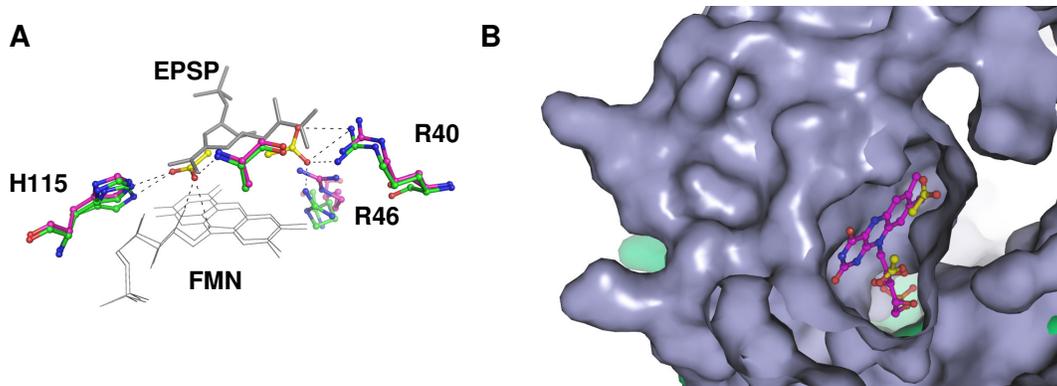
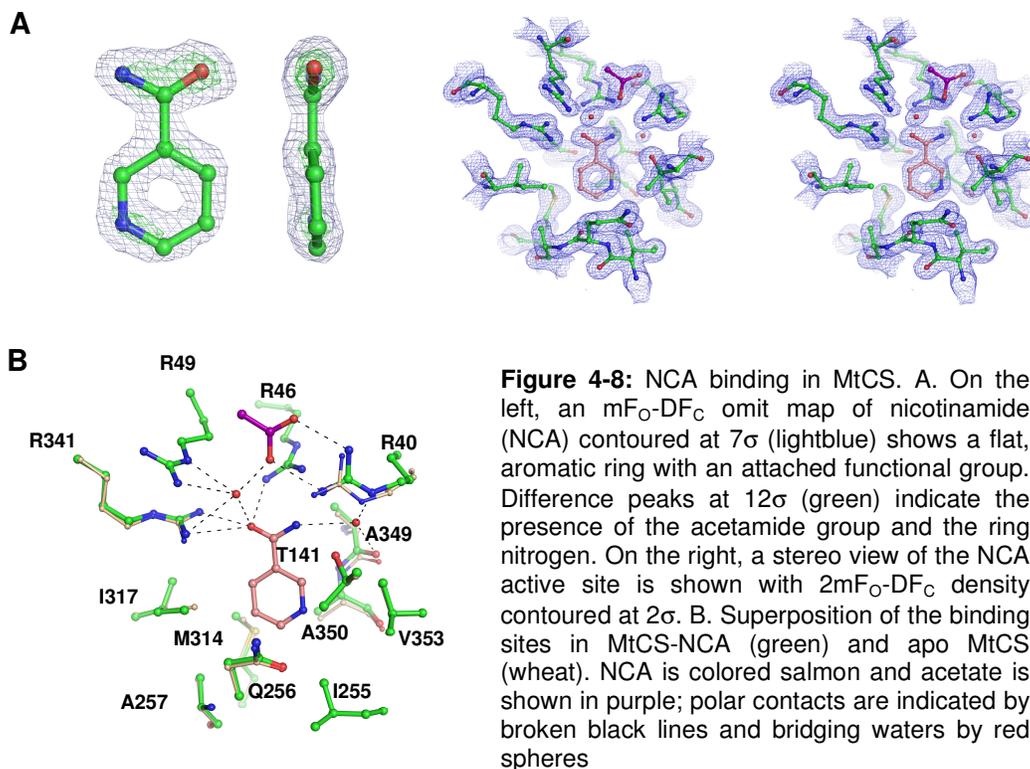


Figure 4-7: Details of FMN binding. A. Details of acetate binding in the active site of the MtCS-FMN complex. The acetates (yellow) occupy the positions of the EPSP carboxyl-moieties (gray sticks) in the ternary complex and are coordinated by conserved residues in both MtCS (green) and SpCS (purple). For reasons of clarity FMN binding in the binary and ternary complexes is indicated by gray lines. B. View from the surface of MtCS-FMN into the binding cavity is shown. The cavity already supplies sufficient space for EPSP binding. Only Arg341 occupies the binding site of EPSP (left out for reasons of clarity). The neighbouring CS monomer is indicated in green.

to the catalytically competent binding modes observed both in the MtCS-FMN and SpCS-FMN-EPSP structures. Inspection of the electron density calculated from the published structure factors of 1UM0 revealed that the FMN in the HpCS-FMN structure was rather poorly defined. Remodelling the FMN ligand in the HpCS-FMN complex in a closely similar binding mode as in MtCS-FMN resulted in an improved fit to the experimental density and similar interactions with the catalytically important residues as in the MtCS and SpCS complexes. Clearly, the different orientation of the FMN in 1UM0 as compared to the modes of FMN binding to MtCS and SpCS represented an artifact resulting from modelling into poor electron density. The remodelled FMN site of the HpCS-FMN complex indicates that HpCS, MtCS and SpCS interact with FMN in closely similar modes permitting subsequent binding of EPSP without substantial reorientational motions of the FMN.

4.2.4. Binding of nicotinamide to MtCS

Crystallization of MtCS under acidic conditions (pH 4.0) resulted in the presence of a bound ligand in the electron density maps (Fig. 4-8a). Due to the high resolution, 1.6 Å, of the structural analysis, the ligand could unequivocally be identified as nicotinamide (NCA). NCA was not added to the crystallization solutions and therefore must be picked up during expression of the recombinant enzyme in *E. coli* cells. The MtCS-NCA structure showed NCA located in a tight pocket (Fig. 4-8b), where the



acetamide group makes direct hydrogen bonds to Arg341 and Arg46. Additional hydrogen bonds, mediated via two crystal waters, bring NCA in contact with Arg40, Arg49, Ala349 and an acetate molecule that is coordinated to Arg40. Corresponding to the hydrophobicity of the aromatic ring, NCA is in the neighborhood of several apolar residues including Val353, Ile317, Met314, Ala138, Ile255 and Ala350. The surface loop L4 is well defined in the MtCS-NCA structure, possibly due to interactions of NCA with Arg49 (Fig. 4-8b). In contrast, the apo-MtCS and MtCS-FMN structures, where this residue is not coordinated, show that the segment comprising the residues 46-54 and 47-52 of L4, respectively, is disordered (Fig. 4-10a/b). For apo MtCS and MtCS-NCA the enzyme conformation around the NCA binding site is essentially identical (Fig. 4-8b). Compared to MtCS-FMN NCA binds to the same hydrophobic part of the co-factor binding pocket as the orthoxylyl ring (Fig. 4-9) and NCA primarily interacts with the invariant Arg341 that blocks part of the FMN binding site.

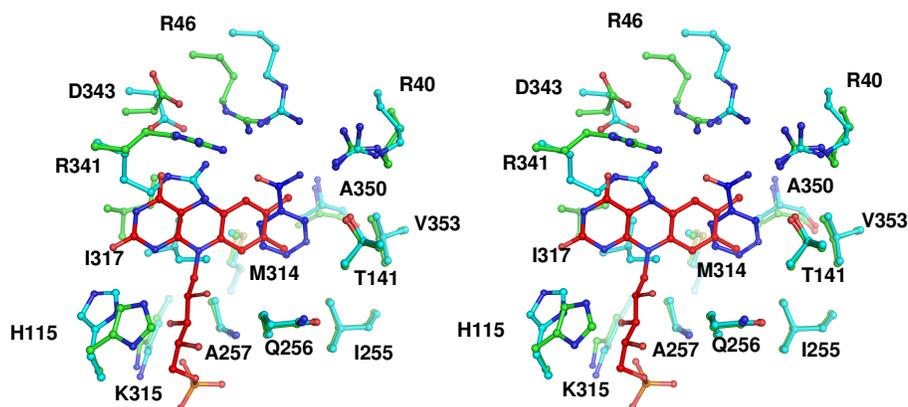


Figure 4-9: NCA vs FMN binding Stereo view of the active site after superposition of MtCS-FMN (green) and MtCS-NCA (cyan). The binding mode of FMN (red) and NCA (dark blue) overlap in the vicinity of the ortho-xylyl moiety of the co-factor.

4.2.5. Interim summary

The structural analysis of *M. tuberculosis* chorismate synthase (MtCS) provides the following key aspects:

- The high-resolution X-ray structures of apo MtCS, MtCS-FMN and MtCS-NCA are of high quality and display an overall structure similar to other known homologue CS.
- The MtCS-FMN complex suggests for the first time that the co-factor binding mode in the binary complex is essentially the same as in the ternary complex and there are no re-orientational motions of FMN involved upon ternary complex formation.
- Comparison of the binary MtCS and ternary SpCS complexes shows essentially identical binding sites in both enzymes, suggesting that FMN binding prepares the active site for EPSP binding.
- The binding site of NCA overlaps with the binding site of FMN and is suggested to support an unproductive active site conformation involving contacts between NCA and the invariant Arg341 in MtCS

4.3. Implications for the enzymatic mechanism

4.3.1. Active site switches during sequential ligand binding

When FMN binds to apo MtCS as the first ligand the two surface loops L25 and L27 undergo motions (Fig. 4-10a/b and Fig. 4-11b/c). Corresponding C_{α} atoms show maximum displacements of 4.1 Å and 3.5 Å for L25 and L27, respectively. In L25 the segment from residue 315 to 319 moves away from the active site thereby increasing the cavity volume for the accommodation of FMN. In contrast the segment comprising the residues 320 to 326 of L25 shows a slightly inward oriented motion that coincides with a similar movement of L27 from residues 334 to 341. The adjacent segment from residue 342 to 344 near FMN again is slightly oriented away from the FMN binding site. Thus, the motions of L25 and L27 near FMN open the active site for accommodation of the isoalloxazine moiety, while the motions of the segments from L25 and L27 farther off from FMN lead to a tighter packing of the loops, thereby probably stabilizing the active site conformation of L25 and L27 in contact to FMN. For FMN binding the side chain of Ile317 has to rotate away from the binding site that it partly occupies (Fig. 4-11a/b). This is accompanied by conformational changes in L25 that permit the insertion of the N1-O2 locus of FMN into a pocket formed by Thr319 and His115 (Fig. 4-11a). Further the conformational change around Ile317 leads to a substantial increase in the hydrophobic surface area for this residue from 15 Å² in the apo form to 38 Å² in the binary complex. This increase in hydrophobicity might facilitate the binding of the isoalloxazine moiety of FMN. Yet, the majority of the hydrophobic residues (Ala346, Ala138, Ala257, Met314, A349, Val353, Thr141, Ile255 and Ala350) involved in the accommodation of the isoalloxazine moiety are already in their respective conformations in the apo form of MtCS. L27 contains a strictly conserved residue, namely Asp343 in MtCS, that was just recently suggested to be participating in the abstraction of the C(6 $proR$) hydrogen from EPSP by the N(5)-locus of FMN [Rauch G et al., 2007]. Together with the invariant Arg341 this aspartate is involved in a conformational switch of the active site upon binary complex formation. In L27 the side chain of Arg341 has to rotate away from the active site during co-factor binding (Fig. 4-10b; Fig. 4-11b). This movement is accompanied by rearrangements of L27 that lead to the release of the invariant Asp343. In the apo form Asp343 is hydrogen bonded via its carboxyl group

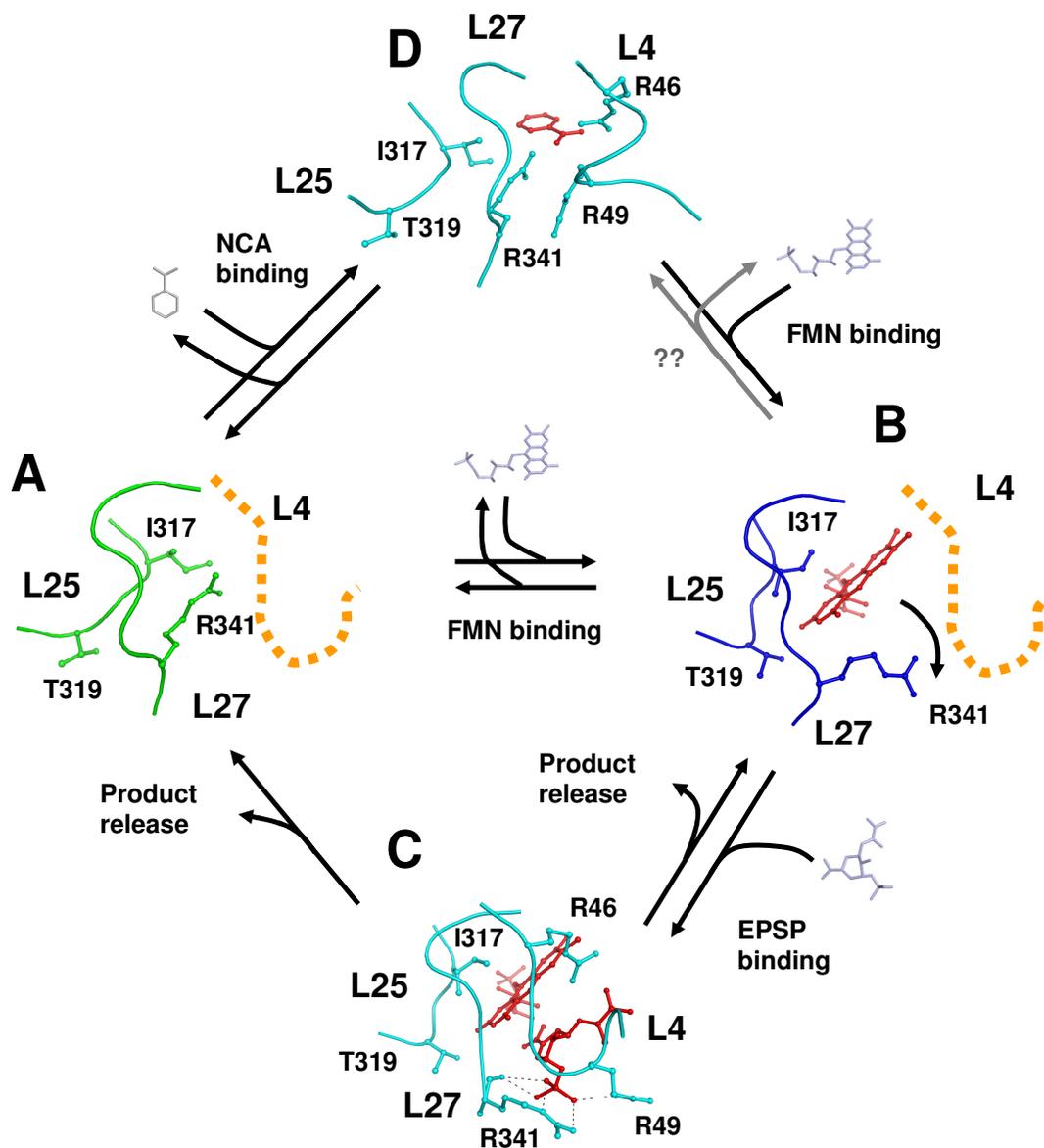


Figure 4-10: A binding cycle for CS and involved structural changes. A. In the apo form L4 is partly disordered but L25 and L27 show distinct conformations. B. Formation of the binary complex goes along with conformational changes in L25 and L27 upon FMN binding, Thr319 and Arg341 move away from the active site, but L4 stays partly disordered. C. Formation of the ternary complex by EPSP binding leads to a closure of the active site and L4 fully adopts a distinct conformation. The phosphate of EPSP is tightly coordinated by Arg341 and Arg49 from L27 and L4, respectively. Product release may go along with disordering of L4 upon loss of Arg341 coordination. D. NCA binding causes a closed alternative active site conformation similar as in the ternary complex, contacting all three loops involved in binding of FMN and EPSP (L4, L25 and L27). Nevertheless binary MtCS-FMN complex formation may occur from the NCA bound state. Vice versa, NCA binding from the binary complex has not been observed.

to the nitrogens NE and NH2 of the Arg341 guanidinium group. Breaking of these hydrogen bonds allows Asp343 to adopt the position necessary for participating in the deprotonation of FMN-N(5) as proposed [Rauch G et al., 2007]. Further, in apo MtCS the conserved water molecule thought to participate in deprotonation of N(5) is not observed. The location of the conserved water is occupied by the NH2 of the bulky guanidinium moiety of Arg341. Thus, only the conformational changes of Asp343 in L27 supported by the movement of Arg341 and the binding of FMN allow the stable coordination of the conserved water molecule to prepare the binding site for catalysis. This suggests a special role for the conserved Arg341 as promoting a switch of the active site from an unproductive to a catalytically competent conformation.

Due to the unavailability of EPSP for this work it was not possible to obtain a ternary complex of MtCS in the crystal. Therefore the following description of conformational changes between binary and ternary complexes will be based on the comparison of MtCS-FMN and SpCS-FMN-EPSP. This is possible, as virtually the entire binding site is conserved between both enzymes. Furthermore, the most important residues necessary for catalysis are generally conserved between all chorismate synthases. For reasons of clarity the following description of conformational changes occurring upon EPSP binding in the binary complex will be based on the numbering of residues and loops of MtCS if not otherwise stated. Comparison of binary and ternary CS complexes suggests that the binding of EPSP to the binary complex involves motions of the loops L4 and L27 in MtCS (Fig. 4-10b/c and Fig. 4-11b/c). Thereby both loops may act as a lid that attaches around the binding site. In this respect the sidechains of the two conserved arginines in L4 (Arg49) and L27 (Arg341) of MtCS interact with the phosphate group of EPSP as seen in SpCS accompanied by the adoption of well ordered conformations by both loops (Fig. 4-10c). These conformational changes are suggested to establish the closed binding site in the ternary complex, whereby the EPSP adopts a position above the non-reactive oxidized FMN ready for catalytic PO_4^{3-} elimination. The invariant aspartate in binary MtCS (Asp343) and in ternary SpCS (Asp339) is in contact to the N(5)-locus of FMN via an additionally conserved water molecule (Fig 4-6b). This allows the aspartate to participate in a water coupled deprotonation of N(5) (see also Fig. 4-2b). Mutation of this asparagine to alanine or aspartate led to a strongly decreased enzymatic activity in NcCS [Rauch G et al.,

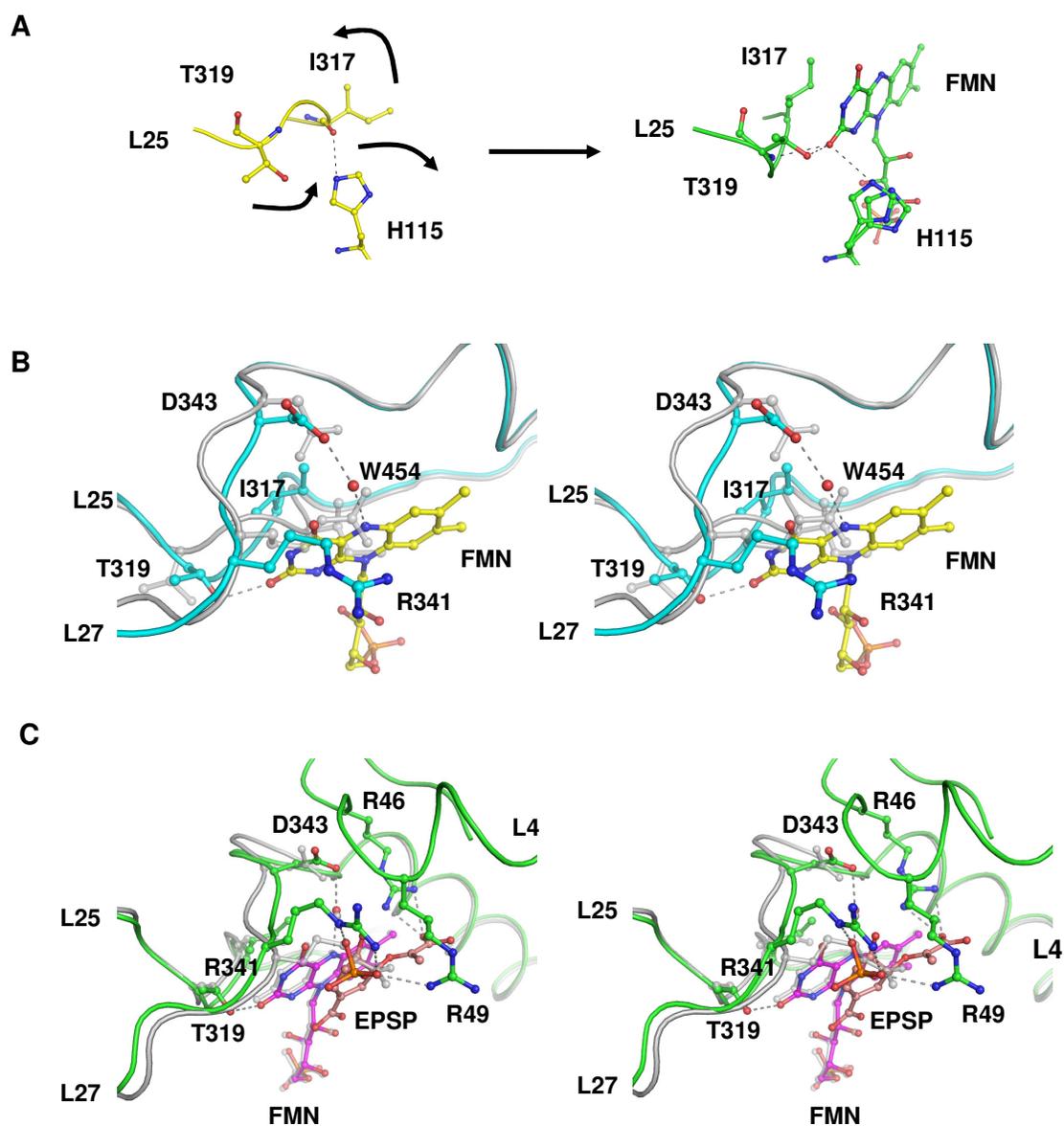


Figure 4-11: Active site rearrangements upon FMN binding. A. Formation of a pocket for the N1-O2-locus of FMN. In the apo form of MtCS (yellow) a hydrogen bond is present between His115 and Ile317 in L25. Disruption of this bond and coordinated movements of Thr319, His115 and Ile317 lead to the formation of a binding pocket for the N1-O2 locus of FMN in the binary complex (green). B/C. Stereo views of the conformational changes upon sequential ligand binding in CS as inferred from comparison of apo MtCS, MtCS-FMN and SpCS-FMN-EPSP. B. Apo MtCS is shown in gray and MtCS-FMN in cyan. Motions in L25 and L27 lead to an accommodation of the conserved water 454 between Asp343 and FMN:N5. Further the N1-O2-locus of FMN is accommodated near Thr319. Arg341 and Ile317 move away from the FMN binding site. C. Binding of EPSP (salmon) in the ternary complex (green) next to FMN (magenta) leads to a coordination of Arg49 and Arg341 by the EPSP phosphate. Binary MtCS-FMN is shown in gray.

2007]. Based on this mutagenic study it was proposed that N(5) is the acceptor of the C(6 $proR$) hydrogen, abstracted in the second step after PO_4^{3-} release from EPSP during chorismate formation. In the binary MtCS-FMN complex both the water molecule and the side chain of Asp343 are already in their respective position necessary for catalysis. Yet, the carboxyl group of Asp343 is in a slightly different conformation than Asp339 in SpCS (Fig. 4-6b). The outcome of this is a considerably less favorable hydrogen bonding angle with the conserved water molecule and MtCS-Asp343 of about 142° compared to nearly 120° for SpCS-Asp339. Therefore the proton abstraction from EPSP might be triggered by binding events occurring upon binding of the substrate. This suggests a special role for the invariant Arg341 that is located near Asp343. While in ternary SpCS the phosphate group of EPSP is in contact to the guanidinium group of the arginine corresponding to Arg341 in MtCS, this strictly conserved residue partly occupies the binding site of EPSP in the binary MtCS complex. Thus upon EPSP binding and phosphate coordination by the invariant arginine at position 341 in MtCS, the induced conformational change in L27 may trigger the Asp343 side chain to adopt a more favorable position for participating in proton abstraction from N(5) of FMN. This mechanism is further supported by an alternative conformation seen for the residue corresponding to Arg341 in SpCS (Arg337). Here the guanidinium group of Arg337 is turned away from the EPSP not making any contacts to its phosphate group. This coincides with an alternate conformation of the invariant aspartate, resulting in an unfavorable hydrogen bonding angle of about 138° . Both conformations seen for L27 in MtCS and the corresponding loop in SpCS suggest an active site switch supported by contacts between the invariant arginine and bound EPSP, fine-tuning the active site of CS for catalysis. Finally the invariant arginine might play a role in product release (Fig. 4-10a/c). After the phosphate group is eventually split off from EPSP during catalysis in the ternary complex, the conserved arginine in position 341 of MtCS would lose its tight coordination and thus the loop corresponding to L27 would regain flexibility. Similar events would happen in MtCS with L4 as the other conserved residue Arg49 is coordinating the phosphate group as well. Yet, the open conformation seen in the ternary complex 1QXO shows, that flexibility of L27 alone would leave sufficient space for chorismate to leave the binding site. Thus, the active site can open again to release the product chorismate. The comparison of all different ligation states of CS from apo over binary to the ternary complex suggests that the invariant arginine

at position 341 in MtCS is involved in several key conformational switches during the formation of a catalytically competent complex. Further the correct enzymatic function of the ternary complex seems to depend on the recognition of the EPSP phosphate group by the invariant arginine. Crystal structure analysis in the binary MtCS and ternary SpCS complexes suggests that only upon coordination of the EPSP phosphate group by Arg341/Arg337 the likewise invariant Asp343/Asp339 adopts a conformation that allows to trigger proton abstraction events leading to the formation of the product chorismate as described [Kitzing K et al., 2003; Rauch G et al., 2007].

4.3.2. Molecular dynamics simulations of CS

The structural observations in this study suggest a special role for the invariant arginine at position 341 in MtCS and 337 in SpCS. This strictly conserved residue not only participates in all key events of the enzymatic functioning of the enzyme (Fig. 4-10) but also seems to support an unproductive active site conformation (Fig. 4-9; Fig. 4-10) that is likely stabilized by the novel CS ligand NCA. Therefore the role of this arginine for the stability and energetics of ligand binding in CS was further investigated using molecular dynamics (MD) simulations. MD simulations were carried out on the ternary complex of the enzyme, as the structural investigations indicate a strong dependence of a stable ternary complex on the coordination of EPSP by the invariant arginine. In order to be able to test this ligation state as close as possible to the experimental observations, the ternary SpCS-FMN-EPSP complexes was used for MD simulations. Test simulations performed on a ternary complex of MtCS build on its homology towards SpCS proved to be considerably less stable than simulations based on SpCS. Stability of all simulations was assessed as root mean square standard deviation (RMSD) towards the energy minimized simulation starting structure. Only C α atoms involved in α -helical or β -sheet secondary structure elements were used for the RMSD calculation. The homology based tetrameric MtCS ternary complex showed an equilibrium RMSD of about 2.5 Å compared to about 1 Å for the tetrameric SpCS ternary complex. Similar was observed in a former MD study of a homology model of *Shigella flexneri* chorismate synthase (SfCS) that was based on the SpCS, HpCS and CjCS crystal structures [Zhou H et al., 2006]. In this study an equilibrium RMSD of about 3.5 Å was reached after less than one nanosecond. However, this study only used a monomer of SfCS

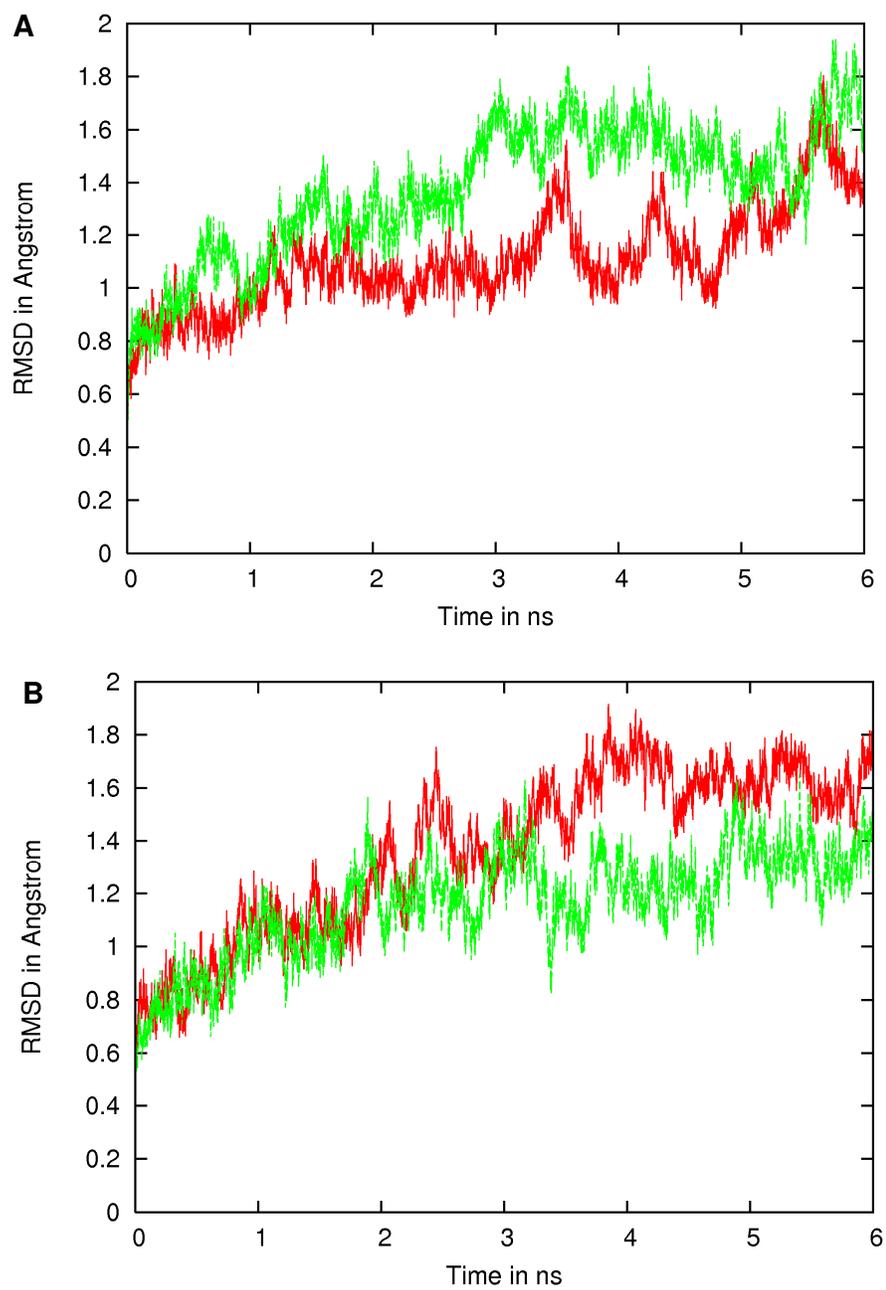


Figure 4-12: RMSD time courses of the dimeric CS simulations: A. RMSD for monomer A (red) and monomer B (green) in the CS2a simulations. B: RMSD for monomer A (red) and monomer B (green) in the CS2b simulations.

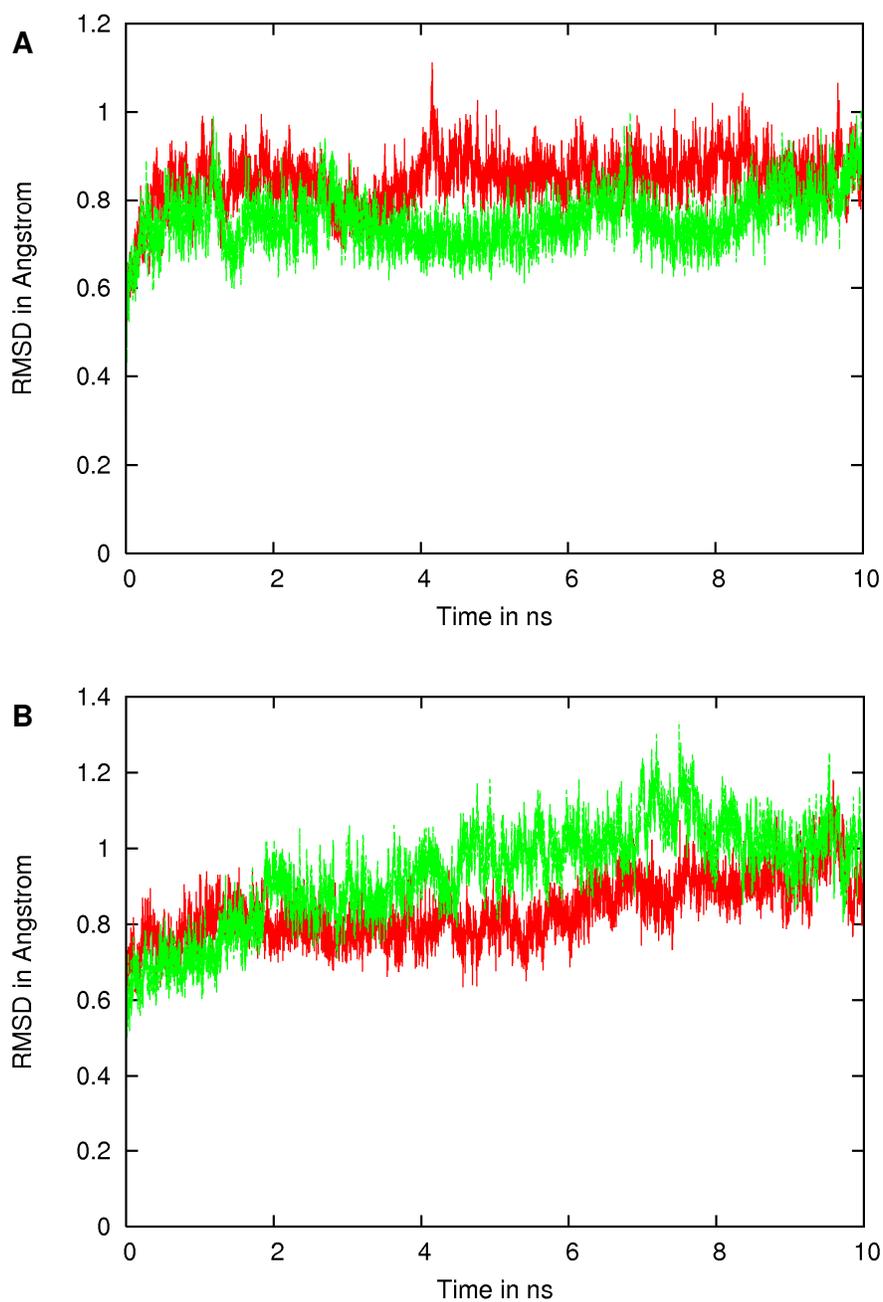


Figure 4-13: RMSD time courses of the tetrameric CS simulations: A: RMSD for the liganded monomers A (red) and B (green) in the CS_{4_{red}} simulations. B: RMSD for the unliganded monomers A (red) and B (green) in the CS_{4_{red}} simulations.

for the simulation. Therefore not only the homology modelling process but also the monomeric setup may be responsible for this particularly high RMSD. For a detailed discussion of the influence of the oligomeric state of CS on the simulation stability please see 4.3.2.1 (page 57). However, the comparison of MD simulations based on homology models and those based on the crystal structure from CS strongly indicates that all atom MD simulations of at least this enzyme are better carried out with a starting model as close to the experimental crystal structure as possible. Thus, MD simulations based on the crystal structure 1QXO are clearly favourable. In total four simulations were performed for the ternary CS complex, two as a dimer and another two as a tetramer. For the dimeric simulations chain A and B from the SpCS crystal structure 1QXO were used. This setup was submitted to simulation twice, each with the active site of either monomer A or B liganded with EPSP and FMN, while the active site of the respective other monomer was left empty. Both monomers differ mainly in the conformation of the loop L20 (corresponding to L25 in MtCS) and the loop L22 (corresponding to L27 in MtCS), the latter carrying both the invariant residues Arg337 and Asp339 (corresponding to Arg341 and Asp343 in MtCS). The conformation of the active site of monomer A represents the open conformation of L22 where Arg337 is not in contact to the EPSP phosphate group [MacLean J and Ali S, 2003]. In contrast, the active site in monomer B has a closed conformation where L22 is folded upon the two ligands and makes them inaccessible from the protein surface. The enzyme bound FMN was modeled in both of the dimeric simulations with all hydrogens, corresponding to the reduced dihydro-state of the co-factor (further called FMN_{red}) necessary for catalysis [Bornemann S, 2003] (see Fig. 4-2). In the following text the dimeric simulations based on SpCS will be called CS2_a and CS2_b, depending on whether monomer A or B was liganded with FMN and EPSP. For the tetrameric simulations all four chains of 1QXO were used. In this setup only monomers A and B were left liganded, while the ligands were removed from the active sites of C and D. The monomers C and D display a closed active site conformation similar as in monomer B. The tetrameric simulations of SpCS differed in the protonation state of FMN. In one simulation the co-factor was modeled as reduced FMN_{red}. In the other tetrameric simulation FMN was modeled in the fully oxidized state FMN_{ox}. Simulations for the tetramers were 10 ns long, while dimers were simulated for 6 ns. The tetrameric simulations based on SpCS will be called CS4_{red}, CS4_{ox}, depending on whether the co-factor was modeled as FMN_{red} or

FMN_{ox}. The former MD study of SfCS in contrast was only based on one monomer in solution and only comprised a simulation time of about 2 ns [Zhou H et al., 2006]. Therefore the here presented MD data of SpCS present the most comprehensive MD study of this enzyme up to date.

All MD simulations of the CS complex were carried out with the software package GROMACS. In contrast to the MtGluRS simulations the GROMACS port of the Amber03 force-field [Sorin EJ et al., 2005] was used with the TIP3P water model [Mahoney MW and Jorgensen WL, 2000]. As the Amber force fields do not contain parameters for EPSP and FMN, MM descriptions for the CS ligands were prepared using the software package ANTECHAMBER [Wang J et al., 2006]. Parameters for the ligands were derived from the General Amber Force Field (GAFF) [Wang J et al., 2004]. The great advantage of GAFF is its compatibility to the AMBER force field and that it has parameters for almost all organic molecules made of C, N, O, H, S, P, F, Cl, Br and I. Partial atomic charges were generated with DIVCON (QuantumBio, Inc.⁶) using the semi-empirical Austin Model 1 (AM1) method [Dewar et al, 1985] with bond-charge corrections (BCC) [Jakalian A et al., 2002]. Usually *ab initio* QM methods are preferred over semi-empirical ones for partial charge generation for MD simulations. However, recent studies suggest that AM1-BCC charges for small molecules are equally well performing as those derived from higher level QM methods with the great advantage of considerably less computational expense [Jakalian A et al., 2002; Mobley DL et al., 2007]. MD simulations of CS complexes were performed at 300 K and 1 bar in the NPT ensemble with a coupling time of 0.1 ps. All bonds were constrained with the LINCS algorithm and an integration time step of 2 fs was used. Cut-offs for Lennard-Jones and electrostatic interactions were 10 Å and 9 Å, respectively; long-range electrostatic interactions were calculated by particle-mesh Ewald summation. For the simulation setup each starting model was immersed in a dodecahedral water box with the box edges 8 Å away from the protein surface and counter-ions were added to the system to neutralize all charges. Each system (i.e. enzyme-ligands complex, water and ions) was subjected to steepest descent energy minimization until convergence of the maximum force below 1000 kJ mol⁻¹ nm⁻¹. This was followed by a 1 ns MD simulation at the target temperature using harmonic position restraints on the heavy atoms of the protein with a force

⁶ <http://www.quantumbioinc.com/>

constant of $k = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ to thoroughly equilibrate water and ions. Subsequently, each system was submitted to an unconstrained MD simulation between 6 and 10 ns.

4.3.2.1. Dimeric vs. tetrameric CS simulations

Comparison of the dimeric and tetrameric simulations should give insight on whether a stable ternary CS complex depends on a specific oligomerization state of the enzyme. For *E. coli* CS (EcCS) it has been proposed that the enzyme exists in solution as a tetramer [White PJ et al., 1988; Macheroux P et al., 1998], while for MtCS and SpCS also the dimeric state was observed [Dias MVB et al., 2006; MacLean J and Ali S, 2003]. While for the dimer – tetramer equilibrium a dissociation constant (K_d) in SpCS of only $0.8 \mu\text{M}$ was observed [MacLean J and Ali S, 2003], the observed K_d in the MtCS case was about 50 M [Dias MVB et al., 2006]. Despite the amazing differences of the two K_d values, the question might be posed whether a dimeric CS is capable of forming a catalytically competent ternary complex. A stable monomeric ternary complex can be excluded as the FMN binding site is only properly formed in the dimeric state of the enzyme (see also 4.2.). This was further proven by test simulations with a monomeric setup (data not shown). For comparison of the dimeric and tetrameric simulations only CS2_a, CS2_b and CS4_{red} will be used as in all three simulations FMN was modelled in its reduced state.

In the CS2 simulations both monomers A and B show comparable stabilities throughout the entire simulation. This can be seen by the time course of the root mean square deviation (RMSD) from the initial structure (Fig. 4-12a/b). After an initial equilibration period of about 2 ns the liganded monomers A and B fluctuate in the CS2 simulations stably between RMSD values of 1 and 1.4 \AA . Both monomers show a somewhat better stability in the liganded than in the unliganded state, the latter exhibiting a slightly elevated equilibrium RMSD of about 1.6 \AA . Furthermore, the equilibration period for the unliganded monomers is longer with about 3 - 4 ns. In contrast the CS4_{red} simulation shows substantially decreased RMSD values throughout the entire simulation for the liganded monomers (Fig. 4-13a). Both equilibrate very fast in about 1 ns and RMSD values stay close to 0.8 \AA throughout the entire simulation time. Additionally the RMSD values fluctuate considerably less

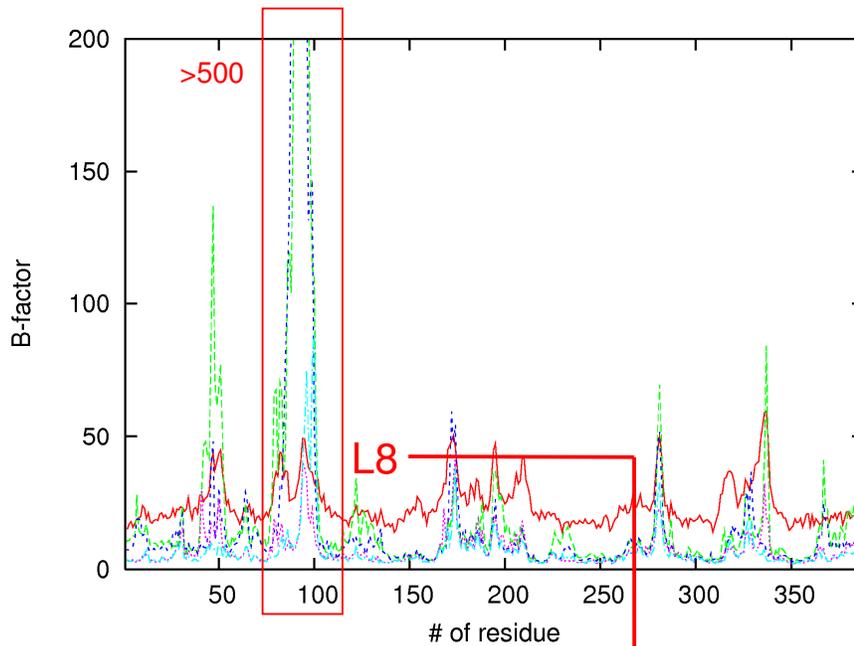
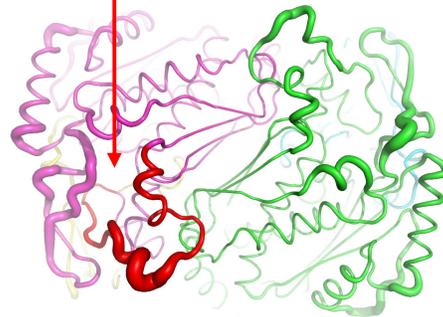


Figure 4-14: A. Comparison of B-factors derived from the SpCS crystal structure (PDB-ID 1QXO) and from the CS2 simulations: 1QXO (red); CS2_a liganded monomer (green); CS2_b liganded monomer (blue); CS4_{red} liganded monomer A (magenta); CS4_{red} liganded monomer B (cyan). The highlighted region corresponds to loop L8 that shows B-factors values > 500 in the CS2 simulations. B. Main-chain of the two monomers A and B of the CS2 simulations shown in purple and green, respectively. The thickness of the chain indicates the level of fluctuation according to the B-factors from the MD simulations. L8 of monomer A lies at the interface between the monomers and is shown in red.



about their average values in the CS4_{red} simulation. This demonstrates that the tetrameric state of the CS-FMN-EPSP complex displays a significantly higher stability as the dimeric state. The unliganded monomers of the CS4 simulations display a somewhat higher RMSD time course with an equilibrium value of about 1 – 1.2 Å (Fig. 4-13b). The observation from both dimeric and tetrameric simulations that unliganded monomers deviate stronger from the liganded starting model points to a stabilizing effect of the ligands on the overall shape of the protein. Small angle scattering observations of tetrameric CS in solution have shown before that ligand binding has a slight influence on the overall shape of the complex [Macheroux P et al., 1998] rendering it more compact.

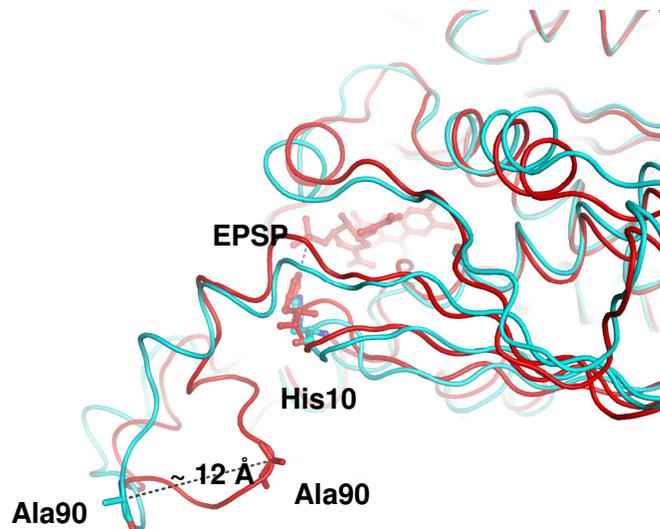


Figure 4-15: Conformational changes of L8 in the CS2 MD simulations: The CS2 starting model (red); conformation of CS in the CS2 simulation at 6ns (blue). See text for details

Comparison of the time averaged RMS fluctuations (RMSF) of the C α atoms shows that all liganded monomers both in the CS dimer and tetramer show essentially the same behaviour as the respective monomers A and B from the SpCS crystal structure 1QXO (Fig. 4-14). However there is a segment in the liganded monomers of the CS2 simulations ranging from residue 80 to 110 (i.e. loop L8) that shows enormously high fluctuations. The loop L8 is at the interface of two dimers that form the tetramer and already shows elevated B-factors in the crystal structure of SpCS (Fig. 4-14). In the CS2 simulations L8 shows large B-factor values most likely due to the missing contacts to the second dimer. The loop L8 adjoins a short 3_{10} -helix that cradles part of the loop L1 containing the invariant His10 (His11 in MtCS). This residue participates in the charge transfer events during catalysis [Kitzing K et al, 2003]. Due to the missing dimer – dimer contacts of L8, the loop shows a large conformational change that causes an outward oriented rotation, leading to displacements of up to 12 Å for single residues (Fig. 4-15). This is resulting in a loss of the contacts to L1 that in turn leads in both CS2 simulations to the almost immediate disruption of contacts between His11 and the EPSP phosphate group present in 1QXO. As this contact is crucial for catalysis [Kitzing K et al., 2003; Rauch G et al., 2007] a tetrameric organization of CS seems to be mandatory for a stable formation of a catalytically competent ternary complex. Only the tetrameric simulations show a stable contact between the EPSP phosphate group and His10

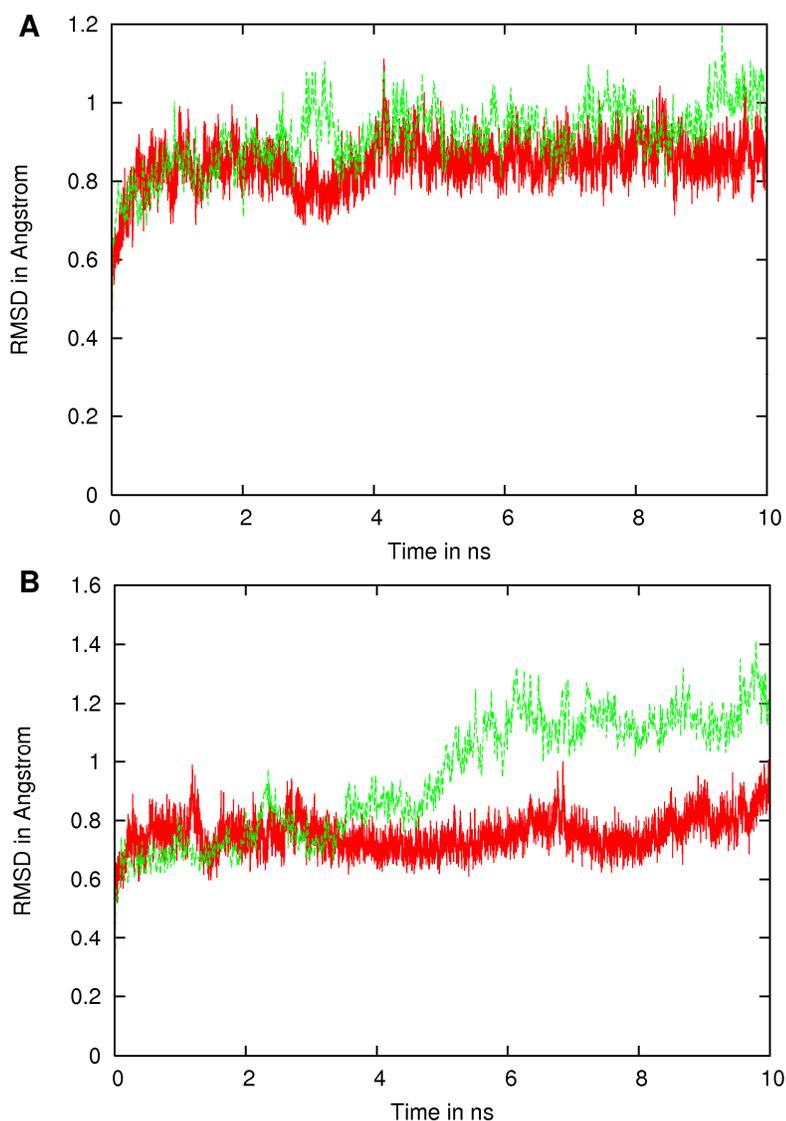


Figure 4-16: Backbone RMSD time courses of the simulations of CS4_{red} (red) and CS4_{ox} (green). A. Liganded monomers A; B. Liganded monomers B

throughout the entire simulation. Therefore, in the dimer a catalytically competent binding site can not be established regardless of the active site being in an open or closed state as observed in crystal structures. Biochemical data shows, that CS predominantly exists in a tetrameric state [White PJ et al., 1988; Macheroux P et al., 1998] and the observed K_d for the dimer – tetramer equilibrium for SpCS was very low with 0.8 μM [MacLean J and Ali S, 2003]. Thus, the tetrameric state of CS dominates and ensures the correct formation of the ternary complex upon ligand binding.

4.3.2.2. Binding site stability in the CS4 simulations

The CS2 simulations prove that a dimeric state of CS clearly does not agree with the description of a catalytically competent ternary complex of the enzyme. Thus the further discussion of ligand binding stability will focus solely on the CS4 simulations. FMN was modeled in the reduced (FMN_{red}) and oxidized form (FMN_{ox}) in the different CS4 simulations. In the ternary complex of SpCS the co-factor FMN was assumed to be bound in its oxidized state [MacLean J and Ali S, 2003]. Since the reduced co-factor is necessary for catalysis [Bornemann S, 2003], it is important to consider whether reduction of enzyme-bound FMN will involve a conformational change of the active site and the present ligands. Therefore FMN was modeled in both states and submitted to MD simulations in an otherwise unchanged setup. The liganded monomers A display both in the CS4_{red} and CS4_{ox} simulation similar stabilities with RMSD time courses that stay well below 1 Å (Fig. 4-16a). In contrast the RMSD of monomer B in the CS4_{ox} simulation shows an increase of its RMSD from 0.8 Å to about 1.2 Å between 5 and 6 ns, while in the CS4_{red} simulation the RMSD fluctuates steadily around 0.8 Å during the entire simulation (Fig. 4-16b). This increase goes along with conformational changes for EPSP and FMN in the closed active site of monomer B. During the first 4 ns EPSP is bound in a stable conformation showing an RMSD of only about 0.2 Å towards the starting conformation (Fig 4-17a). Around 4 ns the RMSD of EPSP increases first to about 0.6 Å and at 5 ns suddenly shows a jump to 1.2 Å. The co-factor FMN displays almost from the beginning of the simulation an RMSD of about 1 Å. Coinciding with the sudden change of the conformation of EPSP, FMN displays a similar jump in its RMSD course to about 2 Å. Notably, the conformational changes of FMN in the closed active site involve only the ribityl-chain but not the isoalloxazine moiety (Fig. 4-17b). This involves an unbinding motion of FMN; the co-factor starts to move towards the protein surface, leaving the deeply buried binding pocket it occupies at the beginning. Moreover, EPSP is involved in a similar motion leaving the active site of monomer B that is involved in a closed to open transition with an endpoint similar to the open conformation in monomer A. In monomer A detachment of EPSP from the active site can be observed, too, but particularly FMN shows only a minor change in conformation. Significant in this respect is that FMN_{red} and EPSP show almost no conformational changes of their

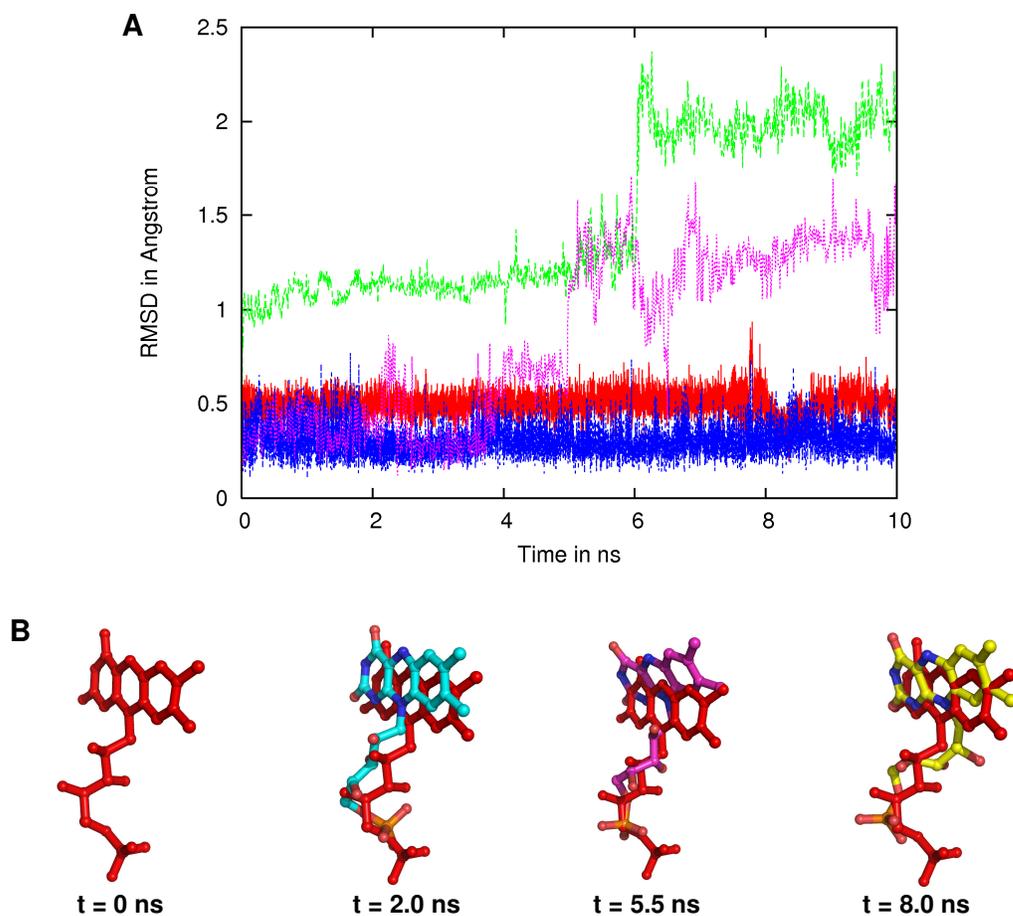


Figure 4-17: A: RMSD time courses of EPSP and FMN in the closed active site of the CS4_{red} and CS4_{ox} simulations. FMN_{red} (red); FMN_{ox} (green); EPSP in CS4_{red} (blue); and EPSP in CS4_{ox} (magenta) B: FMN binding modes in monomer B of the CS4_{ox} simulation; FMN_{red} at time 0 ns is shown in red

respective binding mode in monomer B of the CS4_{red} simulation (Fig. 4-17a). Both ligands show RMSD values of only 0.2 - 0.5 Å throughout the entire simulation. Therefore the binding mode of the two ligands is virtually unchanged in the reduced state of FMN while the oxidized state involves larger conformational changes in the binding mode of both ligands. This suggests that the active site conformation in the crystal structure 1QXO for FMN might be energetically more favourable for the reduced form of the co-factor. However, the FMN bound in the crystal structure 1QXO is expected to be bound in its oxidized state [MacLean J and Ali S, 2003]. Yet, the crystal containing FMN was exposed for a long time to the strong radiation from a synchrotron source during the measurement. As the strong radiation might exhibit a reducing effect on the co-factor, FMN might be bound in a reduced state in the crystal structure, as indicated by the MD simulations in this work. This however would leave

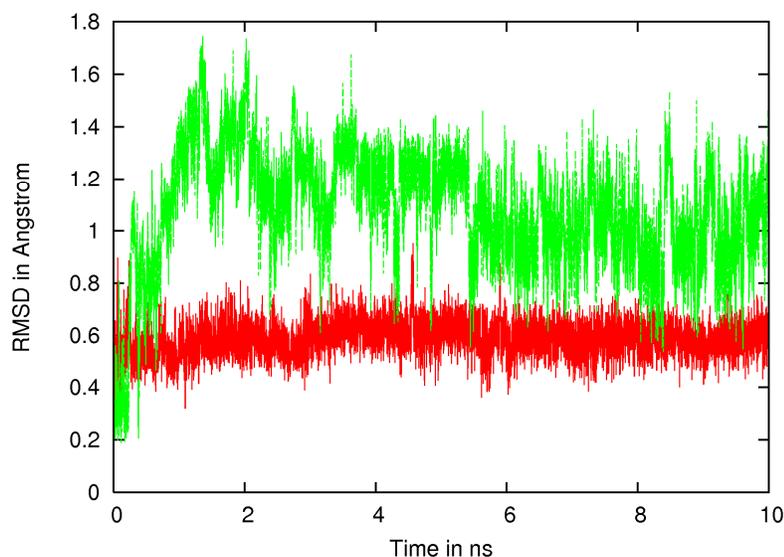


Figure 4-18: RMSD time courses of FMN (red) and EPSP (green) in monomer A of the CS4_{red} simulation.

open the question, why FMN upon reduction by the synchrotron radiation does not participate in chorismate formation from EPSP in the crystal.

In the liganded monomer A of CS4_{red} containing the open state of the active site a conformational deviation is only observed for EPSP and not for FMN (Fig. 4-18). Here EPSP shows a rapid increase of the RMSD to almost 1.6 Å within the first 1.5 ns. This deviation decreases to an equilibrium value of about 1 Å and generally describes a loss of the proper interactions with FMN. During the first 1.5 ns EPSP loses almost all of its initial contacts in the open active site and starts to drift away from FMN. Visual inspection of the trajectories of the liganded monomer A in the CS4_{red} simulation shows the establishment of contacts between the EPSP phosphate group and Arg337 at around 4 ns. Simultaneously L22 folds into a conformation that is similar to that in the closed state of the active site of monomer B (Fig 4-19). Therefore a transition from the open to the closed state of L22 seems to coincide with formation of hydrogen bonds between Arg337 and the phosphate of EPSP. This conformational change is only of transient nature in CS4_{red}, where contacts are lost again at around 7 ns. During the initial contact formation between the EPSP phosphate and Arg337 EPSP is already far apart from its initial position and therefore from FMN. This might be the reason for the transient closing movement of L22. Because EPSP is not properly positioned next to FMN the closure of L22 upon the binding site might not fully occur. In the average structure calculated for the CS4_{red}

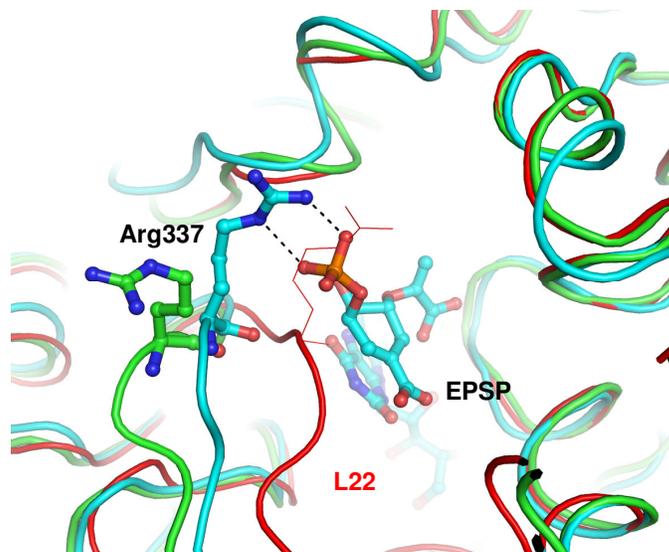


Figure 4-19: Transient closure of L22 in monomer A during the CS4_{red} simulation. Closed conformation of L22 in chain B of 1QXO is shown in red and the starting conformation of monomer A (i.e. 1QXO chain A) of the CS4_{red} simulation in green. Conformation of L22 at 3.4 ns is shown in blue, hydrogen bonding contact between Arg337 and EPSP is indicated by broken black lines.

simulation EPSP is only loosely bound to the backbone atoms of Ala133 and Arg134 at the active site. This dynamical behavior of EPSP in the open active site might shed light on a possible entry route for EPSP to the binary complex. For this, EPSP would be attracted to the active site and may loosely attach somewhere near it. Subsequently established contacts to the side chain of Arg337 then might induce binding events during which EPSP descends into the binding site and makes its proper contacts to FMN. This is indicated by the decrease of the RMSD time course for EPSP. However, while the detachment of EPSP from the binding site occurs on a very small time scale of about 2 ns, binding events are likely happening on a much longer time scale as monitored during this simulation.

In contrast the liganded monomer B does not display a transition of the active site conformation. The active site in monomer B stays in the closed conformation throughout the entire simulation. RMSD time courses of FMN show that the conformation of the bound co-factor is virtually unchanged throughout the CS4_{red} simulation regardless of the active site adopting an open or closed conformation (Fig. 4-17a; Fig. 4-18). In both cases the conformation of FMN shows an equilibrium RMSD of only 0.5 – 0.6 Å. Further, in the closed active site EPSP shows RMSD values of only 0.2 Å throughout the entire simulation time. Ligand - enzyme contacts

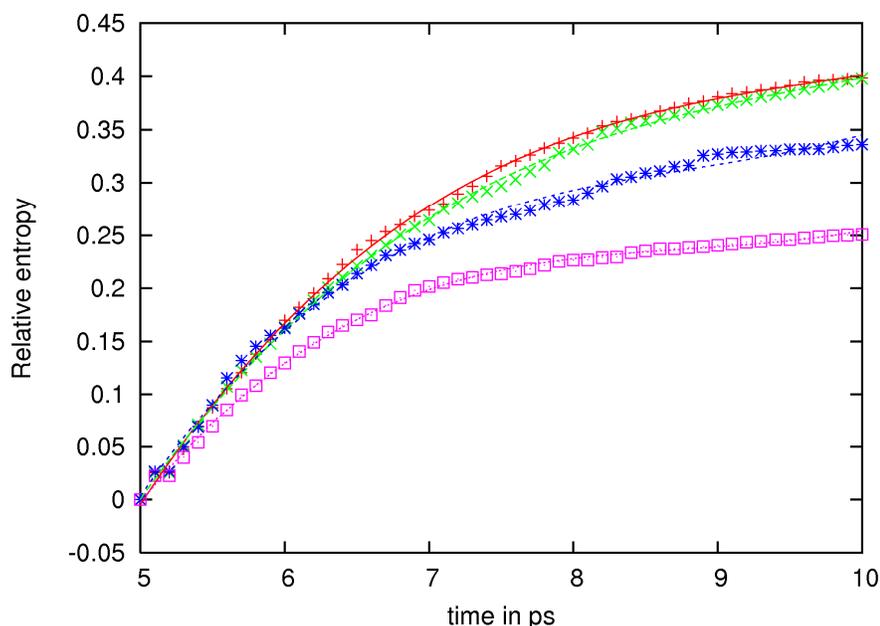


Figure 4-20: Relative entropies of EPSP in open (blue) and closed (magenta) active sites and of FMN in open (red) and closed (green) active sites.

in the binding site of monomer B prove to be highly stable in the CS4_{red} simulation. The main difference in the starting conformation of the liganded monomers is the missing contact between Arg337 in monomer A with respect to monomer B. The MD simulations clearly show that the missing contacts render the ternary complex unstable and that a closed conformation of L22 is mandatory for proper ligand binding. This can also be observed in the energetics of ligand binding. For the last five nano seconds of the simulation the configurational entropy was calculated for the open and closed configurations of SpCS liganded with FMN and EPSP, as well as for the ligands FMN and EPSP in the open and closed active sites separately. All calculations were based on a quasi-harmonic entropy approximation which relates the entropy to the determinant of a mass-weighted covariance matrix [Schäfer H et al., 2000; Andricioaei I and Karplus M, 2001]. To do so, eigenvalues of the mass-weighted covariance matrix were calculated for the entropy approximation. While all atoms were used for the ligands, for the protein the covariance matrix was constructed only from the C α atoms. Relative entropies were achieved by dividing the entropies for the ligands by the entropies of the respective liganded complexes. Calculation of the relative configurational entropies of the ligands for the last 5 ns clearly shows decreased entropy for EPSP in the closed active site (Fig. 4-20). In contrast, the entropy for FMN is similar in both open and closed active site

conformations. Therefore the closed active site conformation and the contacts of Arg341 to the EPSP phosphate group seem to be energetically more favourable than the open active site conformation for EPSP binding, while it does not influence the entropy of the bound FMN.

4.3.3. Interim summary

The structural analysis of MtCS and SpCS and the molecular dynamics simulations of SpCS imply the following for the enzymatic mechanism:

- Comparison of the MtCS structures presented in this work and the structure of SpCS allows for the first time to present a sequential ligand binding cycle for the enzymatic function that might be generally applicable to CS.
- This binding cycle mainly involves movements of the surface loops L4, L25 and L27 (numbering according to MtCS, see also Fig. 4-4b), of which L27 carries an invariant arginine (Arg341 in MtCS and Arg337 in SpCS) that is involved in all key conformational switches during ternary complex formation.
- The invariant arginine is suggested to be essential for the enzymatic mechanism as it mainly participates in the establishment of the closed active site in the ternary complex and might trigger the catalytic reaction by fine tuning the orientation of an invariant aspartate (Asp343 in MtCS and Asp339 in SpCS) for participation in proton abstraction from the N(5) of FMN.
- Molecular dynamics (MD) calculations of the ternary SpCS complex validate the importance of the invariant arginine for the correct formation of a catalytically competent ternary complex of CS and that additionally a tetrameric state of CS might be necessary for the correct enzymatic function of the enzyme.

4.4. Discussion

Structural analysis on *M. tuberculosis* chorismate synthase was carried out using X-ray crystallographic studies combined with molecular dynamics (MD) simulations.

High resolution structures of the enzyme in unliganded and liganded states and comparison to a highly homologue crystal structure allowed for the first time the detailed description of a ligand binding cycle that is suggested to be generally applicable to CS (Fig. 4-10). This binding cycle describes all the key conformational events associated with ligand binding to chorismate synthase and describes the role of several invariant residues present in all CS. Notably, the presented structural data point to the dependency of all key events of sequential ligand on conformational changes of a special invariant arginine (Arg341 in MtCS and Arg337 in SpCS; see also Fig. 4-4). This arginine is of clear necessity for the correct enzymatic mechanism by fine-tuning the active site for catalysis during formation of contacts to EPSP (Fig. 4-11b/c). Further investigations with the help of MD simulations clearly underlined the importance of this residue for the formation of a catalytically competent ternary enzyme complex. The role of the invariant arginine 341 in MtCS is of special importance considering a novel ligation state with nicotinamide that this work presents for the first time. Structural data indicate that the binding mode of NCA might support an unproductive conformation of the active site in the apo form that involves Arg341 and therefore a side route to the formation of the binary complex is provided. However, the possibility exists to replace NCA by FMN in MtCS crystals by soaking (Fig. 4-10b/d).

Yet, the observed formation of a stable MtCS-NCA complex may be of further interest considering the known antimicrobial activity of nicotinamide [Chorine V, 1945]. NCA previously was used as an anti-TB drug, until antagonism between NCA and the first-line drug isoniazid was noticed [Murray MF, 2003]. Combined treatment with NCA and the analogue pyrazinamide (PZA), another first-line drug, may cause cross-resistance [Somoskovi A et al., 2001] due to the fact that both compounds are substrates of the mycobacterial enzyme nicotinamidase/pyrazinamidase (PZAse), which converts both prodrugs into the active forms nicotinic acid and pyrazinoic acid (POA). Recently, evidence has been provided for anti-HIV activity of nicotinamide, suggesting its potential use both as an antimicrobial and an antiretroviral agent [Murray MF, 2003]. However, the mechanism of the antimicrobial action of nicotinamide has not been elucidated until now. The observation presented in this work that NCA firmly binds to chorismate synthase suggests that this influence on the shikimate pathway of *M. tuberculosis* may be part of the antimicrobial activity of NCA.

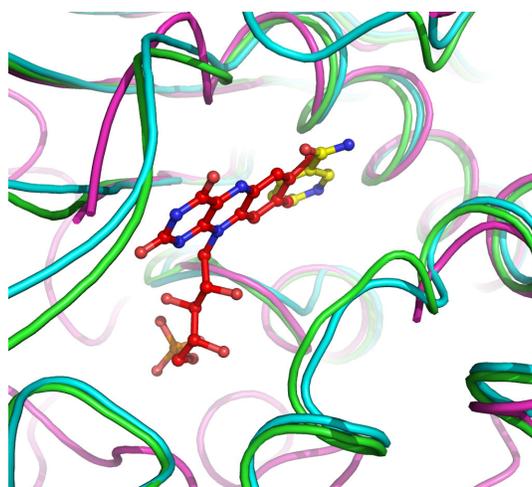


Figure 4-21: Superposition of mono- and bifunctional CS. Bifunctional ScCS (PDB-ID 1R52) is shown in purple, MtCS-NCA in green and MtCS-FMN in cyan; FMN is colored in red and NCA in yellow.

Notably, the MtCS-NCA complex displayed substantially higher stability in acid medium (pH ~ 5) than at conditions near neutral pH, as shown by the crystallization studies for this work. Notably, acidic pH characterizes the environment of mycobacteria in macrophages. Similarly it has been shown that the analogous compound PZA requires acidic pH values to be active against *M. tuberculosis* [Mitchison DA, 1985]. Yet, the observation that *M. tuberculosis* strains lacking PZAse are resistant to both PZA and NCA [Konno K et al., 1967] appears to contradict the hypothesis that the interaction of NCA with chorismate synthase may represent a major contribution to the antimicrobial activity of NCA. Alternatively, there is the possibility that partial interruption of the shikimate pathway in mycobacteria ingested by phagocytes might be bypassed by the uptake of aromatic amino acids from the host. However, resistance against PZA or NCA may not necessarily be related to the absence or mutations of PZAse [Martin A et al., 2006; Singh P et al., 2006]. In total, the results presented in this work suggest taking the effect on chorismate production into account when further investigating the anti-TB activities of NCA and PZA.

The orientation of NCA in the complex with MtCS does not provide a model for the yet unknown orientation of the nicotinamide moiety of NADPH in interactions with bifunctional chorismate synthases. Crystal structures (1R52, 1R53) of the bifunctional chorismate synthase from *Saccharomyces cerevisiae* in unliganded form [Quevillon-

Cheruel S. et al., 2004] show a closely similar conformation as MtCS in the binding regions of FMN and NCA (Fig. 4-21). Constructing a meaningful structural model of possible NADPH binding to ScCS is not feasible due to complete disorder for 23% of the amino acids in 1R52 and 1R53. However, a superposition of NADPH onto NCA in the MtCS-NCA complex shows that binding of NADPH with a similar orientation of the nicotinamide moiety would require major intra-domain conformational changes of an extended part of the enzyme (Fig. 4-22a). Similar is true for ScCS. This possibility may be excluded considering the fact that crystal structures of other enzymes exhibiting interactions between FMN and NADPH ligands (e.g. 1HE4) show a different orientation, allowing the binding of the nicotinamide moiety with respect to the isoalloxazine ring of FMN in a similar orientation as EPSP (Fig. 4-22b). Based on spectroscopic and kinetic characterizations a similar arrangement has been proposed in earlier studies on bifunctional enzymes [Kitzing K et al., 2001] .

The binding mode of NCA proposes a way of stabilizing the unproductive active site conformation in MtCS and may provide a starting point for possible rational design of novel chorismate synthase inhibitors. This newly observed unproductive active site conformation is strongly dependent on the interaction of NCA with an invariant arginine that in turn prevents the correct enzymatic action of CS due to an interaction

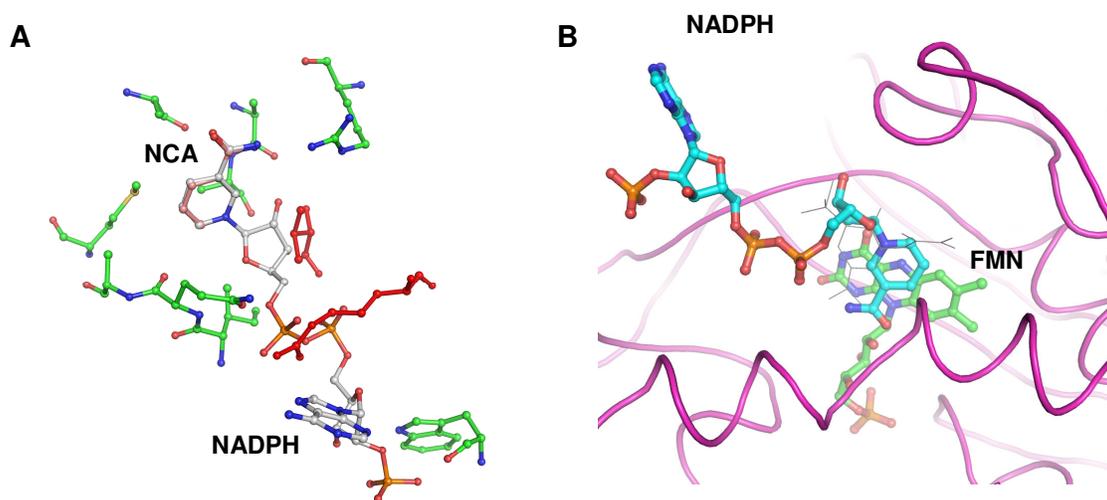


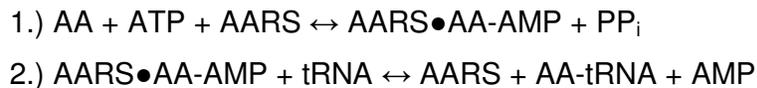
Figure 4-22: NADPH binding model in CS. A. Superposition of NADPH (grey) onto NCA (wheat) in the MtCS-NCA complex (green). Binding of NADPH would require major intra-domain conformational changes of an extended part of the enzyme. Clashing residues are shown red. B. Interactions between FMN and NADPH ligands in the crystal structure of 1HE4 allow the binding of the nicotinamide moiety with respect to the isoalloxazine ring of FMN in a similar orientation as EPSP

with an invariant aspartate. The mentioned arginine is likely involved in all key conformational changes of the enzymatic functioning and therefore may play a central role in ternary complex formation during the catalytic process. This special role of the invariant arginine has been validated by the application of MD simulations. Therefore the structural and functional investigations presented on chorismate synthase in this work not only give detailed new insights into the biological function but also provide a basis for rational drug design studies on this pharmaceutically important enzyme.

5. GlytamyI-tRNA synthetase from *M. tuberculosis*

5.1. Role of AARSs in protein synthesis

Protein biosynthesis is the process of translating a messenger RNA (mRNA) into a protein sequence [Crick F, 1970; reviewed by e.g. Ibba M and Söll D, 1999]. This happens at the ribosomes, where the codons of the mRNA (except for stop-codons) are paired with the specific anticodons of transfer RNAs. (Fig. 5-1). Any utilized transfer RNA (tRNA) in turn is specifically charged with the amino acid that is encoded by the matching codon of the mRNA. As the amino acid is incorporated in the growing peptide chain, the fidelity of correct protein biosynthesis is strongly dependent on two processes: a) codon-anticodon recognition [reviewed by Yarus M and Smith D, 1995] and b) aminoacyl-tRNA synthesis [Ibba M et al., 1997]. In this work the focus is on the latter case, where correct charging of a given tRNA is carried out by an aminoacyl-tRNA synthetase (AARS). The catalysis of aminoacyl-tRNA synthesis involves two steps that are essentially the same for all AARSs [reviewed by Arnez JG and Moras D, 1997]:



First the aminoacid (AA) and ATP, or rarely a different nucleotide, bind to the protein. In the enzyme bound state the α -carboxylate of the amino acid attacks the α -phosphate of ATP by an in-line nucleophilic displacement mechanism. Subsequently the resulting protein-bound aminoacyl-adenylate (AA-AMP) is nucleophilically attacked by the 2'- or 3'-hydroxyl of the terminal adenosine of the tRNA. This results in the formation of an aminoacyl-tRNA (AA-tRNA) that participates in protein biosynthesis.

The AARSs are a large and diverse family of enzymes. Each single AARS is able to specifically attach a particular amino acid to the corresponding tRNA. As full length names for AARSs can be quite long, usually an acronym is assigned to any given enzyme. This is formed by adding the three letter code of the utilized amino acid as a prefix to RS. Similar is used for the corresponding tRNA, to which the three letter

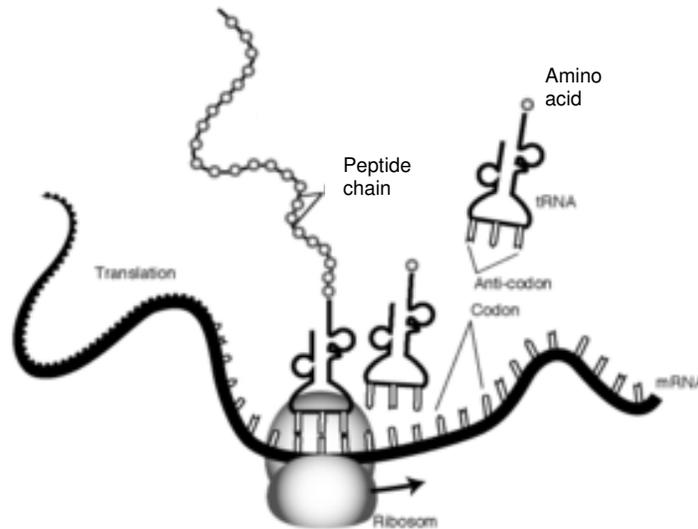


Figure 5-1: Protein biosynthesis at the ribosomes is depending on correctly charged tRNAs for peptide chain elongation (taken from <http://www.biocrawler.com/>)

code is added as a suffix. For example glutamyl-tRNA synthetase charges glutamate onto the glutamyl-tRNA and hence the enzyme is abbreviated GluRS and the tRNA gets the acronym tRNA^{Glu} . A charged tRNA^{Glu} would additionally get a prefix for the added amino acid and would thus be called $\text{Glu-tRNA}^{\text{Glu}}$. Generally the assignment of one AARS to one amino acid and one tRNA would imply that most cells make twenty different AARSs, one for each amino acid. However genomic studies have shown that some organisms do not have genes for twenty different AARSs [Bult CJ et al., 1996]. Since these organisms are still using twenty different amino acids for protein biosynthesis, there are other ways of preparing those tRNAs without a specific AARS counterpart [Tumbula D et al., 1999]. For example in some organisms glutamyl-tRNA (tRNA^{Gln}) is first charged with glutamate, giving mis-acylated $\text{Glu-tRNA}^{\text{Gln}}$. Subsequently amidotransferases modify the glutamyl-moiety to glutaminyl giving properly charged $\text{Gln-tRNA}^{\text{Gln}}$ [Ibba et al., 2000] (see also Fig. 5-4).

All AARSs are grouped into two classes that share the same reaction described above. Differences between the two classes include their quaternary structure and the way they attach amino acids to the tRNA [Ibba M and Söll D, 2000]. Class I enzymes are (with exceptions) monomeric and attach the amino acid to the 2'-hydroxy group of the respective tRNA, while class II enzymes are almost exclusively dimeric or tetrameric and charge the amino acid onto the 3'-hydroxy group of the

tRNA. Further, division into the two classes is based on mutually exclusive sequence motifs, reflecting distinct active site topologies [Eriani G et al., 1990]. As shown by structural studies, the active site of class I AARSs contains a Rossman fold. This motif is normally seen in dinucleotide binding proteins and contains three parallel beta-sheets connected by two alpha-helices in a β - α - β - α - β topological arrangement (Fig. 5-2a). Class II AARSs do not carry this motif and instead contain a novel anti-parallel β -fold (Fig. 5-2b). The different nucleotide binding folds entail a difference in the way ATP is bound in the active site. While class I enzymes bind ATP in an extended conformation, the nucleotide is observed in a bent conformation in class II enzymes (Fig. 5-2).

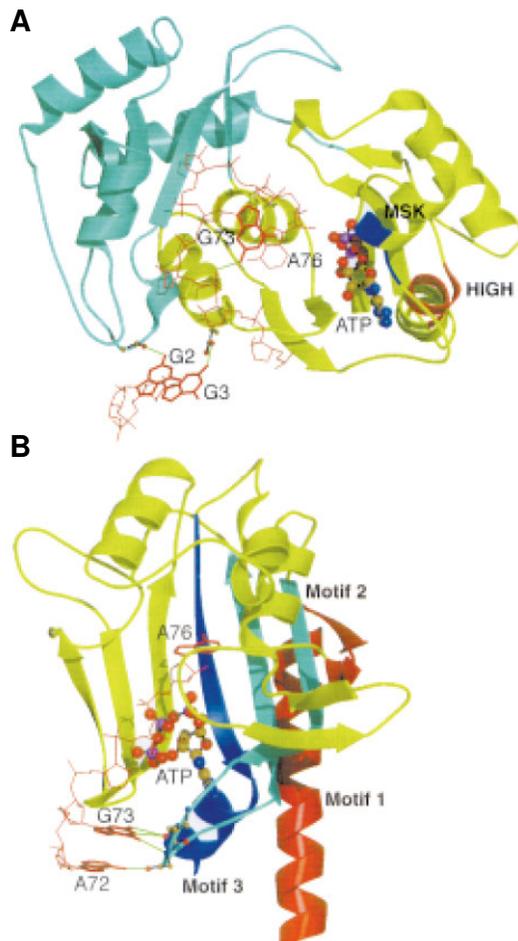


Figure 5-2: Common tRNA motifs. A. The active site of class I AARSs contains the Rossman fold consisting of three parallel beta-sheets connected by two alpha-helices in a β - α - β - α - β topological arrangement. ATP is seen in an extended conformation. B. Class II AARSs do not carry this motif and instead contain a novel anti-parallel β -fold. ATP is observed in a bent conformation. (Taken from Arnez JG and Moras D, 1997)

Whereas the amino acid binding motifs of AARSs are well conserved between the members of one class, generalizations about tRNA binding modes are complicated. This is due to the absence of common RNA-binding motifs in the two classes. Still they differ in the way they approach the tRNA. Class I enzymes approach the acceptor stem of the tRNA from the minor groove side and class II enzymes approach the tRNA from the major groove side. While in the former case the variable loop is facing the solvent, it is facing the protein in the latter case. Yet it is not clear whether these differences are true for all enzymes, as the observations are strictly based on structural studies and only the minority of AARS structures was solved together with its cognate tRNA [Ibba M and Söll D, 2000].

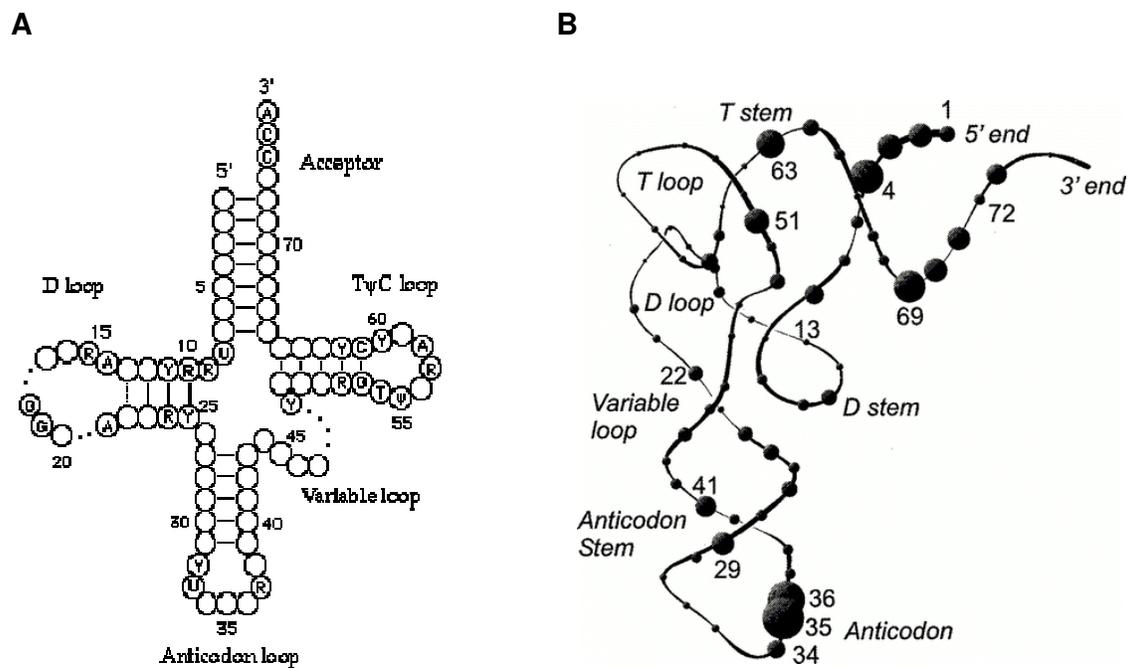


Figure 5-3: Common tRNA representations. A. Cloverleaf representation (taken from <http://bass.bio.uci.edu>). B. L-shaped conformation. Sphere size indicates observed frequencies of nucleotide-AARS interactions. The important structural elements are shown. (Taken from McClain WH and Nicholas HB, 1987)

As any given AARS must be able to recognize its cognate tRNA from a cellular pool of similar molecules, this aspect of aminoacyl-tRNA synthesis has been the subject of intensive research [reviewed e.g. by Saks ME et al., 1994]. Single tRNA molecules are around 70-80 nucleotides long and their secondary structure may be displayed in a clover-leaf presentation showing the five important structural segments (Fig. 5-3a):

- 1.) Acceptor stem
- 2.) D loop
- 3.) Anticodon loop
- 4.) Variable loop
- 5.) TΨC loop

The segments connecting the loops of the cloverleaf are called stems and are stabilised by Watson-Crick base pairing. The acceptor stem is the only one that does not connect to a loop and instead contains both the 3'- and 5'-ends of the tRNA. At the 3'-end of any tRNA the CCA motif is found, where the amino acid is attached to. Due to interactions between the variable loop and the D stem and between the D and the TΨC loop the tRNA folds into an L-shaped molecule in solution (Fig. 5-3b). This L-shaped topology allows the AARS to make contact to the tRNA about its entire length, in this way recognizing both the 3' end of the acceptor stem as well as the anticodon region. This ensures the specific recognition of the proper tRNA. To find out which contacts are the most important for the mutual recognition of tRNA and AARS, systematic mutation studies have been carried out by replacing nucleotides in the tRNA and subsequent characterisation of their aminoacylation competence by AARSs [Saks ME et al., 1994]. These studies made it possible to generate identity sets of nucleotides that are most important for the correct interaction of tRNAs and AARSs. Most often the AARSs interact specifically with the discriminator base (N73), the acceptor stem and the anticodon (Fig 5-3b) [Ibba M and Söll D, 2000]. Yet, there can be other strong determinants for recognition including other general structural elements like the D stem [Senger B et al., 1995] or structural elements that are unique to specific tRNAs such as a variety of modified nucleotides [Björk GR, 1995]. Further some tRNAs carry anti-determinants that prevent interaction with non-cognate AARSs [Giegé R et al., 1998]. It becomes clear that through the intricate, numerous possibilities of contacts between tRNA and AARS the charging of only the cognate tRNA by its assigned AARS is guaranteed. Interestingly, by "transplanting" entire identity sets from one tRNA to the other it is possible to entirely change its specificity from one AARS to another [Giegé R et al., 1993]. Notably, sometimes this is possible by changing only so much as a single nucleotide [Saks ME et al., 1998].

The introduction on the role of AARS clearly shows why aminoacyl-tRNA synthesis displays a high fidelity. As a matter of fact AARSs have an overall error rate of only about 1 out of 10,000 [Yarus M, 1979; Kurland CG, 1992]. While this is already guaranteed by the mechanisms described above, further proof-reading is employed to strengthen the fidelity. This is especially important in the case of amino acid discrimination and binding, as these small molecules are significantly less complex than the tRNAs. However, this work mainly focuses on the recognition of a tRNA by its cognate AARS. Therefore the aminoacylation and proofreading concepts are out of the scope. Interested readers may be referenced to reviews from e.g. Jakubowki H and Goldmann E., 1992.

5.2. Relaxed specificity and non-discriminating GluRS

As indicated in 5.1. exceptions from the rule of one AARS per tRNA exist. Examples for these non-canonical AARSs are AspRS and GluRS that both display relaxed specificity towards their cognate tRNAs [Ibba M and Söll D, 2000]. While in several organisms both GluRS and GlnRS exist, that catalyze the attachment of glutamate and glutamine, respectively, onto their corresponding tRNAs, several organisms do not contain GlnRS [LaPointe J et al., 1986; Schön A et al., 1988; Gagnon Y et al., 1996]. Similar is true for AspRS and AsnRS. Therefore in the case of non-existent AsnRS/GlnRS both mis-aminoacylated Asp-tRNA^{Asn} and Glu-tRNA^{Gln} are produced by a special AspRS and GluRS, respectively (Fig. 5-4). Yet, these mis-charged tRNAs are not proof-read by the AARSs themselves. Still they are not used in protein biosynthesis as they are not recognized by the translation elongation factor [Stanzel M et al., 1994]. Synthesis of these mis-charged tRNAs means that in some organisms special GluRSs and AspRSs do not discriminate between tRNA^{Gln}/tRNA^{Glu} and tRNA^{Asn}/tRNA^{Asp}, respectively. Therefore these enzymes are called non-discriminating AARSs. Non-discriminating GluRSs are found in mitochondria, chloroplasts and a group of microorganisms, such as gram-positive bacteria, cyano- and archaeobacteria that lack GlnRS [LaPointe J et al., 1986; Schön A et al., 1988; Gagnon Y et al., 1996]. Here the non-discriminating AARSs coexist with amidotransferases (AdT) that help to rescue the mis-aminoacylated tRNAs via the transamidation pathway (Fig 5-4). During transamidation the respective carboxyl

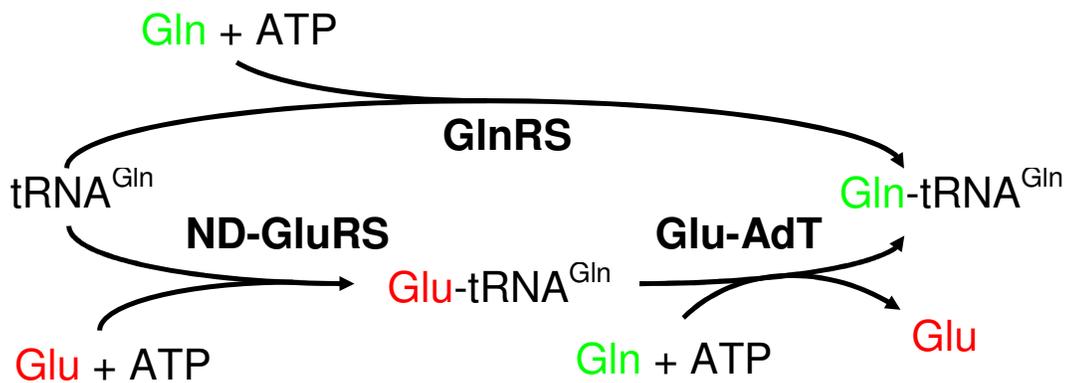


Figure 5-4: Non-discriminating GluRS and the transamidation pathway. In the absence of GlnRS tRNA^{Gln} is charged with Glu by ND-GluRS. Mis-charged Glu-tRNA^{Gln} is rescued by transamidation catalyzed by Glu-AdT resulting in properly charged Gln-tRNA^{Gln}. The mechanism of action is similar for ND-AspRS.

groups of the Asp/Glu sidechains are amidated to generate properly charged Gln-tRNA^{Gln} and Asn-tRNA^{Asn}.

This work will focus on the non-discriminating GluRS from *M. tuberculosis* (EC 6.1.1.17, Glutamate-tRNA ligase). Both discriminating and non-discriminating GluRSs are class I enzymes and therefore contain a Rossmann fold and approach the tRNA from the minor groove side to attach the tRNA to the 2' hydroxyl. Three-dimensional structures have been presented to date for discriminating GluRS from *Thermus thermophilus* (TtGluRS) [Nureki O et al., 1995; Sekine S et al., 2001; Sekine S et al., 2003] and non-discriminating GluRS from *Thermosynechococcus elongatus* (TeGluRS) [Schulze JO et al., 2006]. On the basis of both tRNA-bound and tRNA-free structure from TtGluRS it was proposed that a single arginine at the position 358 in the AARS is the major determinant for tRNA discrimination [Sekine S et al., 2001]. This Arg358 was found in the crystal structure of TtGluRS-tRNA^{Glu} to make hydrogen bonding contacts to a cytosine nucleotide in position 36 in the anticodon of tRNA^{Glu} (₃₄CUC₃₆). In the anticodon of tRNA^{Gln} (₃₄CUG₃₆) this position is occupied by a guanine. As the base of G36 is considerably more bulky it was proposed that the Arg358 can not make hydrogen bonding contacts to it and presents a sterical hindrance for this base, thereby explaining the discrimination of tRNA^{Gln}. Furthermore, it was shown that a R358Q mutant of TtGluRS was able to modify both tRNA^{Glu} and a variant tRNA^{Glu}, where C36 was mutated to G36 [Sekine S et al., 2001]. Thus the mutant enzyme was losing its discriminating power against a

guanine at position 36 in the anticodon upon the presence of a smaller sidechain in the position of Arg358. The tRNA free crystal structure of TeGluRS agreed well with this mechanism. In this non-discriminating enzyme the position corresponding to that of Arg358 is occupied by a glycine (Gly366). Due to the small Gly366 it was proposed by superposition with TtGluRS that both cytosine as well as guanine present in the anticodon can be tolerated by the enzyme, explaining its non-discriminating nature [Schulze JO et al, 2006]. Yet this does not explain in general how discrimination and non-discrimination are working. That is mainly due to the fact, that the arginine identified as the major discriminating factor in TtGluRS is conserved in many non-discriminating GluRSs, as shown by sequence comparison (data not shown). Classification of GluRSs into non-discriminating and discriminating enzymes is hereby based on the genomic criterion of the absence or presence of genes for GlnRS and AdTs, respectively. In this respect the GluRS of a given organism is termed non-discriminating in the absence of a gene for GlnRS and the presence of the genes necessary for AdTs. On the contrary GluRSs of organisms that do have a gene for GlnRS and for AdTs are termed discriminating. Notably, there is a third group of organisms that carry both a gene for GlnRS and the genes for AdTs. The GluRS of *T. thermophilus* belongs to the latter group.

This work presents the crystal structure of GluRS from *M. tuberculosis* (MtGluRS). The genome of *M. tuberculosis* does not contain a gene for GlnRS but instead contains the genes GatA, GatB and GatC coding for an amidotransferase (AdT). Therefore MtGluRS is a non-discriminating AARS and further has an arginine at the position of the discriminating arginine from TtGluRS. As both discriminating TtGluRS and non-discriminating MtGluRS show high sequence homology and interact with structurally similar tRNAs, complexes of MtGluRS with tRNA^{Glu} and tRNA^{Gln} were modelled. This was done on the basis of the complex TtGluRS-tRNA^{Glu} (1N77). Furthermore, for comparison a model of TtGluRS-tRNA^{Gln} was modelled. The four models were submitted to molecular dynamics (MD) calculations to get detailed information on the stability and dynamic behaviour of the tRNA-AARS contacts. On the basis of this combination of experimental and theoretical results a structural mechanism is proposed that may explain the functional differences in anticodon recognition by the GluRSs. Additionally, the identification of the probable interaction

sites of tRNA and GluRS for *M. tuberculosis* provides the basis for rational drug design directly targeting protein biosynthesis via inhibition of tRNA charging.

5.3. Structural analysis

5.3.1. Model quality

The crystal structure of MtGluRS in the apo state was solved in space group $P2_12_12_1$. Final statistics for data processing and refinement are presented in Table 5-1. The crystal structure of the ternary TtGluRS complex 1J09 was used as a search model for molecular replacement with the program MOLREP [Vagin A and Teplyakov A, 1997]. Based on the solution from molecular replacement the apo MtGluRS structure was modelled and refined at high resolution of 1.6 Å (Fig. 5-6). The model converged to low R-/R_{free}-values of 21 % and 25 %, respectively. Stereochemical quality was assessed by PROCHECK and showed 99,8% of all residues were in the most favoured or allowed regions of the Ramachandran plot, while Glu77 was observed in a disallowed region.

Table 5-1: Data collection and refinement statistics

Data set	Apo
Wavelength (Å)	0.9756
Space group	$P2_12_12_1$
Unit cell (Å)	a=50.1, b=61.4, c=169.5 $\alpha=\beta=\gamma=90^\circ$
A. Diffraction statistics	
Resolution	20.00 – 1.65 (1.70 – 1.65)
No. of observations	152,258
No. of unique reflections	56,328
Completeness (%)	88 (91.5)
I/sigma(I)	10.9 (2.3)
Rmerge (%)	0.07 (0.51)
B. Refinement statistics	
Resolution range (Å)	19.52 – 1.65
No. of reflections (working/test)	53,368/2,840
R/Rfree (%)	21/25
Mean B-factor (Å ²)	25.6
Rms bond length deviation	0.017
Rms bond angle deviation	1.7

5.3.2. Overall structure

The final model consists of the residues 2 to 485 of the MtGluRS protein sequence (Fig. 5-5), showing some disordered residues at the C-terminus, including the residues 486 to 490 and the His-tag. The enzyme consists of five distinct domains (Fig. 5-6):

- 1.) Rossman Fold (RF) domain (red)
- 2.) Connective peptide (CP) domain (green)
- 3.) Stem contact (SC) fold domain (cyan)
- 4.) Anticodon binding (AB) domain 4 (blue)
- 5.) Anticodon binding (AB) domain 5 (magenta)

The five domains appear in the given order in the MtGluRS sequence (Fig. 5-5) with the exception that the CP domain (residues 77 to 195) is inserted into the RF domain, dividing the latter into two parts. Thus the RF domain consists of two peptide segments ranging from 2 to 76 and from 196 to 248, respectively. The successive SC fold domain starts at residue 249 with the short β_{10} -strand that is followed by the loop $^{257}\text{KLSKR}^{261}$ and ends at residue 335. The aforementioned loop corresponds to the KMSKS signature motif of AARSs that is involved in ATP binding. At the end come the AB domains 4 and 5 that mainly consist of α -helices and range from residue 336 to 390 and from residue 391 to 485, respectively.

The two GluRS enzymes from *M. tuberculosis* and *T. thermophilus* show a sequence identity of about 40% in the N-terminal part ranging from residue 6 to 320 and a somewhat smaller sequence identity of about 20% in the C-terminal part from residue 321 to 485. While a superposition of tRNA-free MtGluRS and TtGluRS based on 463 corresponding C α -atoms results in a root-mean-square deviation (RMSD) of 2.8 Å, better agreement is seen upon partial superposition. Alignment of the AB domains shows an RMSD of 1.6 Å and alignment of the RF, CP and SC domains shows an RMSD of 1.4 Å. Nevertheless three segments in the last named domains show substantial conformational differences. The first is the KLSKR loop in the RF domain and the second is a region flanked by the helices α_4 (residues 114 to 124) and α_5 (141 to 149) (Fig. 5-5 and Fig. 5-6). In this region an insertion is present at position 127 to 130 that causes the helix α_5 to protrude further than the corresponding helix

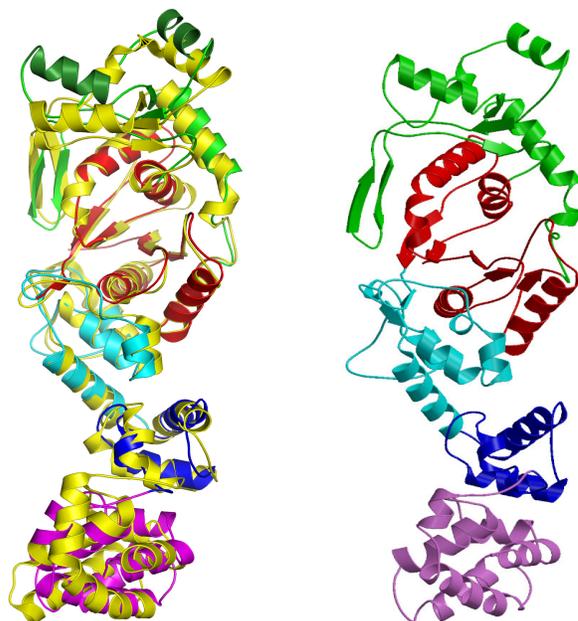


Figure 5-6: MtGluRS topology: Rossman Fold (RF) domain (red); connective peptide (CP) domain (green); stem contact (SC) fold domain (cyan); anticodon binding (AB) domain 4 (blue); anticodon binding (AB) domain 5 (magenta). Left: superposition of MtGluRS onto TtGluRS (PDB-ID 1J09 colored in yellow)

seen in TtGluRS. Finally the loop 291 to 298 in the SC domain shows a different conformation compared to the shorter, homologue loop (277 to 282) in TtGluRS.

The glutamate binding pockets of both MtGluRS and TtGluRS share a similar conformation due to a bound acetate molecule that mimics the γ -carboxyl moiety of glutamate. Acetate was present in the crystallisation condition and the bound acetate ligand forms a hydrogen bonding network with several conserved residues including Arg10, Asn100, Glu46, Tyr196, Ser14 and Arg214. Additionally the semi-conserved Cys12 that is Ala7 in TtGluRS is contacted by the acetate ligand. Next to the acetate molecule a crystal water is present in the binding pocket that is in the position of the α -NH₂-group of glutamate as seen in the ternary TtGluRS complex with glutamate and ATP (PDB-ID 1J09). Therefore the water molecule shows the interactions with the residues Cys12, Ser14 and Glu46 similar to those seen for glutamate in the complex 1J09. This structural similarity allows to model a glutamate molecule in the active site of MtGluRS in which the γ -carboxyl moiety occupies the position of the acetate molecule and the α -NH₂-group the position of the crystal water (Fig. 5-7).

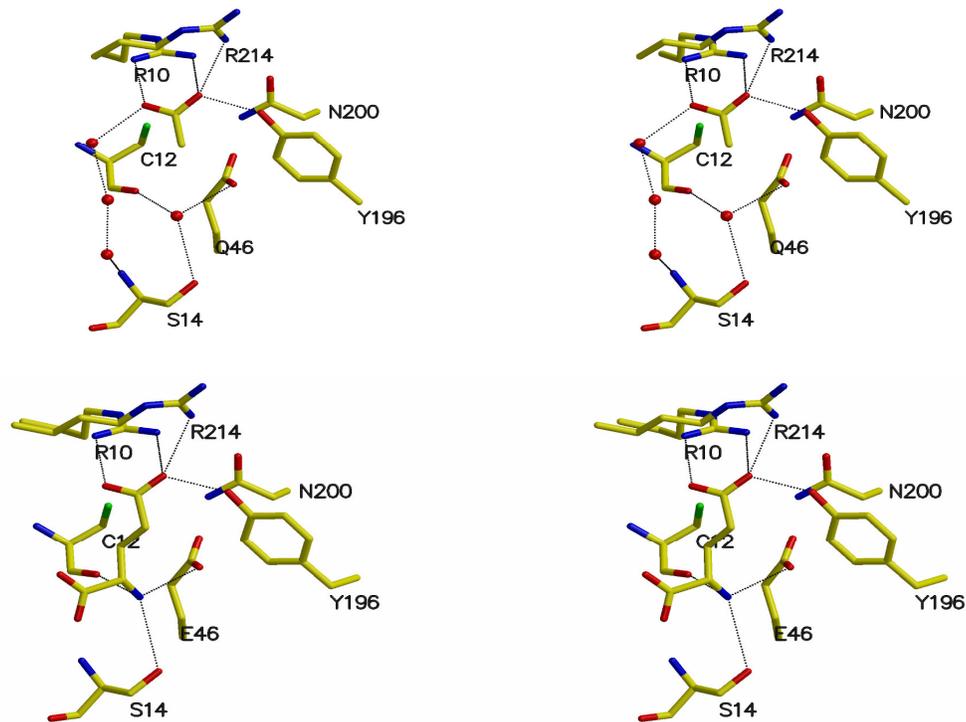


Figure 5-7: Model of glutamate binding in the MtGluRS amino acid binding pocket. Above: Stereoview of acetate bound in the amino acid binding pocket. Below: Stereoview of the modelled glutamate binding. The γ -carboxyl-moiety of glutamate occupies the position of the acetate and the α -NH₂-group the position of a crystal water. Interacting residues are similar in TtGluRS (PDB-ID 1J09).

5.3.3. Interim summary

The structural analysis of the non-discriminating *M. tuberculosis* glutamyl-tRNA synthase (MtGluRS) and its comparison to the discriminating *T. thermophilus* GluRS (TtGluRS) shows the following key aspects:

- The X-ray structural model of MtGluRS shows high quality and displays the typical topology of AARSs consisting of five distinct domains.
- MtGluRS and TtGluRS display a quite high sequence and structural identity despite their fundamentally different function regarding the mechanism of non-discrimination/discrimination between tRNA^{Glu} and tRNA^{Gln}.

5.4. Enzyme-tRNA interactions

5.4.1. Docking models of GluRS and tRNA

Based on the observation that the non-discriminating MtGluRS and the discriminating TtGluRS carry both the same arginine supposed to be able to discriminate between tRNA^{Glu} and tRNA^{Gln} [Sekine S et al., 2001] docking models of both GluRSs with both tRNAs were constructed. This was done in order to validate the mechanism of discrimination with the two different anticodons of tRNA^{Glu} (CUC) and tRNA^{Gln} (CUG) (Fig. 5-8). The conformations of MtGluRS in complexes with tRNA^{Glu} and tRNA^{Gln} were modelled on the basis of the experimental structures of complexes of TtGluRS with tRNA^{Glu} (1N77). To model the tRNA-bound form of MtGluRS the TtGluRS enzyme in chain A from the PDB entry 1N77 served as a template. In the case of TtGluRS no major conformational differences were observed between the tRNA-bound and free enzyme [Sekine S et al., 2003]. Binding of the tRNA molecule mainly involved minor rotations ($\sim 6-8^\circ$) of single domains without changes in their fold. Similar can be expected for MtGluRS and therefore the tRNA-bound state was modelled manually by adapting the relative domain orientations as observed for TtGluRS-tRNA^{Glu}. Starting models of *M. tuberculosis* tRNA^{Glu} (CUC) and tRNA^{Gln} (CUG) were also built manually using the modelling functions of COOT. As a template for the models the tRNA^{Glu} structure from *T. thermophilus* (1N77; chain C) was used. Prior to the modelling the secondary structure of Mt-tRNA^{Glu} and Mt-tRNA^{Gln} was predicted with the program tRNAscan-SE [Lowe TM and Eddy SR, 1997] provided by the Genomic tRNA data base services⁷. Both enzyme and tRNAs were energy minimized prior to docking. After minimization tRNAs and the enzyme were docked by superimposing both onto one TtGluRS-tRNA^{Glu} complex from 1N77 (chains A and C). For superposition of the enzyme and the tRNA C α and phosphate atoms were used, respectively. Finally the contacts at the enzyme-tRNA interfaces were manually adjusted on the basis of contacts conserved between tRNA and enzyme in both TtGluRS and MtGluRS. The completed complex structures were refined and validated with COOT. Ramachandran plots for the docking complexes showed all non-glycine and non-proline residues in allowed regions.

⁷ <http://lowelab.uscsc.edu/GtRNAdb>

All theoretically built models of the tRNAs were further validated using the software X3DNA [Lu XJ and Olson WK, 2003]. Results for the geometry of the modelled tRNAs were compared to those from the tRNA^{Glu} model of obtained by crystallography (PDB-ID 1N77). Upon this comparison the geometries of the theoretically modelled tRNA^{Glu/Gln} for *M. tuberculosis* were found to be in good agreement with the crystallographic data for tRNA^{Glu} from *T. thermophilus*. Subsequently the docking models of GluRS-tRNA^{Glu/Gln} were submitted to energy minimization and the resulting models were found to have good geometries. Differences in the backbone conformations between MtGluRS-tRNA^{Glu} and TtGluRS-tRNA^{Glu} are only observed for insertions or deletions in loop segments. In the docking complex MtGluRS-tRNA^{Glu} the enzyme mainly interacts with the acceptor arm, the augmented D-helix and the anticodon arm. The latter is contacted by the KLSKR loop

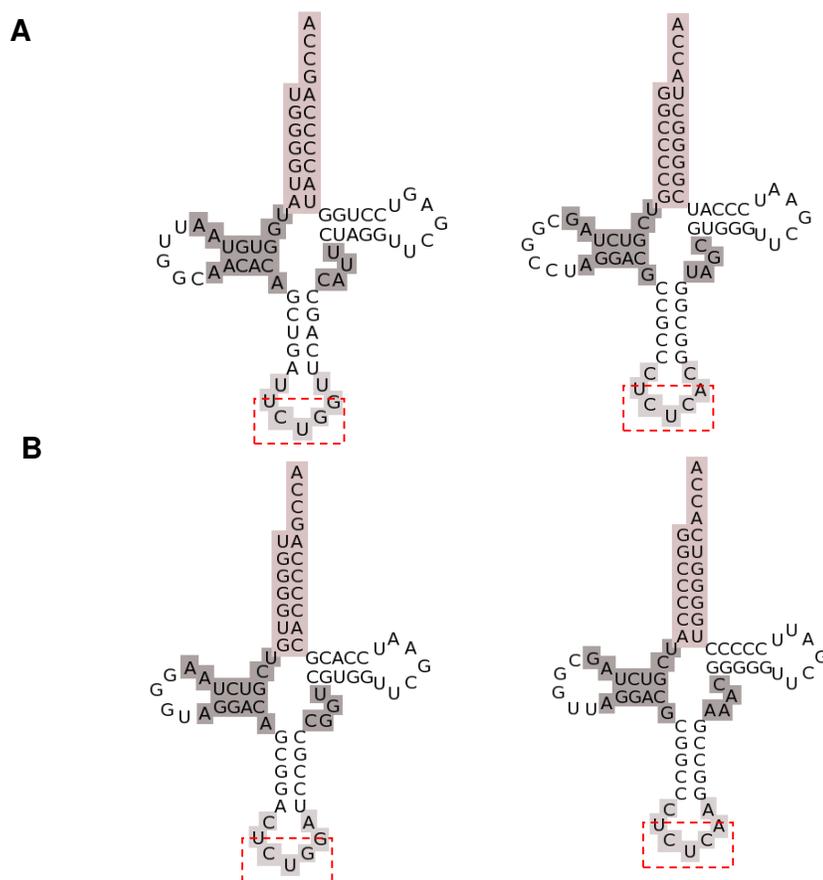


Figure 5-8: Cloverleaf representation of tRNA^{Gln} (left) and tRNA^{Glu} (right) for *M. tuberculosis* (A) and *T. thermophilus* (B). The anticodon is indicated by red boxes. The acceptor arm is highlighted in grey, the D-stem and variable loop in dark grey and the anticodon loop in light grey

as well as by the helices $\alpha 4$ and $\alpha 5$. Both helices show elevated backbone temperature factors (averaged 32 \AA^2 compared to 21 \AA^2 for the domains RF, SC and AB 4) in the MtGluRS apo structure. Thereby the helices indicate conformational flexibility that might facilitate binding of the acceptor arm upon tRNA binding. The conformations of the helices $\alpha 4$ and $\alpha 5$ are similar in the tRNA^{Glu} complexes with both MtGluRS and TtGluRS. Additionally MtGluRS makes contact to the phosphate group of the nucleotides G70 and C71 via the KLSKR loop in the same way the KISKR loop of TtGluRS is involved into contacts with the phosphates of the corresponding nucleotides G70 and U71. These interactions are involved in the rearrangement of the active site upon tRNA^{Glu} binding [Sekine S. et al., 2003]. Further similarities in tRNA^{Glu} binding between MtGluRS and TtGluRS include contacts of residues of the SC domain to the D stem. Especially contacts of the

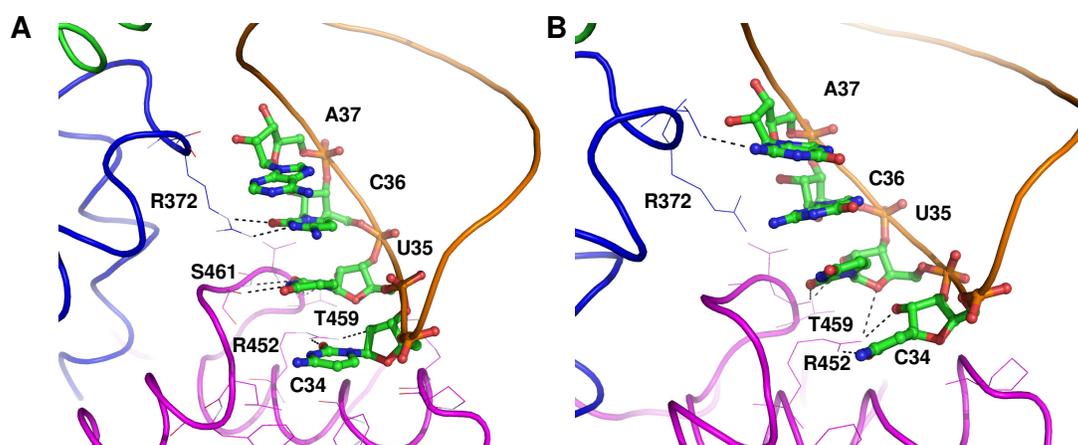


Figure 5-9: Anticodon recognition in the modelled MtGluRS-tRNA^{Glu} (A) and MtGluRS-tRNA^{Gln} (B) complexes. See text for details

nucleotides 11 to 13 in MtGluRS-tRNA^{Glu} to the D stem are similar to those observed in the TtGluRS-tRNA^{Glu}-ATP structure. Here these contacts were reported to play a role in the tRNA dependent switching of the ATP binding mode [Sekine S. et al., 2003]. In MtGluRS-tRNA^{Glu} the loop 291 to 298 is close to the minor groove of the D arm involving base specific contacts for Asp293, Asp296 and Lys323. Here the base specific contact of Asp296 corresponds to the essential direct interaction seen in TtGluRS-tRNA^{Glu} between Glu282 and the tRNA [Nureki O et al., 1995]. In the MtGluRS-tRNA^{Glu} model a further specific interaction between Asp293 and G22 is observed. In total, close similarity in the contacts can be observed in both GluRS-

tRNA^{Glu} complexes from *M. tuberculosis* and *T. thermophilus*. For both complexes all conserved residues in the contact areas between enzyme and tRNA are observed in closely similar conformations.

5.4.2. Anticodon recognition

The AB domains 4 and 5 of GluRS are involved in the recognition of the respective anticodon of tRNA^{Glu/Gln} (Fig. 5-9). In both TtGluRS [Sekine S et al., 2001] and MtGluRS an interdomain cleft between AB 4 and AB5 allows the accommodation of the nucleotides in the anticodon region. In the case of MtGluRS and tRNA^{Glu} these nucleotides are C34, U35, C36 and A37 (Fig. 5-9a) and for tRNA^{Gln} C34, U35, G36 and A37 (Fig. 5-9b) are accommodated by the AB domains 4 and 5. In both anticodons the first position was suggested to play a role in discrimination between tRNA^{Glu} and tRNA^{Asp} [Sekine S et al., 1996]. In MtGluRS C34 is recognized by a base specific interaction with the side chain of Arg452 in the AB domain 5 both in tRNA^{Glu} and tRNA^{Gln}. Interaction of Arg452 with C34:O2 occurs via both nitrogens of the guanidinium moiety, similarly as the interaction of Arg435 to C34 observed in TtGluRS. This arginine is conserved in most GluRSs [Siatecka M et al., 1998], whereas e.g. the non-discriminating GluRS from *Sinorhizobium meliloti* carries a phenylalanine in this position. Like C34, the second anticodon nucleotide U35 is contacted similarly in both tRNA^{Glu} and tRNA^{Gln} in the MtGluRS complex model. The 2-carbonyl group of U35 forms hydrogen bonds to Thr459:O in tRNA^{Glu} and tRNA^{Gln} and to Ser461 in tRNA^{Glu}. In TtGluRS Thr444 corresponds to Ser461 in MtGluRS and plays a similar role in anticodon recognition. Further contacts in TtGluRS to U35 involve the carbonyl-oxygen of Pro445. A similar contact between the corresponding Pro462 in the MtGluRS complexes is not observed.

Finally the major difference in the anticodon of tRNA^{Glu} and tRNA^{Gln} are the nucleotides C36 and G36, respectively. These two nucleotides are glutamate and glutamine identity determinants. In MtGluRS-tRNA^{Glu} the side chain of Arg372 of AB domain 4 adopts a distinct position and forms hydrogen bonds to C36 (Fig. 5-10a). However in the tRNA free enzymes the side chain of Arg372 is flexible, as can be seen by the presence of two alternative conformations, where Arg372 is in inter-

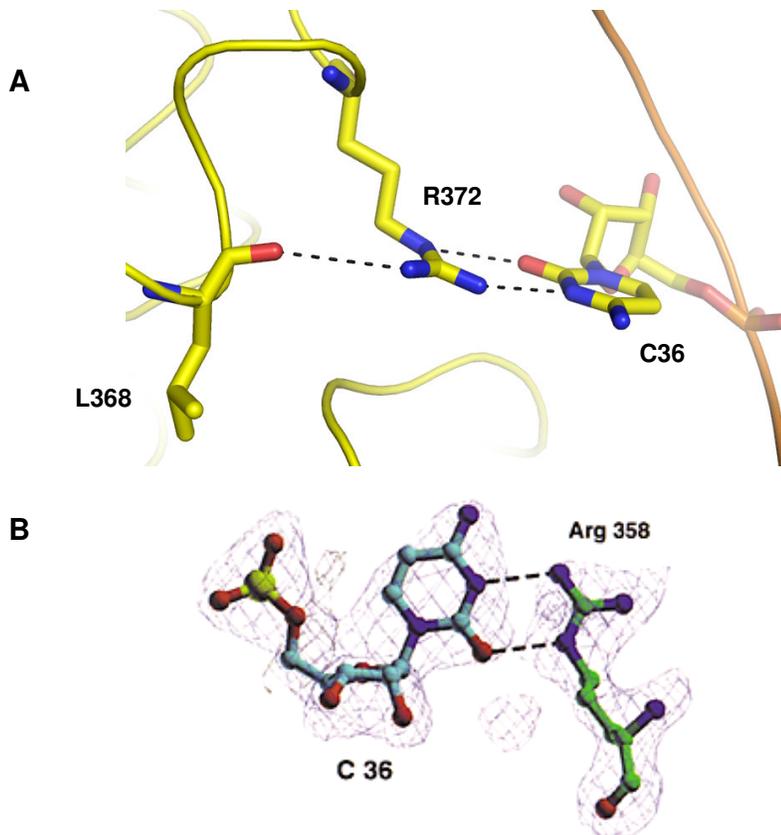


Figure 5-10: Recognition of C36 in the modelled MtGluRS-tRNA^{Glu} complex (A) and in the TtGluRS-tRNA^{Gln} crystal structure 1N77 (B) [Sekine S. et al., 2001]

domain contact with Ser461 and Pro462, respectively. These two contacts are disrupted in the enzyme-tRNA complex in favour of hydrogen bonds between Arg372:NH1 and Arg372:NE with the ring nitrogen and the carbonyl group of C36, respectively (Fig. 5-10a). Additionally, the guanidinium group of Arg372 forms a salt bridge with the conserved residue Leu368. All described interactions for Arg372 are nearly identical to those of Arg358 seen in the complex of TtGluRS with tRNA^{Glu} (Fig. 5-10b). Yet, Arg372 makes a further polar interaction with the side chain oxygen of Thr371, a contact that is not observed in a similar way in TtGluRS.

While recognition of C36 involves base specific hydrogen bonding to Arg372 in the MtGluRS-tRNA^{Glu} model, similar is not possible in the case of G36. Rather the replacement of C36 by G36 would create a sterical clash with the arginine adopting the position necessary for hydrogen bonding contact to C36 (Fig. 5-11a). This was already proposed for Arg358 in the discriminating TtGluRS [Sekine S et al., 2001]. However, in the non-discriminating MtGluRS-tRNA^{Gln} model it is possible that the corresponding Arg372 adopts a different conformation that avoids a sterical clash

with the bulkier base of G36 (Fig. 5-11a). This alternative conformation is very similar to one of the two alternative conformations observed for Arg372 in the tRNA free MtGluRS structure. Similarly a corresponding position can be adopted by Arg358 in the hypothetical TtGluRS-tRNA^{Gln} model that would avoid a sterical clash with G36 as well (Fig. 5-11b). Notably, by adopting this conformation Arg358 would get into hydrogen bonding distance to Met355, involving Arg358:NH1 and Met355:S. Since methionine itself is regarded as a hydrophobic amino acid such hydrogen bonding is highly unlikely. In fact there is only one example of a hydrogen bond involving a methionine in the entire protein data bank, rendering this case probably unfavourable. Thus, it has rather to be expected that the bulky side chain of the Met355 sterically clashes with Arg358. Hence, the protein surrounding near Met355 would prevent Arg358 to adopt an alternative conformation that would in turn avoid a

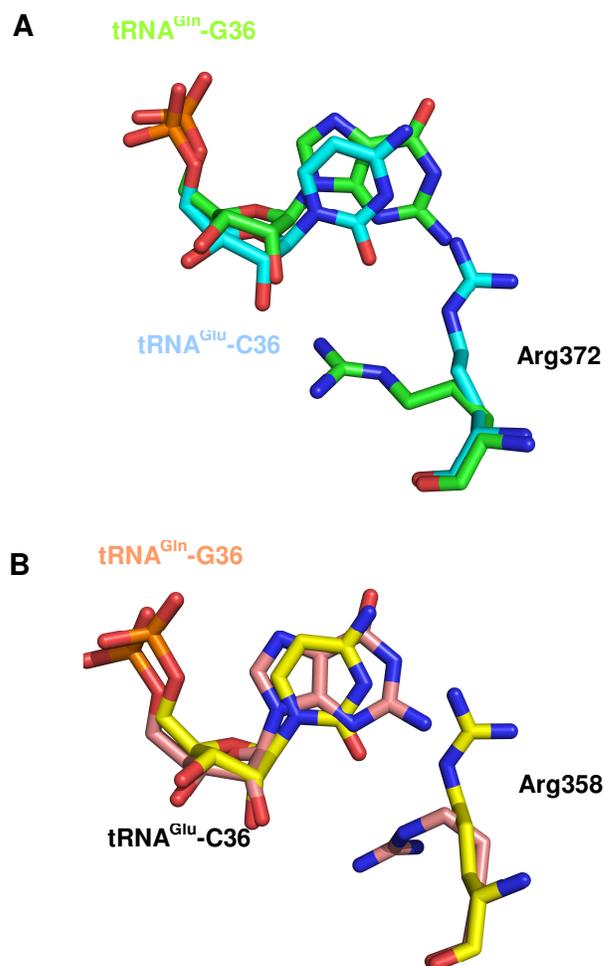


Figure 5-11: Alternative side chain conformations allow the proper accommodation of the bulky base G36 next to Arg372 in MtGluRS (A) and Arg358 in TtGluRS (B)

clash with the bulky base of G36. In contrast, the position of Met355 in MtGluRS is occupied by the considerably less bulky Val369. No comparable situation as with Met355 and Arg358 in TtGluRS would arise from the alternative position of Arg372 with Val369 in MtGluRS. Therefore in MtGluRS the alternative position of Arg372 should be stable and allow the accommodation of the bulky G36 and thus be in accordance to the non-discriminating character of MtGluRS.

5.4.3. MD simulations of the modelled GluRS-tRNA complexes

The docking models of the GluRS-tRNA complexes imply an unstable conformation of the anticodon region of tRNA^{Gln} in contact to TtGluRS while interactions between the anticodon bases of tRNA^{Gln} and MtGluRS should result in a stable accommodation of the bulky G36 near Arg372. In turn both TtGluRS and MtGluRS in complex with their respective tRNA^{Glu} should exhibit stable contacts in the anticodon region, in particular between Arg372 and C36. In order to validate this model of anticodon stability, each GluRS-tRNA^{Glx} model was submitted to 5 ns molecular dynamics (MD) simulations. All MD simulations on the tRNA-enzyme complexes were carried out using the software package GROMACS. The GROMACS port of the Amber94 force-field [Sorin EJ et al., 2005] was used with the TIP3P water model [Mahoney MW and Jorgensen WL, 2000]. Simulations were performed at 300 K and 1 bar in the NPT ensemble with a coupling time of 0.1 ps. All bonds were constrained using the LINCS algorithm and an integration time step of 2 fs was used. Cut-offs for Lennard-Jones and electrostatic interactions were 10 Å and 9 Å, respectively; long-range electrostatic interactions were calculated by particle-mesh Ewald summation. For the simulation setup each starting model was immersed in a rectangular water box of a size such that neighbouring molecules were at least 20 Å apart and counterions were added to the system to neutralize all charges. Each system (i.e. complex, water and ions) was subjected to steepest descent energy minimization until convergence of the maximum force below 1000 kJ mol⁻¹ nm⁻¹. This was followed by a 20 ps MD simulation at the target temperature using harmonic position restraints on the heavy atoms of the protein with a force constant of $k = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ to briefly equilibrate water and ions. Initial MD simulations showed that this equilibration period was too short and resulted in large conformational differences especially of tRNA^{Gln} from its starting conformation. Therefore the equilibration was prolonged. For

further equilibration the following 1000 ps the force constant was applied only to the C α and phosphate backbone and gradually decreased by 100 kJ mol⁻¹ nm⁻² every 100 ps until the free simulation was started at 1 ns. Subsequently, each system was run for another 4 ns, totalling in a 5 ns trajectory per starting model, including the equilibration phase. Root mean square standard deviation (RMSD) values were calculated towards the energy minimized simulation starting structures. Only C α atoms involved in α -helical or β -sheet secondary structure elements were used for the RMSD calculations.

5.4.3.1. TtGluRS complexes

No modelling was applied to the TtGluRS-tRNA^{Glu} complex prior to the MD simulation. Comparison of crystallographic B-factors and those derived from RMSF values for the C α atoms shows that this complex displays comparable coordinate fluctuations in the MD simulation and the crystal structure (Fig. 5-12a). However, the amplitudes of the fluctuations partly differ between crystal and simulation. This is due to the effect that generally atomic motions, especially of flexible loops, are likely more pronounced in solution compared to a crystalline environment. Yet, it can be seen that the behaviour of the single domains is similar both in simulation and in the crystal structure. Furthermore, residues involved in contacts between protein and tRNA are mostly found in the regions displaying the least fluctuation (Fig. 5-12c), obviously due to the tight packing and little exposure to solvent. Out of a total of 70 protein residues involved in the contact area only 10 display fluctuations above 30 Å² during the simulation. Most of these are located in a helix from the CP domain in contact to the 3' acceptor arm of the tRNA. This helix already shows elevated B-factors in the TtGluRS crystal structure. The other residues with elevated B-factor values in the simulation are located in a loop in contact to the anticodon of tRNA^{Glu}. In the tRNA out of 30 nucleotides in the contact area 10 display fluctuations above 30 Å², but are well below 50 Å², except for C72 near the 3' end with an RMSF value of about 75 Å². Notably, the acceptor arm of tRNA^{Glu} bound to TtGluRS seems to exhibit significantly less flexibility in the simulation compared to the B-factors of the crystal structure. While in the crystal structure the B-factors of the phosphate backbone strongly increase towards the 3'-end of tRNA^{Glu}, fluctuations actually level off in the simulation (Fig. 5-12a/b). The latter is in accordance with a tightly bound acceptor arm of the

enzyme-tRNA complex. Furthermore, a decreased flexibility of especially the first segment of the RF domain and the entire CP domain is observed. As both domains are involved in the binding of the acceptor arm, there seems to be a mutual stabilisation of the protein and the tRNA in this area. While the stem contact domain still shows a somewhat decreased flexibility, its fluctuations are virtually identical in both AB domains in simulation and crystal structure, except for two short segments ranging from 350 to 375 and from 440 to 460 in AB4 and AB5, respectively. Both are located at the interdomain cleft in which Arg358 mediates the contact to C36. In total, the comparison of the RMSF values from the simulation and the crystallographic B-factors validate a highly stable simulation of TtGluRS-tRNA^{Glu}. Comparison of the B-value fluctuations between TtGluRS-tRNA^{Glu} and TtGluRS-tRNA^{Gln} shows essentially the same pattern for both models during the simulations (Fig. 5-12b). Yet, amplitudes of the fluctuations are generally bigger for the TtGluRS-tRNA^{Gln} and may be attributed to the purely theoretical nature of this complex as TtGluRS is naturally discriminating. Increased B-factors for the protein are observed in the CP and SC domains. In these two domains the segments 150 to 170 and 260 to 320 show elevated flexibility. The first segment is located in the CP domain and consists of two antiparallel β -sheets that are in contact to the 5'-end of the tRNAs. The second segment in the SC domain forms a cleft that fits the nucleotides 12 to 18 of tRNA^{Gln}. This segment shows also shows increased B-factors. Furthermore, as expected, the anticodon region of tRNA^{Gln} displays somewhat increased fluctuations at the interdomain cleft, where the anticodon is accommodated. However, the observations from the MD simulations do not point to a general structural incompatibility of tRNA^{Gln} towards TtGluRS. That hints on a special importance of the anticodon recognition to be important for tRNA discrimination.

5.4.3.2. MtGluRS complexes

Comparison of the B-factors of the TtGluRS-tRNA^{Glu} X-ray structure to the MtGluRS-tRNA^{Glu} model shows a similar pattern of the fluctuations for both enzymes even if it is slightly shifted due to variations in the respective protein sequences (Fig. 5-13a). A significant difference of the flexibility between TtGluRS and MtGluRS concerns the CP domain of the enzyme. In MtGluRS the helix α 4 is not in contact to the 3' acceptor arm of tRNA^{Glu} and shows only minor fluctuations. The corresponding helix

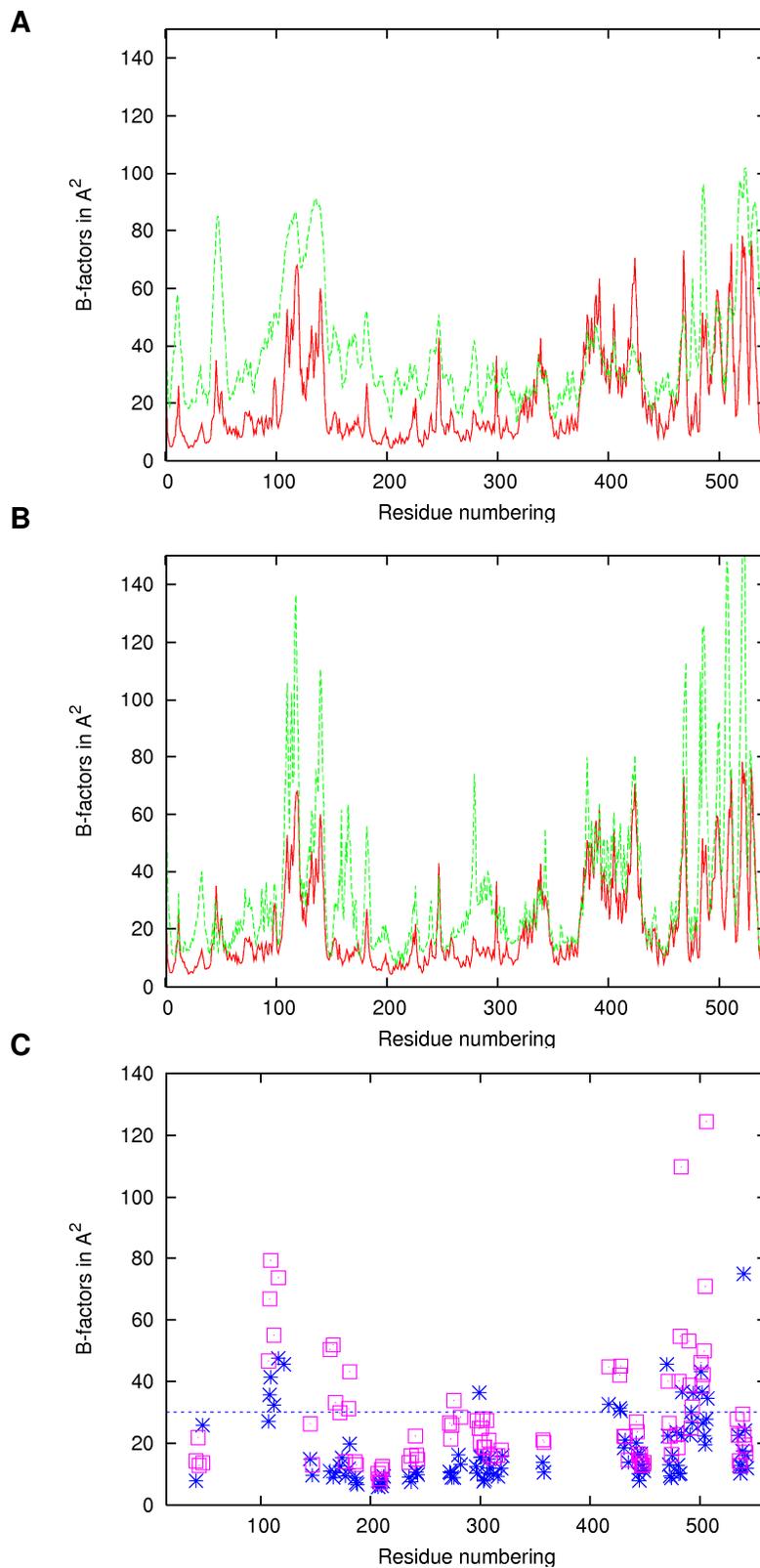


Figure 5-12: Comparison of crystallographic and MD derived B-factors. A. Crystallographic B-factors of 1N77 (green) and MD derived B-factors thereof (red). B. MD derived B-factors of TtGluRS-tRNA^{Glu} (red) and TtGluRS-tRNA^{Gln} (green). C. MD derived B-factors of residues at the tRNA-enzyme interface for TtGluRS-tRNA^{Glu} (blue) and -tRNA^{Gln} (magenta). A B-factor threshold of 30 \AA^2 is indicated by a broken blue line.

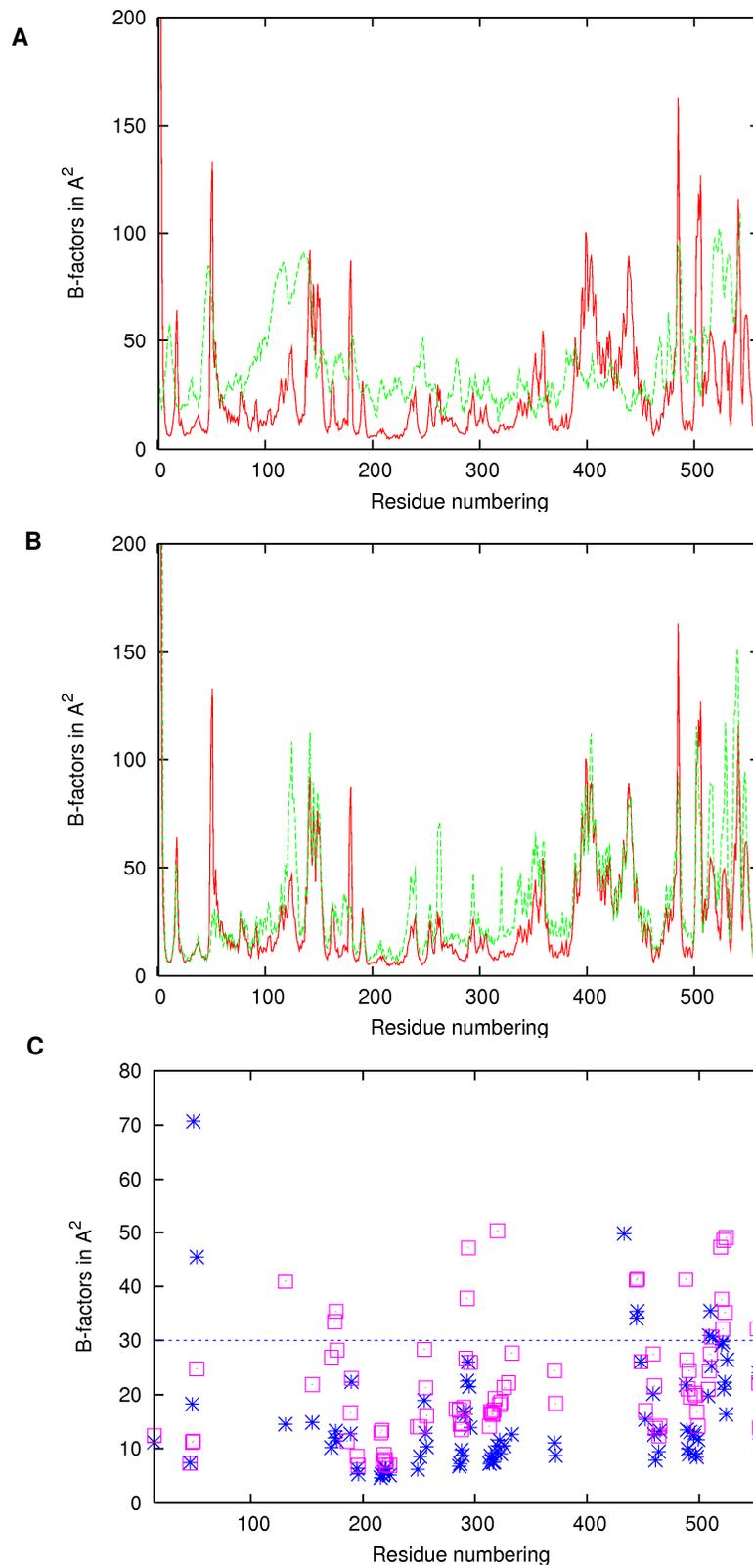


Figure 5-13: Comparison of crystallographic and MD derived B-factors. A. Crystallographic B-factors of 1N77 (green) and MD derived B-factors of MtGluRS-tRNA^{Glu} (red). B. MD derived B-factors of MtGluRS-tRNA^{Glu} (red) and MtGluRS-tRNA^{Gln} (green). C. MD derived B-factors of residues at the tRNA-enzyme interface for MtGluRS-tRNA^{Glu} (blue) and -tRNA^{Gln} (magenta). A B-factor threshold of 30 Å² is indicated by a broken blue line.

in TtGluRS as discussed above shows strong fluctuations both in the crystal and in simulation. However, the acceptor arm of tRNA^{Glu} is tightly bound in MtGluRS and displays only small B-factor values. Therefore tight coordination of the acceptor arm is not necessarily involving the CP domain. Furthermore, the AB domains show increased flexibility. Similarly to the CP domain these regions of the enzyme are not in contact to the tRNA. Thus, the high flexibility of these segments obviously is seen due to their solvent exposure. As in the case of the TtGluRS simulations, residues involved in contacts between tRNA and enzyme show only minor fluctuations (Fig. 5-13c). Most of the residues and nucleotides at the interface show RMSF values well below 30 Å². Further, no major differences of the flexibility pattern in the domains can be observed for MtGluRS-tRNA^{Glu} and MtGluRS-tRNA^{Gln} (Fig. 5-13b). Therefore the MD simulations of the MtGluRS-tRNA complexes point to a general compatibility of the enzyme towards both cognate tRNAs. This is in accordance with the non-discriminating function of the enzyme.

5.4.3.3. Deviations of the complexes from the initial model structure

The time-course of the root-mean-square deviation (RMSD) of all four simulations was monitored for both the C_α backbone of the protein and the phosphate backbone of the tRNA (Fig. 5-14 and Fig. 5-15). In case of the protein backbone all C_α atoms present in loops were excluded, as the stable fold of a protein is usually defined by well ordered structural elements like α-helices or β-sheets. Since at the beginning of the simulations both C_α and P atoms were restrained with gradually decreasing strength, the RMSD is only displayed for the time of the free simulation ranging from 1 to 5 ns. Generally tRNA backbone RMSD values show an initial equilibration after release of the restraints at a time between 1.2 and 1.5 ns. From 1.5 ns on the phosphate RMSD fluctuates stably around 2 – 2.5 Å, while the protein backbone equilibrates at the same time-scale but at a lower RMSD of about 1.5 - 2 Å. Thus, the tRNAs of all models display a stronger deviation from their starting structure, than their respective protein partner, MtGluRS-tRNA^{Glu} being the only exception. This kind of deviation of a tRNA structure from its initial conformation was observed in other studies as well [Yamasaki S. et al., 2006] and was deemed to the consequence of a slight elongation of tRNAs molecules simulated in explicit solvent. Further, the tRNA molecules observed in the crystal structure of TtGluRS-tRNA^{Glu} exhibit an elevated

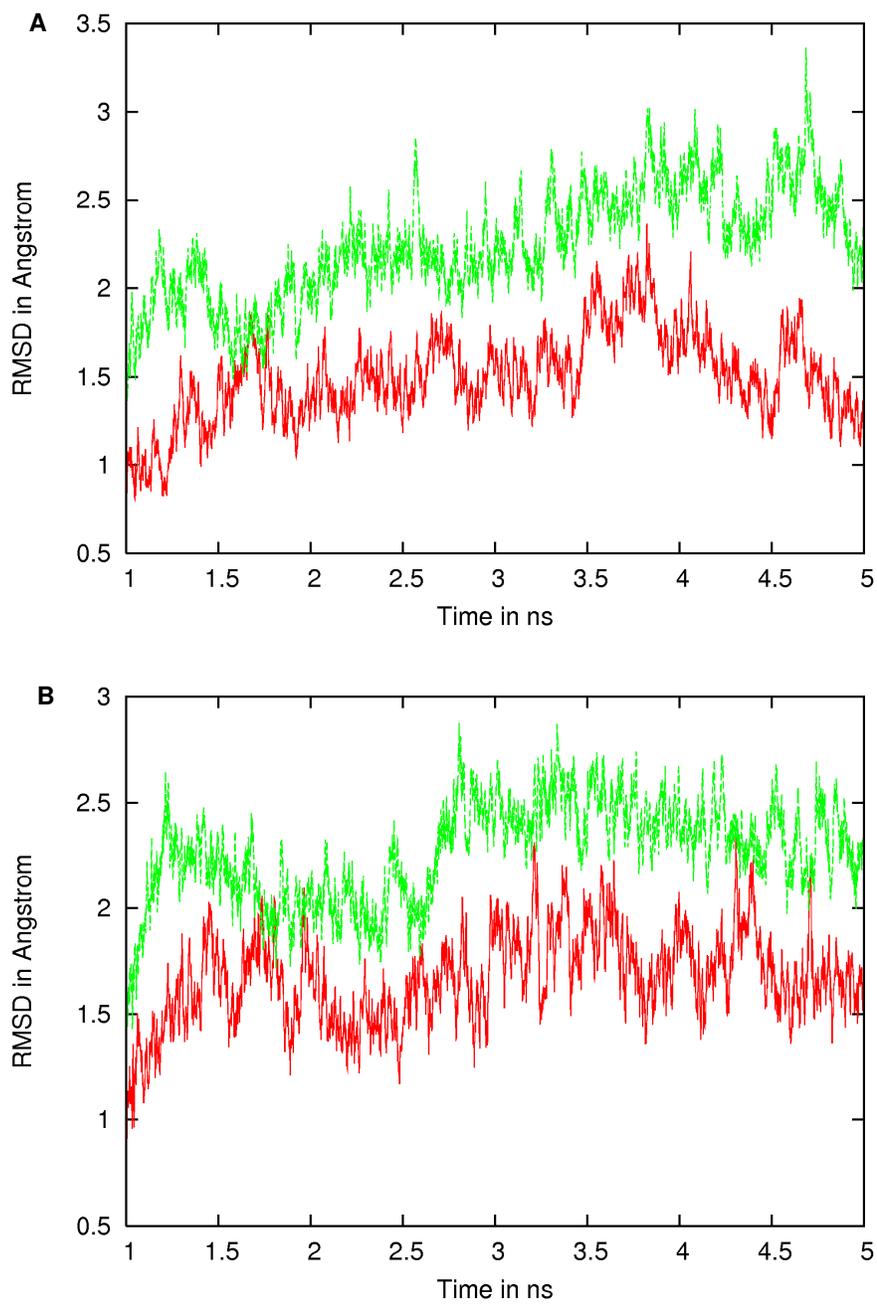


Figure 5-14: RMSD time courses of the C α - (red) and P-backbones (green) of the TtGluRS-tRNA^{Glu} (A) and TtGluRS-tRNA^{Gln} (B) complexes during the MD simulations from 1 ns to 5 ns.

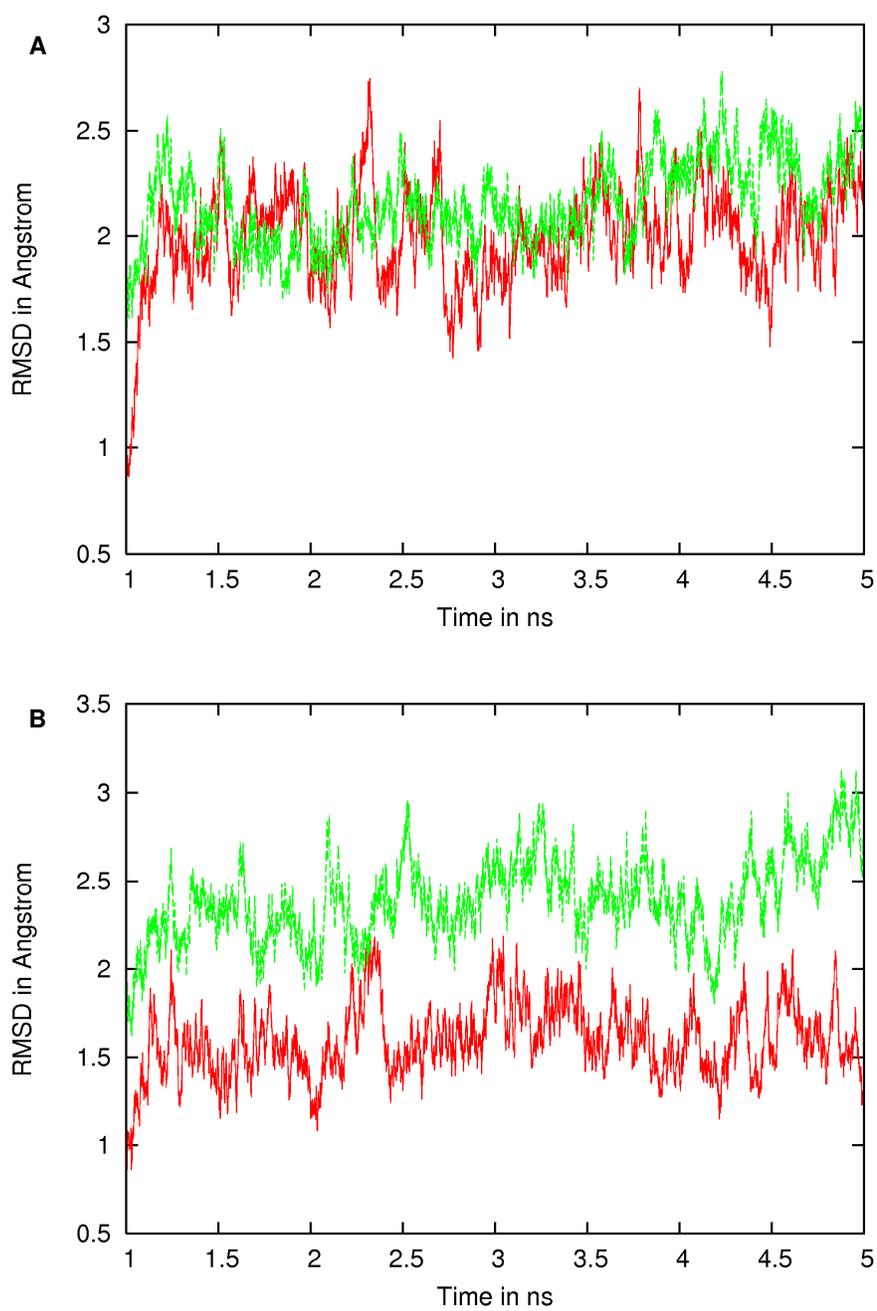


Figure 5-15: RMSD time courses of the C α - (red) and P-backbones (green) of the MtGluRS-tRNA^{Glu} (A) and MtGluRS-tRNA^{Gln} (B) complexes during the MD simulations from 1 ns to 5 ns.

mean B-factor of 60 Å² compared to a mean B-factor of 34 Å² for the protein. This implies an increased flexibility of the tRNA compared to the protein. In turn this increased flexibility might be the cause for most of the structural changes of the different GluRS models compared to their starting structures. Domain decomposition of the RMSD values shows that the single domains generally only display little structural rearrangements over the simulation time. For all GluRS models the single domains except for the CP domain display RMSD values to the starting model of only 0.5 to 1 Å, indicating a strong integrity of the single domains. In all simulations the CP domain shows the highest RMSD with values between 1.2 and 1.5 Å. In contrast the AB domains show the smallest deviation from their initial coordinates with RMSDs between 0.4 and 0.8 Å. Finally, RF and SC domains vary between 0.8 and 1.2 Å in all GluRS complexes. The small deviations from the starting structure of all enzymes point to rigid body movements of the single GluRS domains during the simulation. Most likely due to the elongation of the tRNA, the proteins must adapt to the new conformation and this is done by a rearrangement of the domain orientations relative to one another. In the MD simulations the integrity of the domains is not compromised during this adaptation process. Similar was observed in the crystal structure of TtGluRS of going from the apo to the tRNA liganded state [Sekine S et al., 2003]. The only larger conformational change is observed for the CP domain and likely is a consequence of the larger exposure to solvent. However, comparing all models and especially the tRNA^{Gln} complexes all models prove to be reasonably stable during the simulations. This provides the basis for the detailed study of the anticodon recognition by the conserved arginine and the influence of the protein environment in its vicinity during the MD simulations.

5.4.3.4. Anticodon recognition in the MD simulations

The anticodon recognition for both non-discriminating and discriminating GluRS complexes with tRNA^{Glu} mediated by the contacts between Arg372/Arg358 and C36 is stable throughout the entire simulation as expected. Both NE and NH1 atoms of the respective arginine are in stable contact to the carbonyl group and the ring nitrogen of C36 throughout the entire simulation. Therefore the MD calculations prove that the anticodon of tRNA^{Glu} is properly accommodated in the domain cleft between AB4 and AB5. Furthermore, the MD simulation of the MtGluRS-tRNA^{Gln} complex

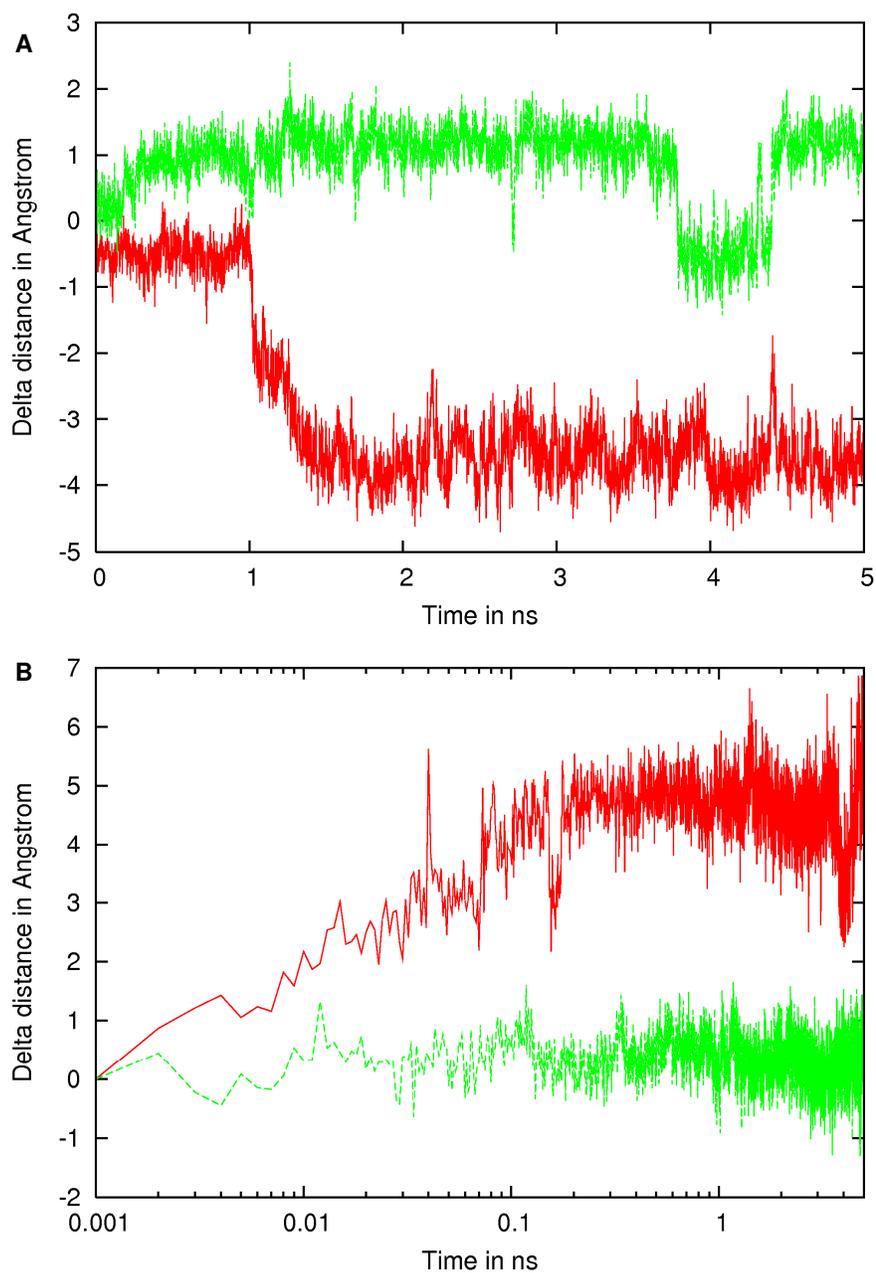


Figure 5-16: Anticodon recognition in MtGluRS-tRNA^{Gln} and TtGluRS-tRNA^{Gln}. A. Change of the distance between C R372:Cα-G36:P (red) and R358:Cα-G36:P (green) during the MD simulation. B. Change of the distance between R372:CZ-V369:CB (red) and R358:CZ-M355:SD (green) during the MD simulation. The time scale is displayed logarithmically.

shows that the alternative conformation of Arg372 is favourable. During the simulation Arg372 is accommodated in a position near the protein surface that allows proper interaction of the AB domains with the anticodon region. This can be seen by the virtually unchanged distance between Arg372 and V369 during the simulation (Fig. 5-16b). As no direct interactions can be observed between Arg372 and G36, the distance of between their C α and P atoms was monitored throughout the simulation. After the release of the restraints Arg372 and G36 approach each other by ~4 Å within the first 500 ps and afterwards stay close throughout the entire simulation (Fig. 5-16a). During this time the side chain of Arg372 is safely accommodated at the protein surface and no sterical hindrance for this conformation can be observed. In contrast, the results of the MD simulation for TtGluRS-tRNA^{Gln} show that almost immediately after the start the interaction between Arg358 and Met355 is broken and the distance between Arg358:NH1 and Met355:SD increases to 8 Å within 100 ps and stays between 7 and 9 Å during the entire rest of the simulation (Fig. 5-16b). This is accompanied by a conformational change of Arg358 turning its side chain into the direction of the base G36 (Fig. 5-17b). This prevents the anticodon region of tRNA^{Gln} to adopt a stable conformation at the cleft between AB4 and AB5. Monitoring the distance between the Arg358:C α and G36:P shows that very rapidly G36 increases its distance to the backbone of Arg358 (Fig. 5-16a). This is likely the result of the movement of the guanidinium group of Arg358 towards the anticodon region. The increased B-factor values of TtGluRS-tRNA^{Gln} in the anticodon region suggest that the orientational change of Arg358 infers a general instability on this segment.

5.4.4. Interim summary

Modelling and/or comparison of the enzyme-tRNA complexes of glutamyl-tRNA synthetases from *M. tuberculosis* (MtGluRS) and *T. thermophilus* (TtGluRS) plus their validation by molecular dynamics (MD) showed the following main points:

- Despite their different enzymatic function both MtGluRS and TtGluRS carry the same arginine (Arg372 and Arg358, respectively) that is believed to be the major determinant for discrimination between tRNA^{Glu} and tRNA^{Gln} in TtGluRS [Sekine S et al., 2001].

- Since both enzymes and the respective tRNAs are structurally closely related, docking complexes of TtGluRS and MtGluRS with tRNA^{Glx} were modelled on the basis of the TtGluRS-tRNA^{Glu} crystal structure.
- In the modelled complexes with tRNA^{Gln} the “discriminating” arginine can adopt a position that theoretically allows both in TtGluRS and MtGluRS the proper accommodation of G36 in the anticodon.
- However, in TtGluRS the bulky side chain of a methionine (Met355) in the vicinity of Arg358 is in sterical conflict with the alternative position of the arginine, while the alternative orientation of Arg372 in MtGluRS is tolerated by a considerably smaller valine (V369).
- The MD simulations show that the tRNA-enzyme complexes are overall stable and both TtGluRS and MtGluRS make stable contacts to the anticodon of tRNA^{Glu}.
- In contrast, the alternative conformation of Arg358 is instable in the vicinity of the methionine and therefore, in agreement with the discriminating function of the enzyme, disfavours the accommodation of the tRNA^{Gln} anticodon.
- However, the alternative conformation of Arg372 is stable and allows the proper accommodation of the G36 of the tRNA^{Gln} anticodon during the MD simulation, thereby making possible the recognition of both tRNA^{Glu} and tRNA^{Gln}, as implied by the non-discriminating function.

5.5. Discussion

MtGluRS is the only GlxRS present in *M. tuberculosis*. Based on the genomic criterion of the absence of any GlnRS or paralogous GluRSs in the genome and the presence of the genes *gatA*, *gatB* and *gatC* encoding for Glu/Asp AdT it may be concluded that the enzyme is non-discriminating. In contrast, the genome of *T. thermophilus* contains genes not only for GluRS and Glu/Asp AdT, but also for GlnRS. As *T. thermophilus* does not have an AsnRS, the Glu/Asp AdT is required for the synthesis of asparagine. This might be the reason that the *gatCAB* genes were

not lost during evolution of this eubacterium. The discriminating character of TtGluRS was described previously [Becker HD and Kern D, 1998], and it was demonstrated that a replacement of the discriminating Arg358 by Gln (R358Q) resulted in a substantial change towards non-discrimination of wild type (C36) and variant (C36G) tRNA^{Glu} transcripts [Sekine S et al., 2001].

Structural and sequence alignments of GluRSs from *M. tuberculosis* and *T. thermophilus* show that both enzymes are closely related to each other (Fig. 5-5 and Fig. 5-6). Both enzymes display a high sequence homology and possess closely similar three-dimensional structures. Further, they interact with tRNAs that are structurally similar. This was used to for the comparison of the docking model of the MtGluRS-tRNA^{Glu} complex to the TtGluRS-tRNA^{Glu} structure. Both models underlined the close similarity in the interactions of both GluRSs with their cognate tRNA^{Glu}. The key amino acids interacting with the anticodon loop bases are practically identical for both enzymes. In particular the interaction of the arginine (Arg372 in MtGluRS, Arg358 in TtGluRS) and the nucleotide C36 of tRNA^{Glu} (CUC) can be modelled in an identical conformation. While this arginine residue is almost invariant in discriminating GluRSs, not only MtGluRS but nearly 50% of all non-discriminating GluRSs contain an arginine in this position. Further, close agreement between MtGluRS and TtGluRS is observed also for key residues involved in the recognition of C34 (Arg435 and Leu447 in TtGluRS, Arg452 and Leu464 in MtGluRS) and U35 (Thr444 in TtGluRS, Ser461 in MtGluRS) on the basis of the modelling.

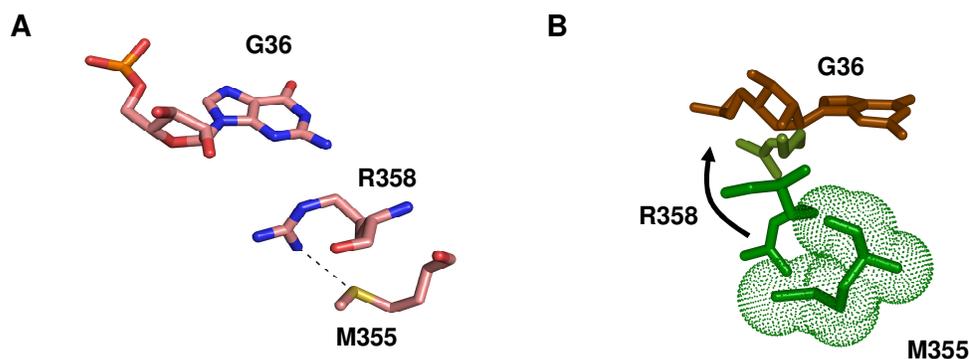


Figure 5-17: Arg358-Met355 hydrogen bonding in the TtGluRS-tRNA^{Glu} complex model. A: Starting conformation, B: Conformational change during the MD simulation

While similarities between both GluRSs are strikingly close, TtGluRS discriminates between glutamate and glutamine, whereas MtGluRS does not. The discriminative

character of TtGluRS was explained essentially by differences in the interactions of Arg358 with C36 of tRNA^{Glu} (CUC) and G36 of tRNA^{Gln} (CUG), respectively [Sekine S. et al., 2001]. The crystal structure of TtGluRS-tRNA^{Glu} (1N77) showed that the Arg358 side chain formed hydrogen bonds with C36. These specific interactions could be modelled in a practically identical conformation for Arg372 in the docking model of MtGluRS-tRNA^{Glu} and were proven to be stable in a 5 ns MD simulation. In contrast, modelling of G36 into the TtGluRS-tRNA^{Glu} structure, replacing C36, was described to result in a steric clash between G36 and the Arg358 side chain [Sekine S. et al., 2001]. Notably, neither Arg358 in TtGluRS nor Arg372 in MtGluRS can form hydrogen bonds between their side chains and G36. The docking models of the complexes of MtGluRS and TtGluRS with tRNA^{Gln} (CUG) presented in this work provided a more detailed picture of the interactions of these arginine residues with their respective environment. In the case of the MtGluRS-tRNA^{Gln} complex, the

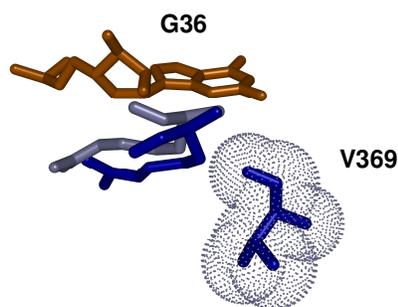


Figure 5-18: Arg372-V369-G36 interactions in the MtGluRS-tRNA^{Gln} complex model. R372 in darkblue: starting conformation, R372 in lightblue: average position in the MD simulation

Arg372 side chain in fact cannot adopt a similar orientation as in the complex with tRNA^{Glu} (CUC), as this would cause a collision with the G36 base (Fig. 5-11a). However, the Arg372 side chain can swing to a different rotamer orientation, which is essentially the same as it was observed in the crystal structure of tRNA-free MtGluRS. This allows Arg372 to form a hydrogen bond via the backbone nitrogen to the backbone carbonyl of Val369. Given this conformation of Arg372 tRNA^{Gln} may dock onto the MtGluRS without any sterical hindrance, maintaining the stacking arrangement around G36. In total, the interactions between MtGluRS and the

anticodon loops of tRNA^{Glu} and tRNA^{Gln}, respectively, do not involve discriminating structural elements. This observation is in agreement with the function of the enzyme.

A different situation is encountered in the case of TtGluRS. The docking model of the complex of TtGluRS with tRNA^{Gln} confirms the previous observation [Sekine S. et al., 2001] that the Arg358 side chain can not adopt a similar orientation as in the TtGluRS-tRNA^{Glu} complex (1N77), due to a sterical clash with G36. However, theoretically the arginine side chain can be modelled in a similar rotamer state as in the MtGluRS-tRNA^{Gln} docking complex (Fig. 5-11b). This orientation brings the side chain into hydrogen-bonding distance with the sulphur atom of the Met355 side chain (Fig. 5-17a). Such an Arg-Met interaction seems highly unusual as there is only one single protein structure in the PDB with a possible hydrogen bond between Arg:NH and Met:SD. That such an interaction is highly unlikely was further seen in a 5 ns molecular dynamics simulation of the interaction of TtGluRS with tRNA^{Gln}. This simulation showed that the tRNA in the vicinity of the anticodon base G36 gradually moved away from the protein surface (Fig. 5-16a). This movement correlates with an increase of the distance between Arg358:NH1 and Met355:SD (Fig. 5-16b). Analysis of the trajectories made clear that the motion of the Arg358 side chain was hindered by possible collisions with the side chains of Met355 and Lys368 in its vicinity. This is in accordance with observations [Sekine S. et al., 2001] of a predominantly discriminative character of wildtype TtGluRS and a change towards a non-discriminative behaviour upon replacement of C36 by G36 in tRNA^{Glu} a change of the R358Q mutant towards non-discriminative behaviour against the tRNA^{Glu} variant C36G. Clearly, the glutamine side chain is considerably smaller than that of arginine. Therefore less conflict is expected with the bulky side chains of Met355 and Lys368. In contrast to the instability in the interactions between TtGluRS and tRNA^{Gln}, a similar 5 ns MD simulation of the MtGluRS-tRNA^{Gln} complex showed high stability in the enzyme-tRNA contacts (Fig. 5-16 and Fig. 5-18). Whereas in the case of TtGluRS Arg358 swings away from the protein surface, Arg372 in MtGluRS actually moves closer to the protein surface. This results in a decrease of the distance between Arg372:CA and G36:P by ~ 4 Å, indicating the good accommodation of the anticodon to the protein surface near Arg372 (Fig. 5-16a). On the other hand, the distance between Arg358:CA and G36:P in TtGluRS increases by almost 2 Å.

The residues Val369 and Leu382 in the non-discriminating MtGluRS are replaced by residues with more bulky side chains, Met355 and Lys368 in the discriminating TtGluRS. Therefore the protein surface close to the arginine recognizing base 36

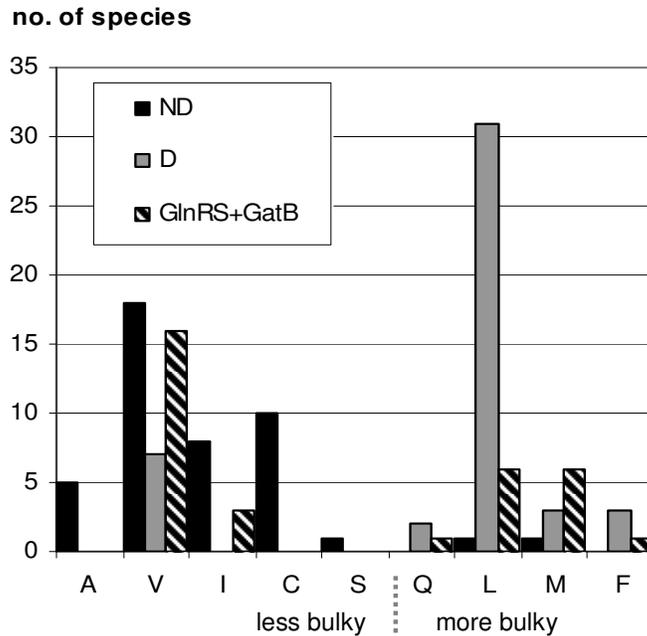
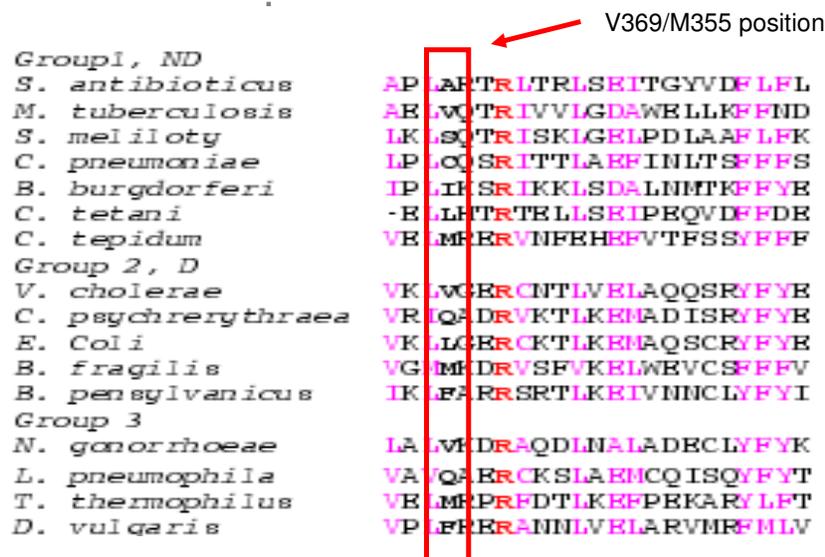


Figure 5-19: Left: Absolute count of species with GluRSs containing less bulky (A, V, I, C, S) or more bulky residues (Q, L, M, F) at the position corresponding to V369/M355 in MtGluRS/TtGluRS. Below: Partial alignment of GluRSs near the positions of V369/M355. For every group of GluRSs (non-discriminating, discriminating and GlnRS+GatB) examples are shown. The position corresponding to V369/M355 is emphasized. The conserved arginine, corresponding to R372 in MtGluRS is indicated in red



differs substantially between the two enzymes. A number of other GluRSs have been identified that suggest that such differences in the protein surface accessible to the arginine may correspond to a more general concept (Fig. 5-19). The identified GluRSs contain a conserved segment in their amino acid sequences that includes the same arginine as MtGluRS and TtGluRS. Therefore a similar interaction with tRNA near the anticodon base 36 can be expected. Notably, the discriminating

GluRSs among them possess residues with bulky side chains at the locations corresponding to Met355 and Lys368 in TtGluRS, the non-discriminating GluRSs have less bulky residues instead (Fig. 5-19). Hence, the discriminating or non-discriminating function of GluRSs may depend not only on the immediate interaction of residues with the anticodon base 36 of the tRNA, but also on the conformation of the protein environment. In this respect the structural models of MtGluRS with tRNA^{Glx} further may provide a basis for identifying new antimicrobial drugs. As drugs targeting a singular interaction might be less effective than those that are specifically designed to target a more complex environment, the model suggested above might be helpful to guide such strategies.

6. Conclusions and outlook

This work presents the high-resolution structures of two enzymes chorismate synthase (MtCS) and glutamyl-tRNA synthetase (MtGluRS) from two important metabolic pathways of the bacterium *M. tuberculosis*. The three dimensional structures of both enzymes were solved with the help of X-ray crystallography and provided detailed new insights into the respective catalytic mechanism. Comparison of the MtCS structures with a highly homologue SpCS structure allowed a detailed view of the structural changes necessary for ternary complex formation and indicates for the first time the important role of an invariant arginine necessary for the correct enzymatic function. This special role was further validated by molecular dynamics (MD) calculations. Additionally a novel ligand binding mode of NCA may support an unproductive active site conformation that also involves the invariant arginine. In the case of GluRS modelling of enzyme-tRNA complexes for MtGluRS and TtGluRS and MD simulations thereof suggest a special non-discriminating/discriminating function of GluRSs: Here the bulkiness of a single residue next to the arginine in contact to the tRNA anticodon may decide upon the discriminating or non-discriminating nature of the enzyme. Currently mutagenic studies are carried out on MtGluRS to test the proposed mechanism in vitro employing kinetic studies.

Despite the difference of the catalytic function of both enzymes they are of high interest for a joint presentation from a structural point of view. Both examples point out how small structural elements such as single side chains may determine the macroscopic function of an enzyme. In the case of MtCS a single invariant arginine is suggested to be important for the correct formation of the closed ternary complex and hence for the correct catalytic function of the enzyme. In a similar way the bulkiness of a single residue seems to determine whether a GluRS is able to discriminate between two highly similar tRNA molecules. This implies a special importance for X-ray crystallographic studies of enzymatic mechanisms, as only these provide sufficient atomic detail for the functional implications presented in this work. Finally, these functional insights may provide a good starting point for the rational design of novel compounds against TB.

References

- Ahn HJ, Yoon HJ, Lee B, and Suh SW (2004) Crystal structure of chorismate synthase: a novel FMN-binding protein fold and functional insights. *J.Mol.Biol.* 336, 903-915.
- Allinger NL (1977) MM2. A Hydrocarbon Force Field Utilizing and Torsional Terms. *J. Amer. Chem. Soc.* 99, 8127-8134.
- Andricioaei I and Karplus M (2001) On the calculation of entropy from covariance matrices of the atomic fluctuations. *Journal of Chemical Physics* 115, 6289-92.
- Arnez JG and Moras D (1997) Structural and functional considerations of the aminoacylation reaction. *Trends Biochem Sci.* 22, 211-6.
- Becker HD and Kern D (1998) *Thermus thermophilus*: a link in evolution of the tRNA-dependent amino acid amidation pathways. *Proc. Natl. Acad. Sci.* 95, 12832–12837.
- Bentley R (1990) The shikimate pathway--a metabolic tree with many branches. *Crit. Rev. Biochem. Mol. Biol.* 25, 307–384.
- Björk GR (1995) Genetic dissection of synthesis and function of modified nucleosides in bacterial transfer RNA. *Prog Nucleic Acid Res Mol Biol.* 50, 263-338.
- Bornemann S (2002) Flavoenzymes that catalyse reactions with no net redox change. *Nat Prod Rep.* Dec;19, 761-72.
- Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NS, and Venter JC (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* 273, 1058-73.
- Camus JC, Pryor MJ, Medigue C, and Cole ST (2002) Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* 148, 2967-2973.
- Centers for Disease Control and Prevention (CDC), Division of Tuberculosis Elimination. (2000) Core Curriculum on Tuberculosis: What the Clinician Should Know. 4th edition.
- Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, and Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31, 3497-500.
- Chorine V (1945) Action de l'amide nicotinique sur les bacilles du genre *Mycobacterium*. *Comp. Ren. Acad. Sc.* 220, 150-151.
- Coggins JR, Abell C, Evans LB, Frederickson M, Robinson DA, Roszak AW, and Laphorn AP (2003) Experiences with the shikimate-pathway enzymes as targets for rational drug design. *Biochem Soc Trans.* 31, 548-52.
- Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE, Tekaija F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail MA, Rajandream MA, Rogers J, Rutter S, Seeger K, Skelton J, Squares S, Squares R, Sulston JE, Taylor K, Whitehead S, and Barrell BG (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537-544.
- Collaborative Computational Project, Number 4 (1994) The CCP4 Suite: Programs for Protein Crystallography. *Acta Cryst. D* 50, 760-763.

- Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, and Kollman PA (1996) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117, 5179-5197.
- Crick F (1970) Central dogma of molecular biology. *Nature* 227, 561-3.
- DeLano W (2002) The PyMOL Molecular Graphics System. DeLano Scientific LLC, San Carlos, CA, USA.
- Dell KA and Frost JW (1993) Identification and removal of impediments to biocatalytic synthesis of aromatics from D-glucose: rate-limiting enzymes in the common pathway of aromatic amino acid biosynthesis. *J. Am. Chem. Soc.* 115, 11581 – 11589.
- Dewar MJS, Ziegler EG, Healy EF, and Stewart JJP (1985) AM1: A new general purpose Quantum Mechanical Model. *Journal of the American Chemical Society* 107, 3902-3909.
- Dias MVB, Borges JC, Ely F, Pereira JH, Canduri F, Ramos CHI, Frazzon J, Palma MS, Basso LA, Santos DS, and de Azevedo Jr. WF (2006) Structure of chorismate synthase from *Mycobacterium tuberculosis*. *J. Struct. Biol.* 154, 130-143.
- Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang J, and Kollman PA (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J Comput Chem.* 24, 1999-2012.
- Emsley P and Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Cryst. D* 60, 2126-32.
- Eriani G, Delarue M, Poch O, Gangloff J, and Moras D (1990) Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* 347, 203-6.
- Fiser A, Do RK, and Sali A (2000) Modeling of loops in protein structures. *Protein Science* 9, 1753-1773.
- Forouhar F, Lee I, Vorobiev SM, Xiao R, Acton TB, Montelione GT, Tong L, and Hunt JF (To be published) Crystal Structure of the Chorismate Synthase from *Campylobacter jejuni*, Northeast Structural Genomics Target BR19.
- Friesner RA and Guallar V (2005) Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. *Annu Rev Phys Chem.* 56, 389-427.
- Gagnon Y, Lacoste L, Champagne L, and Lapointe L (1996) Widespread use of the glu-tRNA^{Gln} transamidation pathway among bacteria. A member of the alpha purple bacteria lacks glutaminyl-tRNA synthetase. *J. Biol. Chem.* 271, 14856-14863.
- Giege R, Puglisi JD, and Florentz C (1993) tRNA structure and aminoacylation efficiency. *Prog Nucleic Acid Res Mol Biol.* 45, 129-206.
- Giege R, Sissler M, and Florentz C (1998) Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.* 26, 5017-35.
- Gogonea V, Suarez D, van der Vaart A, and Merz KM (2001) New developments in applying quantum mechanics to proteins. *Curr Opin Struct Biol.* 11, 217-23.
- Gouet P, Courcelle E, Stuart DI, and Metz F (1999) ESPript: multiple sequence alignments in PostScript. *Bioinformatics* 15, 305-8.

- Haslam E (1993) *Shikimic Acid: Metabolism and Metabolites*, J. Wiley and Sons, Chichester.
- Hassell AM, An G, Bledsoe RK, Bynum JM, Carter HL 3rd, Deng SJ, Gampe RT, Grisard TE, Madauss KP, Nolte RT, Rocque WJ, Wang L, Weaver KL, Williams SP, Wisely GB, Xu R, and Shewchuk LM (2007) Crystallization of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr.* 63, 72-9.
- Herrmann KM (1995) The shikimate pathway as an entry to aromatic secondary metabolism. *Plant Physiol.* 107, 7-12.
- Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG, and Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673-4680.
- Hooft RWW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. *Nature* 381, 272-272.
- Ibba M and Soll D (1999) Quality control mechanisms during translation. *Science* 286, 1893-7.
- Ibba M, Curnow AW, and Soll D (1997) Aminoacyl-tRNA synthesis: divergent routes to a common goal. *Trends Biochem Sci.* 22, 39-42.
- Ibba M, Becker HD, Stathopoulos C, Tumbula DL, and Soll D (2000) Adaptor hypothesis revisited. *Trends Biochem Sci.* 25, 311-6.
- Jakalian A, Jack BD, and Bayly CI (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem* 23, 1623-41.
- Jakubowski H and Goldman E (1992) Editing of errors in selection of amino acids for protein synthesis. *Microbiol Rev.* 56, 412-29.
- Jaworski EG (1972) Mode of action of N-phosphonomethylglycine: inhibition of aromatic amino acid biosynthesis. *J. Agric. Food Chem.* 20, 1195-1198.
- Jorgensen WL and Tirado-Rives J. (1988) The OPLS Force Field for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* 110, 1657-1666
- Karplus M and McCammon JA (1983) Dynamics of proteins: elements and function. *Annu Rev Biochem.* 52, 263-300.
- Kitzing K, Auweter S, Amrhein N, and Macheroux P (2003) Mechanism of chorismate synthase. Role of the two invariant histidine residues in the active site. *J Biol Chem.* 279, 9451-61.
- Konno K, Feldman FM, and McDermott W (1967) PZA susceptibility and amidase activity of tubercle bacilli. *Am. Rev. Respir. Dis.* 95, 461-469.
- Krissinel E and Henrick K (2005). Detection of Protein Assemblies in Crystals. In: M.R. Berthold et.al. (Eds.): *CompLife LNBI* 3695, 163-174.
- Kurland CG (1992) Translational accuracy and the fitness of bacteria. *Annu Rev Genet.* 26, 29-50.
- Lapointe J, Duplain J, and Proulx M (1986) A single glutamyl-tRNA synthetase aminoacylates tRNA^{Glu} and tRNA^{Gln} in *Bacillus subtilis* and efficiently misacylates *Escherichia coli* tRNA^{Gln1} in vitro. *J. Bacteriol.* 165, 88-93.
- Laskowski RA, MacArthur MW, Moss DS, and Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* 26, 283-291.
- Lindahl E, Hess B, and van der Spoel D (2001) GROMACS: A package for molecular simulation and trajectory analysis. *J. Mol. Mod.* 7, 306-317.
- Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955-64.

- Lu XJ and Olson WK (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* 31, 5108-21.
- Macheroux P, Schonbrunn E, Svergun DI, Volkov VV, Koch MH, Bornemann S, and Thorneley RN (1998) Evidence for a major structural change in *Escherichia coli* chorismate synthase induced by flavin and substrate binding. *Biochem J.* 335, 319-27.
- Macheroux P, Schmid J, Amrhein N, and Schaller A (1999) A unique reaction in a common pathway: mechanism and function of chorismate synthase in the shikimate pathway. *Planta* 207, 325-34.
- Mackerell AD (2004) Empirical force fields for biological macromolecules: overview and issues. *J Comput Chem.* 25, 1584-604.
- Maclean J and Ali S (2003) The structure of chorismate synthase reveals a novel flavin binding site fundamental to a unique chemical reaction. *Structure* 1, 1499-1511.
- Mahoney MW and Jorgensen WL (2000) A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.* 112, 8910-8922.
- Martin A, Takiff H, Vandamme P, Swings J, Palomino JC, and Portaels F (2006) A new rapid and simple colorimetric method to detect pyrazinamide resistance in *Mycobacterium tuberculosis* using nicotinamide. *J. Antimicrobial Chemotherapy* 58, 327-331.
- McClain WH and Nicholas HB (1987) Differences between transfer RNA molecules. *J Mol Biol.* 194, 635-42.
- Mitchison DA (1985) The action of antituberculosis drugs in short-course chemotherapy. *Tubercle* 66, 219-225.
- Mobley DL, Dumont E, Chodera JD, and Dill KA (2007) Comparison of charge models for fixed-charge force fields: small-molecule hydration free energies in explicit solvent. *J Phys Chem B.* 111, 2242-54.
- Mozzarelli A and Rossi GL (1996) Protein function in the crystal. *Annu Rev Biophys Biomol Struct.* 25, 343-65.
- Murray MF (2003) Nicotinamide: An oral antimicrobial agent with activity against *Mycobacterium tuberculosis* and human immunodeficiency virus. *Clin. Infect. Dis.* 36, 453-460.
- Murshudov GN, Vagin A, and Dodson EJ (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst. D* 53, 240-255.
- Nureki O, Vassilyev DG, Katayanagi K, Shimizu T, Sekine S, Kigawa T, Miyazawa T, Yokoyama S, and Morikawa K. (1995) Architectures of class-defining and specific domains of glutamyl-tRNA synthetase. *Science* 267, 1958-1965.
- Olafson BD, States DJ, Swaminathan S, and Karplus M (1983) CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comp. Chem.* 4, 187-217.
- Onyebujoh P and Rook GAW (2004) World Health Organization Disease Watch: Focus: Tuberculosis.
- Otwinowski Z and Minor W (1996) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* 276, 307-326.
- Parish T and Stoker NG (2002) The common aromatic amino acid biosynthesis pathway is essential in *Mycobacterium tuberculosis*. *Microbiology* 148, 3069-77.
- Perrakis A, Morris RM, and Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* 6, 458-463.
- Quevillon-Cheruel S, Leulliot N, Meyer P, Graille M, Bremang M, Blondeau K, Sorel I, Poupon A, Janin J, and van Tilbeurgh H (2004) Crystal structure of the bifunctional chorismate synthase from *Saccharomyces cerevisiae*. *J.Biol.Chem.* 279, 619-625.

- Ramjee MK, Coggins JR, and Thorneley RN (1994) A continuous, anaerobic spectrophotometric assay for chorismate synthase activity that utilizes photoreduced flavin mononucleotide. *Anal Biochem.* 220, 137-41.
- Rauch G, Ehammer H, Bornemann S, and Macheroux P (2007) Mutagenic analysis of an invariant aspartate residue in chorismate synthase supports its role as an active site base. *Biochemistry* 46, 3768-74.
- Read RJ (2005) Liberating crystallographers. *Structure* 13, 1236-7.
- Saks ME, Sampson JR, and Abelson JN (1994) The transfer RNA identity problem: a search for rules. *Science* 263, 191-7.
- Saks ME, Sampson JR, and Abelson J (1998) Evolution of a transfer RNA gene through a point mutation in the anticodon. *Science* 279, 1665-70.
- Sali A and Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* 234, 779-815.
- Schäfer H, Mark AE, and Van Gunsteren WF (2000) Absolute entropies from molecular dynamics simulation trajectories. *Journal of Chemical Physics.* 113, 7809-17.
- Schlichting I (2005) X-ray crystallography of protein-ligand interactions. *Methods Mol Biol.* 305, 155-66.
- Schön A, Kannangara CG, Gough S, and Söll D (1988) Protein biosynthesis in organelles requires misaminoacylation of tRNA. *Nature* 331, 187-190.
- Schulze JO, Masoumi A, Nickel D, Jahn M, Jahn D, Schubert WD, and Heinz DW (2006) Crystal structure of a non-discriminating glutamyl-tRNA synthetase. *J. Mol. Biol.* 361, 888-897.
- Sekine S, Nureki O, Sakamoto K, Niimi T, Tateno M, Go M, Kohno T, Brisson A, Lapointe J, Yokoyama S.(1996) Major identity determinants in the "augmented D helix" of tRNA^{Glu} from *Escherichia coli*. *J. Mol. Biol.* 256, 685–700.
- Sekine S, Nureki O, Shimada A, Vassylyev DG, and Yokoyama S (2001) Structural basis for anticodon recognition by discriminating glutamyl-tRNA synthetase. *Nature Struct. Biol.* 8, 203-206.
- Sekine S, Nureki O, Dubois DY, Bernier S, Chenevert R, Lapointe J, Vassylyev DG, and Yokoyama S (2003) ATP binding by glutamyl-tRNA synthetase is switched to the productive mode by tRNA binding. *EMBO J.* 22, 676-88.
- Senger B, Aphasizhev R, Walter P, and Fasiolo F (1995) The presence of a D-stem but not a T-stem is essential for triggering aminoacylation upon anticodon binding in yeast methionine tRNA. *J Mol Biol.* 249, 45-58.
- Siatecka M, Rozek M, Barciszewski J, and Mirande M (1998) Modular evolution of the Glx-tRNA synthetase family--rooting of the evolutionary tree between the bacteria and archaea/eukarya branches. *Eur. J. Biochem.* 256, 80–87.
- Singh P, Mishra AK, Malonia SK, Chauhan DS, Sharma VD, Venkatesan K, and Katoch VM (2006) The paradox of pyrazinamide: an update on the molecular mechanisms of pyrazinamide resistance in *Mycobacteria*. *J. Commun. Dis.* 38, 288-298.
- Singh UC and Kollman PA (1984) An approach to computing electrostatic charges for molecules. *J. Comp. Chem.* 5, 129-145.
- Somoskovi A, Parsons LM, and Salfinger M (2001) The molecular basis of resistance to isoniazid, rifampin and pyrazinamide in *Mycobacterium tuberculosis*. *Respir. Res.* 2, 164-168.
- Sorin EJ, Rhee YM, Pande VS (2005) Does water play a structural role in the folding of small nucleic acids? *Biophys J.* 88, 2516-24.

- Srinivasan PR, Shigeura HT, Sprecher M, Sprinson DB, and Davis BD (1956) The biosynthesis of shikimic acid from D-glucose. *J Biol Chem.* 220, 477-97.
- Stanzel M, Schon A, and Sprinzl M (1994) Discrimination against misacylated tRNA by chloroplast elongation factor Tu. *Eur J Biochem.* 15, 435-9.
- Steinrucken HC and Amrhein N (1980) The herbicide glyphosate is a potent inhibitor of 5-enolpyruvyl-shikimic acid-3-phosphate synthase. *Biochem Biophys Res Commun.* 94, 1207-12.
- Tao J, Wendler P, Connelly G, Lim A, Zhang J, King M, Li T, Silverman JA, Schimmel PR, and Tally FP (2000) Drug target validation: lethal infection blocked by inducible peptide. *Proc Natl Acad Sci* 97, 783-6.
- TDR (2004) *Nature Reviews Microbiology Disease Watch.*
- Thomas MG, Lawson C, Allanson NM, Leslie BW, Bottomley JR, McBride A, and Olusanya OA (2003) A series of 2(Z)-2-benzylidene-6,7-dihydroxybenzofuran-3[2H]-ones as inhibitors of chorismate synthase. *Bioorg Med Chem Lett.* 13, 423-6.
- Tumbula D, Vothknecht UC, Kim HS, Ibba M, Min B, Li T, Pelaschier J, Stathopoulos C, Becker H, and Soll D (1999) Archaeal aminoacyl-tRNA synthesis: diversity replaces dogma. *Genetics* 152, 1269-76.
- Vagin A and Teplyakov A (1997) MOLREP: an Automated Program for Molecular Replacement. *J. Appl. Crystallogr.* 30, 1022-1025.
- van Gunsteren WF (1987) GROMOS. Groningen Molecular Simulation Program Package; University of Groningen: Groningen, 1987.
- Viola CM, Saridakis V, and Christendat D (2004) Crystal structure of chorismate synthase from *Aquifex aeolicus* reveals a novel beta alpha beta sandwich topology. *Proteins: Struct.,Funct.,Genet.* 54, 166-169.
- Vriend G (1993) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577-637.
- Wang J, Wolf RM, Caldwell JW, Kollman PA, and Case DA (2004) Development and testing of a general AMBER force field. *Journal of Computational Chemistry*, 25, 1157-1174.
- Wang J, Wang W, Kollman PA, and Case DA (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model.* 25, 247-60.
- White PJ, Millar G, and Coggins JR (1988) The overexpression, purification and complete amino acid sequence of chorismate synthase from *Escherichia coli* K12 and its comparison with the enzyme from *Neurospora crassa*. *Biochem J.* 251, 313-322.
- Yamasaki S, Nakamura S, Terada T, and Shimizu K (2007) Mechanism of the difference in the binding affinity of *E. coli* tRNA^{Gln} to glutaminyl-tRNA synthetase caused by noninterface nucleotides in variable loop. *Biophys J.* 92, 192-200.
- Yarus M (1979) The accuracy of translation. *Prog Nucleic Acid Res Mol Biol.* 23, 195-225.
- Yarus M and Smith D (1995) In tRNA: Structure, Biosynthesis and Function. Edited by: Soll D and Raj Bhandary TL, American Society for Microbiology, 443 – 469.
- Zhou H, Singh NJ, and Kim KS (2006) Homology modelling and molecular dynamics study of chorismate synthase from *Shigella flexneri*. *J Mol Graph Model* 25, 434-41.

Curriculum vitae

Personal details:

Name: Marc Bruning

Address:

Office: Biozentrum, University of Basel
Klingelbergstrasse 50/70
CH-4056 Basel / Switzerland

Home: Mittlere Straße 148
CH-4056 Basel

Contact: m.bruning@unibas.ch (Office)
(+41) 61-267-2081
m.bruning@gmail.com (Private)
(+41) 61-321-1925

Date and place of birth: August 29, 1975 Lohne, Germany

Marital status: unmarried

Academic education:

2007	Doctoral Degree ("Doktor-Ingenieur") Hamburg University of Technology, Germany
2004 – 2007	Ph. D. Thesis at the Max-Planck Unit for Structural Molecular Biology c/o DESY, Hamburg, Germany Thesis: "Structural-functional studies of chorismate synthase and glutamyl-tRNA synthase from <i>M. tuberculosis</i> "
2003	Engineering Degree ("Diplom-Ingenieur") in Biotechnology, Berlin University of Technology, Berlin, Germany
2001 – 2002	Diploma thesis at the Max-Planck Institute for Molecular Genetics, Berlin, Germany Thesis: "Genome wide analysis of Expressed Sequence Tags for the identification of tissue specific gene expression"
1996 – 2003	Study of Biotechnology, TU Berlin, Germany

Employment:

since 2007	University of Basel, Biozentrum, Basel
2004 – 2007	Max-Planck Unit for Structural Molecular Biology c/o DESY, Hamburg
2001 – 2002	Max-Planck Institute for Molecular Genetics, Berlin