



Julius-Maximilians-Universität Würzburg



**Approaching antimicrobial resistance –
Structural and functional characterization of the fungal
transcription factor Mrr1 from *Candida albicans* and the bacterial
 β -ketoacyl-CoA thiolase FadA5 from *Mycobacterium tuberculosis***

**Auf den Spuren der antimikrobiellen Resistenz –
Strukturelle und funktionelle Charakterisierung des
Transkriptionsfaktors Mrr1 aus *Candida albicans* und der
bakteriellen β -ketoacyl-CoA thiolase FadA5 aus
*Mykobakterium tuberculosis***

Dissertation

for a doctoral degree at the Graduate School of Life Sciences,

Julius-Maximilians-Universität Würzburg,

Section Biomedicine

submitted by

Christin Marliese Schäfer

from Miltenberg am Main

Würzburg, November 2014

Submitted on:

Office stamp

Members of the *Promotionskomitee*:

Chairperson: Ulrike Holzgrabe

Primary Supervisor: Caroline Kisker

Supervisor (Second): Joachim Morschhäuser

Supervisor (Third): Nicole Sampson

Date of Public Defence:

Date of Receipt of Certificates:

Table of content

Table of content

Table of content	1
Summary	1
Zusammenfassung.....	4
1 Introduction.....	7
1.1 The emergence of antimicrobial resistances	7
1.2 Fungal infections, resistance in <i>Candida albicans</i> and the role of the transcription factor Mrr1	7
1.2.1 Fungal infections	7
1.2.2 <i>Candida</i> species and <i>Candida albicans</i>	8
1.2.3 Virulence factors.....	9
1.2.4 Antimycotics	10
1.2.4.1 Polyenes	11
1.2.4.2 Fluoropyrimidines	11
1.2.4.3 Azoles	12
1.2.4.4 Echinocandins.....	13
1.2.4.5 Others	11
1.2.5 Resistance mechanisms in <i>C. albicans</i>	14
1.2.5.1 Structure specific and independent resistance mechanisms	14
1.2.5.2 Regulation of efflux pump expression	16
1.2.6 Multidrug resistance regulator 1 (Mrr1)	17
1.3 <i>Mycobacterium tuberculosis</i> – FadA5 involved in a new druggable pathway?	21
1.3.1 <i>Mycobacterium tuberculosis</i> and the disease	21
1.3.1.1 Tuberculosis	21
1.3.1.2 <i>Mycobacterium tuberculosis</i>	22
1.3.1.3 Infection, granuloma formation and persistence	23
1.3.2 Treatment of tuberculosis, anti-TB drugs and resistance	24
1.3.3 Approaching new ways – The cholesterol metabolism pathway.....	27
1.3.4 The β -ketoacyl-CoA thiolase FadA5.....	30
1.3.4.1 Role in <i>M. tuberculosis</i> infection	30
1.3.4.2 Thiolase activity and reaction of FadA5	32
1.3.4.3 What about human thiolase homologs?	33
1.4 Research objective	34
2 Materials.....	35
2.1 Chemicals.....	35

Table of content

2.2	Equipment	35
2.3	Enzymes, Kits and Additives	37
2.4	Bacteria and Yeast culture.....	38
2.4.1	Bacterial strains, yeast strains and plasmids.....	38
2.4.2	Agarose gel electrophoresis	40
2.4.3	Antibiotics, media and additives	40
2.5	Protein purification	41
2.5.1	Buffers and solutions.....	41
2.5.2	Protein gel electrophoresis	43
2.5.3	Western blotting	43
2.6	Crystallization screens and compounds	43
2.7	Software and Databases.....	45
3	Methods	46
3.1	Molecular biology and microbiology techniques	46
3.1.1	Polymerase chain reaction (PCR)	46
3.1.2	DNA agarose gel electrophoresis	47
3.1.3	DNA purification	47
3.1.4	Transformation and genetic integration	48
3.1.4.1	Transformation into <i>E. coli</i> cells	48
3.1.4.2	Transformation into <i>M. smegmatis</i> cells – electroporation	49
3.1.4.3	Genetic integration into <i>P. pastoris</i>	50
3.1.5	Plasmid amplification and isolation.....	51
3.1.6	Cloning Techniques	52
3.1.6.1	Restriction digestion and ligation	52
3.1.6.2	Sequence and ligation independent cloning (SLIC)	54
3.1.6.3	Site-directed mutagenesis	55
3.1.6.4	Colony PCR – <i>E. coli</i>	56
3.1.6.5	Colony PCR – <i>P. pastoris</i>	56
3.1.7	Protein expression.....	57
3.1.7.1	<i>E. coli</i>	57
3.1.7.2	<i>M. smegmatis</i>	57
3.1.7.3	<i>P. pastoris</i>	57
3.2	Protein purification and characterization	59
3.2.1	Protein purification.....	59
3.2.1.1	Cell lysis	60

Table of content

3.2.1.1.1	FadA5 from <i>M. smegmatis</i>	60
3.2.1.1.2	Mrr1 variants from <i>E. coli</i>	60
3.2.1.1.3	Mrr1 variants in <i>P. pastoris</i>	61
3.2.1.2	Metal-affinity chromatography	61
3.2.1.3	Anion-exchange chromatography	63
3.2.1.4	TEV-cleavage reaction	64
3.2.1.5	Size-exclusion chromatography	65
3.2.2	Protein characterization	66
3.2.2.1	SDS-PAGE gel electrophoresis	66
3.2.2.2	Western Blotting	66
3.2.2.3	Thermofluor assay	67
3.2.2.4	Circular-dichroism (CD) spectroscopy	68
3.2.2.5	Dynamic light scattering (DLS)	69
3.2.2.6	Size-exclusion chromatography coupled multiangle laser light scattering (SEC-MALLS)	69
3.2.2.7	Isothermal titration calorimetry (ITC)	70
3.3	Protein crystallization.....	71
3.3.1	Vapor diffusion method	71
3.3.2	Sample preparation.....	72
3.3.3	Initial screening	73
3.3.4	Hit optimization.....	74
3.4	X-ray crystallography.....	75
3.4.1	X-ray diffraction and important factors	75
3.4.2	Data collection and processing.....	77
3.4.3	Phase determination and molecular replacement (MR).....	79
3.4.4	Model building, refinement and validation.....	80
3.4.5	Modeling and Hotspot analysis	82
4	Results	83
4.1	Mrr1 – a transcription factor from <i>C. albicans</i>	83
4.1.1	Cloning.....	83
4.1.1.1	Secondary structure analysis	83
4.1.1.2	Mrr1 constructs for expression in <i>E. coli</i>	84
4.1.1.3	Mrr1 constructs for expression in <i>P. pastoris</i>	86
4.1.2	Expression in <i>E. coli</i> and purification experiments.....	88
4.1.2.1	Mrr1 Full length	88

Table of content

4.1.2.1.1	Expression test	88
4.1.2.1.2	Affinity chromatography purification tests	89
4.1.2.1.3	Thermofluor analysis	90
4.1.2.1.4	Size-exclusion chromatography tests and general purification procedure	91
4.1.2.2	Mrr1 Armadillo repeat	93
4.1.2.2.1	Expression test	93
4.1.2.2.2	Expression and purification	94
4.1.2.3	Mrr1 '250' and variants	96
4.1.2.3.1	Expression tests	96
4.1.2.3.2	Thermofluor analysis	97
4.1.2.3.3	Expression and purification of the wild type protein	98
4.1.2.3.4	Expression and purification of the K335N, P683S, G878E and G997V Mrr1 '250' variants	101
4.1.2.3.5	Expression and purification of the Mrrr1 '250' wild type construct with a C-terminal His ₆ -tag	104
4.1.3	Expression tests in <i>P. pastoris</i>	105
4.1.4	Biophysical characterization.....	109
4.1.4.1	Dynamic light scattering (DLS) analysis of Mrr1 Full length protein	109
4.1.4.2	Circular-dichroism experiments	111
4.1.4.3	Size-exclusion coupled multiangle laser light scattering (SEC-MALLS)	113
4.1.5	Crystallization	115
4.1.5.1	Mrr1 '250' wild type protein	115
4.1.5.2	Mrr1 '250' mutant proteins	117
4.2	FadA5 – a thiolase from <i>M. tuberculosis</i>	119
4.2.1	Expression and purification of the FadA5 variants.....	119
4.2.2	Biochemical and biophysical characterization	121
4.2.2.1	Size-exclusion coupled multiangle laser light scattering (SEC-MALLS)	121
4.2.2.2	Isothermal titration calorimetry	123
4.2.3	Crystals structures of FadA5 apo.....	125
4.2.3.1	Crystallization	125
4.2.3.2	Data collection, structure solution and validation	126
4.2.3.3	Structural overview	129
4.2.3.4	The other 'apo' structure (apo II)	132

Table of content

4.2.4	Crystal structures of FadA5 in complex with ligands	134
4.2.4.1	Crystallization	134
4.2.4.2	Acetylated FadA5 WT(-acetyl)-CoA complex	136
4.2.4.2.1	Data collection, structure solution and validation	136
4.2.4.2.2	Structure overview and ligand binding	137
4.2.4.3	Acetylated FadA5 C93S-CoA complex	138
4.2.4.3.1	Data collection, structure solution and validation	138
4.2.4.3.2	Structure overview and ligand binding	140
4.2.4.4	FadA5 C93S-OPC complex	141
4.2.4.4.1	Data collection, structure solution and validation	141
4.2.4.4.2	Structure overview and ligand binding	142
4.2.5	Comparison of the FadA5 structures.....	146
4.2.5.1	Active site	146
4.2.5.2	Ligand binding	149
4.2.6	Modeling and comparison with human thiolases	151
5	Conclusion and perspectives	156
5.1	Mrr1 – a transcription factor from <i>C. albicans</i>	156
5.1.1	Mrr1 constructs – expression and purification	156
5.1.2	Biophysical characterization and crystallization	159
5.1.3	Modeling.....	161
5.1.4	Perspective	163
5.2	FadA5 – a β -ketoacyl-CoA thiolase from <i>M. tuberculosis</i>	164
5.2.1	Purification and crystallization	164
5.2.2	Structure analysis and comparison of the enzyme in the prominent dimeric form ...	165
5.2.3	Steps towards the determination of a lead compound	167
5.2.4	The other dimer - a redox switch?	170
6	References.....	172
7	Copyright	181
8	Appendix.....	185
8.1	Supplemental Figures.....	185
8.2	Supplemental Tables	187
8.3	Abbreviations	195
8.4	List of Figures, Tables and Equations	196
8.5	DNA and amino acid sequences of protein constructs	199
	Acknowledgements	204

Table of content

Curriculum Vitae.....	205
Affidavit	206
List of publications.....	207

Summary

The spreading of antimicrobial drug resistances among bacteria, fungi, viruses or parasites combined with often occurring negative drug interferences during treatments of more than one disease generate an urgent need for the understanding of the respective drug resistance processes and for the development of new therapeutics. This thesis addresses two examples to approach the development of new antimicrobial drugs: First, the characterization of the transcription factor Mrr1, which regulates the expression of a multidrug resistance pump in *Candida albicans*. The second example focuses on the structural characterization of the thiolase FadA5, an important enzyme from *Mycobacterium tuberculosis*.

The number of fungal infections is rising in Germany and worldwide. These infections are mainly caused by the opportunistic fungal pathogen *C. albicans*, which especially harms immunocompromised people. With increasing numbers of fungal infections, more frequent and longer lasting treatments are necessary and lead to an increase of drug resistances, for example against the clinically applied therapeutic fluconazole. Drug resistance in *C. albicans* can be mediated by the Multidrug resistance pump 1 (Mdr1), a membrane transporter belonging to the major facilitator family. However, Mdr1-mediated fluconazole drug resistance is caused by the pump's regulator, the transcription factor Mrr1 (Multidrug resistance regulator 1). It was shown that Mrr1 is hyperactive without stimulation or further activation in resistant strains which is due to so called gain of function mutations in the *MRR1* gene.

To understand the mechanism that lays behind this constitutive activity of Mrr1, the transcription factor should be structurally and functionally (*in vitro*) characterized which could provide a basis for successful drug development to target Mdr1-mediated drug resistance caused by Mrr1. Therefore, the entire 1108 amino acid protein was successfully expressed in *Escherichia coli*. However, further purification was compromised as the protein tended to form aggregates, unsuitable for crystallization trials or further characterization experiments. Expression trials in the eukaryote *Pichia pastoris* neither yielded full length nor truncated Mrr1 protein. In order to overcome the aggregation problem, a shortened variant, missing the N-terminal 249 amino acids named Mrr1 '250', was successfully expressed in *E. coli* and could be purified without aggregation. Similar to the wild type Mrr1 '250', selected gain of function variants were successfully cloned, expressed and purified with varying yields and with varying purity. The Mrr1 '250' construct contains most of the described regulatory domains of Mrr1. It was used for crystallization and an initial comparative analysis between the wild type protein and the variants. The proposed dimeric form of the transcription factor, necessary for DNA binding, could be verified for both, the wild type and the mutant proteins. Secondary structure analysis by circular dichroism measurements revealed no significant differences

Summary

in the overall fold of the wild type and variant proteins. *In vitro*, the gain of function variants seem to be less stable compared to the wild type protein, as they were more prone to degradation. Whether this observation holds true for the full length protein's stability *in vitro* and *in vivo* remains to be determined. The crystallization experiments, performed with the Mrr1 '250' constructs, led to few small needle shaped or cubic crystals, which did not diffract very well and were hardly reproducible. Therefore no structural information of the transcription factor could be gained so far.

Infections with *M. tuberculosis*, the causative agent of tuberculosis, are the leading cause of mortality among bacterial diseases. Especially long treatment times, an increasing number of resistant strains and the prevalence of for decades persisting bacteria create the necessity for new drugs against this disease. The cholesterol import and metabolism pathways were discovered as promising new targets and interestingly they seem to play an important role for the chronic stage of the tuberculosis infection and for persisting bacteria.

In this thesis, the 3-ketoacyl-CoA thiolase FadA5 from *M. tuberculosis* was characterized and the potential for specifically targeting this enzyme was investigated. FadA5 catalyzes the last step of the β -oxidation reaction in the side-chain degradation pathway of cholesterol. We solved the three dimensional structure of this enzyme by X-ray crystallography and obtained two different apo structures and three structures in complex with acetyl-CoA, CoA and a hydrolyzed steroid-CoA, which is the natural product of FadA5. Analysis of the FadA5 apo structures revealed a typical thiolase fold as it is common for biosynthetic and degradative enzymes of this class for one of the structures. The second apo structure showed deviations from the typical thiolase fold. All obtained structures show the enzyme as a dimer, which is consistent with the observed dimer formation in solution. Thus the dimer is likely to be the catalytically active form of the enzyme. Besides the characteristic structural fold, the catalytic triad, comprising two cysteines and one histidine, as well as the typical coenzyme A binding site of enzymes belonging to the thiolase class could be identified. The two obtained apo structures differed significantly from each other. One apo structure is in agreement with the characteristic thiolase fold and the well-known dimer interface could be identified in our structure. The same characteristics were observed in all complex structures. In contrast, the second apo structure followed the thiolase fold only partially. One subdomain, spanning 30 amino acids, was in a different orientation. This reorientation was caused by the formation of two disulfide bonds, including the active site cysteines, which rendered the enzyme inactive. The disulfide bonds together with the resulting domain swap still permitted dimer formation, yet with a significantly shifted dimer interface. The comparison of the apo structures together with the preliminary activity analysis performed by our collaborator suggest, that FadA5 can be inactivated by oxidation and reactivated by reduction. If this redox switch is of biological importance requires further evaluation, however, this would be the first reported example of a bacterial thiolase employing redox regulation.

Summary

Our obtained complex structures represent different stages of the thiolase reaction cycle. In some complex structures, FadA5 was found to be acetylated at the catalytic cysteine and it was in complex with acetyl-CoA or CoA. These structures, together with the FadA5 structure in complex with a hydrolyzed steroid-CoA, revealed important insights into enzyme dynamics upon ligand binding and release. The steroid-bound structure is as yet a unique example of a thiolase enzyme interacting with a complex ligand. The characterized enzyme was used as platform for modeling studies and for comparison with human thiolases. These studies permitted initial conclusions regarding the specific targetability of FadA5 as a drug target against *M. tuberculosis* infection, taking the closely related human enzymes into account. Additional analyses led to the proposal of a specific lead compound based on the steroid and ligand interactions within the active site of FadA5.

Zusammenfassung

Sowohl die zunehmende Verbreitung von Resistenzen gegen die eingesetzten Antimikrobiotika in Bakterien, Pilzen, Viren oder Parasiten, als auch häufig auftretende negative Wechselwirkungen zwischen Medikamenten verstärken die Notwendigkeit Resistenzprozesse zu verstehen und neue Medikamente zu entwickeln. In dieser Arbeit wurden zwei Ansätze zur Entwicklung neuer antimikrobieller Wirkstoffe anhand ausgesuchter Proteine verfolgt: Zum einen wurde der Transkriptionsfaktor Mrr1 charakterisiert, der die Expression eines Transporters in *Candida albicans* reguliert, welcher für multiple Resistenzen verantwortlich ist. Zum anderen wurde die strukturelle Charakterisierung der Thiolase FadA5 verfolgt, welche ein wichtiges Enzym im Cholesterin-Metabolismus von *Mykobakterium tuberculosis* darstellt.

Die Zahl der Pilzinfektionen, welche hauptsächlich durch den opportunistisch-pathogenen Pilz *C. albicans* verursacht werden, ist nicht nur in Deutschland, sondern weltweit steigend. Die auftretenden Infektionen betreffen vor allem immunsupprimierte Personen. Dieser Anstieg an Pilzinfektionen verursacht häufigere und immer länger andauernde Behandlungen und resultiert auch im vermehrten Auftreten von Resistenzen gegen Antimykotika, unter anderem gegen das klinisch eingesetzte Fluconazol. Eine Möglichkeit der Resistenzbildung in *C. albicans* ist die Expression der ‚Multidrug resistance pump 1‘ (Mdr1), einer Membranpumpe, die zur Major-Facilitator-Superfamilie zählt. Diese durch Mdr1-vermittelte Fluconazolresistenz wird durch den Mdr1 regulierenden Transkriptionsfaktor Mrr1 (‚Multidrug resistance regulator 1‘) gesteuert. In resistenten *C. albicans* Stämmen befindet sich Mrr1 ohne weitere Stimulation oder externe Aktivierung bereits in einem hyperaktiven Zustand, der durch Mutationen mit Funktionsgewinn im *MRR1* Gen verursacht wird.

Um die Mechanismen, die sich hinter der konstitutiven Aktivität von Mrr1 verbergen, zu entschlüsseln, sollte dieser Transkriptionsfaktor *in vitro* strukturell und funktionell charakterisiert werden. Diese Charakterisierung könnte im Anschluss genutzt werden, um Wirkstoffe gegen die von Mrr1 gesteuerte und von Mdr1-vermittelte Resistenz zu entwickeln. Zu diesem Zweck, wurde das gesamte, 1108 Aminosäuren umfassende, Protein in *Escherichia coli* exprimiert. Die anschließende Proteinreinigung war allerdings durch Aggregatbildung beeinträchtigt, welche Kristallisationsansätze oder eine weitere Charakterisierung dieses Proteinkonstruktes verhinderten. Im Eukaryot *Pichia pastoris* durchgeführte Expressionsanalysen, waren leider erfolglos und weder die Expression des Volllängen-Mrr1 noch seiner verkürzten Proteinvarianten konnte nachgewiesen werden. Um Proteinaggregation zu umgehen, wurde deshalb ein N-terminal, um 249 Aminosäuren, verkürztes Proteinkonstrukt, Mrr1 ‚250‘, in *E. coli* exprimiert und erfolgreich, ohne Aggregation, gereinigt. Zusätzlich zum wildtypischen Mrr1 ‚250‘ Protein wurden auch ausgewählte Varianten kloniert,

Zusammenfassung

exprimiert und gereinigt, allerdings mit unterschiedlicher Ausbeute und Reinheit. Da das verkürzte Mrr1 ,250' Protein noch immer fast alle in der Literatur beschriebenen Regulierungsdomänen besitzt, wurde es zur Kristallisation und für einen initialen Vergleich zwischen Wildtyp und Varianten genutzt. So konnte zum Beispiel die vermutete Dimerisierung des Transkriptionsfaktors sowohl für das Wildtypprotein als auch für die Varianten gezeigt werden. Eine weiterführende Untersuchung der Sekundärstruktur mittels zirkularer Dichroismus Messungen zeigte keine signifikanten Unterschiede zwischen den Mutanten und dem Wildtypprotein. Allerdings erscheinen die Funktionsgewinn Varianten von Mrr1 *in vitro* instabiler als das Wildtypprotein, was sich durch stärkeren Abbau der Variantenproteine zeigt. Ob diese Beobachtungen allerdings vom verkürzten Protein auf das Gesamtprotein und dessen *in vitro* und *in vivo* Stabilität übertragbar sind, ist derzeit noch unklar. Kristallisationsansätze, die mit den verschiedenen Varianten des Mrr1 ,250' Konstrukts durchgeführt wurden, führten zu sehr wenigen, nadelförmigen oder kubischen Kristallen, die kaum reproduzierbar waren und schlecht diffraktierten. Bisher konnten deshalb keine strukturellen Daten für den untersuchten Transkriptionsfaktor erhalten werden.

Noch immer sind Infektionen, die durch *M. tuberculosis*, dem Erreger der Tuberkulose, verursacht werden die Haupttodesursache im Bereich der bakteriellen Infektionen. In diesem Zusammenhang stellen vor allem lange Behandlungszeiten, das vermehrte Auftreten resistenter Stämme und das Vorkommen persistierender Bakterien, die Jahrzehnte in ihrem Wirt überdauern können, nach wie vor große Herausforderungen dar und die Entwicklung neuer Tuberkulosemedikamente ist dringend erforderlich. Sowohl der Cholesterinimport als auch dessen Stoffwechselweg wurden als vielversprechende Wirkstoffziele identifiziert. Nicht zuletzt, da beide Mechanismen eine wichtige Rolle während der chronischen Phase der Tuberkuloseinfektion und für persistierende Bakterien zu spielen scheinen.

Im Laufe dieser Arbeit wurde die 3-ketoacyl-CoA Thiolase FadA5 aus *M. tuberculosis* strukturell charakterisiert und auf ihre Tauglichkeit als spezifisches Wirkstoffziel hin untersucht. FadA5 katalysiert den letzten Schritt der β -Oxidation im Zuge des Seitenkettenabbaus von Cholesterin. Wir konnten die Proteinstruktur des FadA5 Proteins mittels Röntgenkristallographie ermitteln und erhielten zwei unterschiedliche apo-Strukturen sowie drei Komplexstrukturen. In den Komplexstrukturen waren entweder Acetyl-CoA, CoA oder ein hydrolisiertes Steroid-CoA, welches das natürliche Produkt von FadA5 darstellt, an das Enzym gebunden. Die Strukturanalyse der apo-Strukturen lies für eine der beiden Modelle die typische Thiolasefaltung erkennen, welche für biosynthetische und degradative Enzyme dieser Klasse üblich ist. In der zweiten apo-Struktur konnte diese Faltung nur teilweise identifiziert werden. Das Protein liegt in allen erhaltenen Strukturen als Dimer vor, was auch in Lösung beobachtet werden konnte und darauf hinweist, dass das Dimer die katalytisch aktive Form des Proteins darstellt. Neben der charakteristischen Faltung, wurde das

Zusammenfassung

aktive Zentrum, bestehend aus zwei Cysteinen und einem Histidin, sowie die für Thiolasen übliche Coenzym A Bindetasche identifiziert. Die erhaltenen apo-Strukturen unterschieden sich deutlich voneinander. Die zuvor beschriebene typische Dimer-Interaktionsfläche wird auch in den Komplexstrukturen beobachtet. Dahingegen war die Thiolasefaltung in der zweiten Apo-Struktur nur teilweise vorhanden, da beispielsweise eine Domäne, die 30 Aminosäuren umfasst, umorientiert vorlag. Die Bildung zweier Disulfidbrücken, welche beide katalytischen Cysteine involviert, verursachte die beschriebene Umorientierung und damit gepaart eine wahrscheinliche Inaktivität des Enzyms. Trotz der beschriebenen Umorientierung und Disulfidbrückenbildung liegt das Protein noch immer als Dimer vor, allerdings mit einer deutlich verschobenen Interaktionsfläche. Der Vergleich der beiden apo-Strukturen in Kombination mit einer vorläufigen Aktivitätsanalyse, die von unseren Kollaborationspartnern durchgeführt wurde, lassen vermuten, dass FadA5 durch Oxidation inaktiviert und durch Reduktion reaktiviert werden kann. Ob diese Redoxregulierung biologisch relevant ist, muss noch geklärt werden, allerdings wäre dies der erste beschriebene Fall einer redoxregulierten bakteriellen Thiolase.

Die Komplexstrukturen stellen verschiedene Stufen der Thiolasereaktion dar. In einigen dieser Strukturen lag FadA5 am katalytischen Cystein acetyliert vor und befand sich im Komplex mit acetyl-CoA oder CoA. Durch eine weitere Struktur, in der FadA5 im Komplex mit einem hydrolysierten Steroid-CoA vorlag, konnten wichtige Einblicke in die Enzymdynamik während der Ligandenbindung und Freisetzung gewonnen werden. Die Steroid gebundene Struktur stellt derzeit ein einzigartiges Beispiel einer Thiolase im Komplex mit einem großen, mehrere Ringsysteme umfassenden Liganden dar. Das charakterisierte Enzym diente als Ausgangspunkt für Modellierungsversuche und Vergleiche mit humanen Thiolasen. Diese Analysen erlaubten initiale Schlussfolgerungen bezüglich einer Verwendung von FadA5 als spezifisches Wirkstoffziel gegen Tuberkuloseinfektionen, im Kontext verwandter humaner Enzyme. Zusätzliche Untersuchungen ermöglichten die Ausarbeitung einer spezifischen Leitsubstanz, die auf den analysierten Interaktionen zwischen dem aktiven Zentrum von FadA5 und den gebundenen Liganden basiert.

1 Introduction

1.1 The emergence of antimicrobial resistances

The World Health Organization recently published the first public report to address antimicrobial resistance. This report elucidates the emerging resistances of bacteria, parasites, viruses or fungi. The resistance of these organisms has already caused far-reaching consequences, including generating a substantial health threat for future generations with a subsequent permanent increase in healthcare costs. The report describes an increasing number of resistances not only for already-known candidates, such as *Mycobacterium tuberculosis* or *Staphylococcus aureus*, but also for the human immunodeficiency virus (HIV), plasmodium species and *Candida albicans*. (WHO, 2014)

The increasing number of fungal infections of *Candida* or other species (*Aspergillus fumigatus*, *Cryptococcus neoformans*) is ironically caused by improvements in health care. The surge in antibiotic treatments and the rising number of immunocompromised patients lead to an increasing number of infections of the above mentioned opportunistic fungi. (Morschhauser, 2010a) Resistance development in the above-mentioned organisms are natural processes that are accelerated by selective pressure, which is established by the extensive use and misuse of antimicrobial agents. Naturally, a victory over bacteria, fungi, viruses or parasites is impossible but an increase of surveillance programs, the ability to decipher resistance mechanisms and the intensified development of antimicrobial agents will at least provide mankind an opportunity to survive. (Fair and Tor, 2014; WHO, 2014)

1.2 Fungal infections, resistance in *Candida albicans* and the role of the transcription factor Mrr1

1.2.1 Fungal infections

Fungal infections were for a long time a hardly recognized disease type due to the prevalence of bacterial infections such as the plague and tuberculosis. With the advent of antibacterial therapy and later immunosuppressing therapies, new niches for fungal expansion in the human host were created and this is documented by an increase in fungal infections since the 1960s. The wide prevalence of immuno-compromising diseases, such as AIDS, cancer or diabetes, as well as an increasing number of transplantations (organ or bone marrow) contribute to a rise in fungal infections. (Vandeputte et al., 2012)

Introduction

In general, infections are caused by opportunistic or primary pathogens. Primary pathogens are able to infect healthy humans unlike opportunistic fungi which are restricted to immuno-compromised persons. Fungal infections can be classified by the infected tissue type, where superficial and cutaneous infections with primary pathogens, such as *Malassezia subspecies*, *Epidermophyton ssp.* or *Microspora ssp.*, are defined by damage caused to keratinized structures e.g. skin, hair or nails. Mucosal or systemic infections are mainly caused by opportunistic fungal pathogens, such as *Aspergillus ssp.*, *Cryptococcus ssp.*, *Pneumocystis ssp.*, *Candida ssp.* and many others, that colonize mucosal surfaces or organs. In this work, the genus *Candida*, specifically *Candida albicans* were of interest. (Vandeputte et al., 2012)

1.2.2 *Candida* species and *Candida albicans*

More than 200 different ascomycetous yeasts are classified in the genus *Candida*; however, less than 20 are opportunistic human pathogens. (Karkowska-Kuleta et al., 2009) All members of this genus grow as yeasts, but some of them are in addition forming pseudohyphae or true hyphae structures.



Figure 1 - Image of a SABHI agar plate culture of the fungus *Candida albicans* grown at 20 °C

Kindly provided by CDC (Kaplan, 1969)

Pseudohyphae are built from yeast cells or hyphae structures by budding without a final separation of the daughter cell from the parent cell. (Calderone, 2002) The *candida ssp.* are ubiquitous yeasts and the family members, which are not pathogenic to humans, are environmental saprophytes. In contrast to the non-pathogenic species, the pathogens are rarely found in the natural environment but are mainly localized on or in their according host. (Calderone, 2002)

The most prominent member of the *Candida* genus, which causes the majority of all *Candida* infections, is *Candida albicans*, it is named after its white (*lat. albus*), creamy colony color (**Figure 1**). However, recently the percentage of *Candida albicans* caused *candida* infections is decreasing and other genus members, such as *C. glabrata* or *C. krusei*, are on the rise. (Karkowska-Kuleta et al., 2009) *C. albicans* is mainly associated to warm-blooded animals, for example humans, and there to the colonization of the skin, mucosal surfaces or the gastrointestinal tract. This fungus is a diploid organism that contains eight chromosomes, it normally appears in unicellular form but can also build multicellular, filamentous hyphae or pseudohyphae structures, and therefore it is considered to be polymorphic. (Calderone, 2002; Feldmann) Another remarkable characteristic of *C. albicans* is the unusual codon usage. A CUG triplet, which normally translates to the amino acid leucine, encodes a serine in *C. albicans*. (Santos and Tuite, 1995)

Introduction

The infections caused by *C. albicans* can be of superficial character, such as esophageal, oropharyngeal, vaginal or urinary tract candidiasis, where mucosal surfaces are affected. However, in addition to the rather harmless mucosal infections, life-threatening systemic infections, also referred to as invasive candidiasis, can occur. (Calderone, 2002) Initial bloodstream infections provide the means for the fungi to reach every inner organ of the human body, to colonize it and, thereby, to cause severe tissue damage and life-threatening disease. (Karkowska-Kuleta et al., 2009)

1.2.3 Virulence factors

To assess the virulence of *C. albicans*, many factors, such as phenotype, morphology, adhesion, biofilm formation or secretion processes, are highly important.

Phenotypic switching describes processes that lead to subpopulations with different morphology, appearance or surface properties. (Karkowska-Kuleta et al., 2009) A process called white-opaque switching, in which normally white, round cells change their appearance into elongated, gray-colored cells, seems to be an important phenotype switch for the colonization of specific niches. (Slutsky et al., 1987) The opaque phenotype is less virulent, but represents the only mating form of *C. albicans*, which before was regarded as asexual. *In vitro* and *in vivo* sexual reproduction has thus far been observed; however, natural mating of *C. albicans* is yet to be recorded. (Hull and Johnson, 1999; Magee and Magee, 2000) White-opaque switching is a reversible process that is dependent on the homozygosity of the mating type locus (MTL). (Miller and Johnson, 2002) Homozygosity might be an important factor not only in niche colonization, but also for the persistence of coexisting opaque and white cells in infected tissue. (Morschhauser, 2010b)

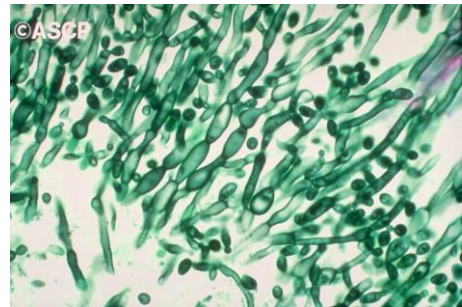


Figure 2 - Histopathology of Candida albicans infection.

Pseudohyphae and true hyphae shown by methenamine silver stain. Kindly provided by the American Society for Clinical Pathology (ASCP) for copyright see 7 (Pathologists)

The morphological dimorphism of *C. albicans* refers to its ability to form unicellular yeast cells and to also build pseudohyphae or hyphae structures (Figure 2). (Karkowska-Kuleta et al., 2009) This morphology change is dependent on environmental conditions, such as temperature, gas content, pH values, and the presence or absence of serum or other additives. Different studies investigated the regulation of dimorphism and identified for example the involvement of distinct activating (Cph1p (Lo et al., 1997)) and repressing (Tup1p or Rbp1p (Calderone and Fonzi, 2001)) transcription factors. The different morphological forms are very important for the organism as, on the one hand, the yeast cells are likely to contribute mainly to dissemination, and, on the other hand, the hyphae form seems to be responsible for virulence and invasion processes. (Karkowska-Kuleta et al., 2009)

Introduction

Experiments with *C. albicans* yeast and hyphae gene knockouts showed less virulent phenotypes upon deletion. (Yang, 2003) Pseudohyphae are thought to contribute to the exploitation of 'distant' nutrient sources. (Berman and Sudbery, 2002)

The initial step for colonization of a new host is the yeasts adhesion to the surface, e.g. the human skin, followed by biofilm formation. This process is achieved by the interplay of many different fungal adhesins with host receptors. A large variety of fungal adhesins have already been identified but many mechanisms and interplayers remain unknown, especially those of the host organism. Proteins described as Als (agglutinin-like sequence) proteins have been identified as being responsible for the adherence to endothelial or epithelial cells as well as to collagen, fibronectin or laminin. (Karkowska-Kuleta et al., 2009) Other adhesins are described which can bind to similar cells or proteins but can also adhere to platelets. An example of known human interaction partners are the Toll-like receptors 2 and 4, and other such receptors located on cells of the human immune system. Adhesion and biofilm formation are important factors for fungal pathogenicity. (Karkowska-Kuleta et al., 2009) Especially fungal biofilms on artificial surfaces, such as catheters or medical devices, are causing severe problems. (Baillie and Douglas, 2000)

Further examples of important virulence factors are proteins that help to overcome reactive oxygen species released by the host or proteins that enable iron requisition during host colonization, a process which is needed by the fungus for different catalytic reactions, such as ergosterol synthesis. Secreted hydrolytic enzymes, such as phospholipases, lipases and secreted aspartyl proteinases (SAPs), are important virulence factors as they contribute to dissemination processes, tissue damage or support fungal nutrition. (Karkowska-Kuleta et al., 2009)

1.2.4 Antimycotics

The current antifungal therapies, nicely reviewed by Vandeputte *et al.* and Sanglard *et al.*, comprise antimycotics that belong to different molecular classes and that are active against different yeast strains. (Sanglard, 2002b; Vandeputte et al., 2012) Because amphotericin B was the only available drug and fungal infections were out of the public focus for a very long time, severe infections were mainly treated with amphotericin B, a polyene with problematic side effects. Since the 1980's the range of clinically-approved antimycotics has increased following the acceptance of azoles, base analogs, such as fluoropyrimidines, and, most recently, of the echinocandins. (Ghannoum and Rice, 1999)

Introduction

1.2.4.1 Polyenes

Although polyenes constitute a molecular class with an abundance of compounds, very few have found an application as an antifungal agent. The polyenes were originally isolated from bacteria of the *Streptomyces* genus and, due to economic constraints, the primary production of polyenes for clinical applications is still dependent on bacterial organisms. As mentioned above, polyenes, especially amphotericin B (**Figure 3A**), were formerly the predominant antimycotics applied against systemic fungal infections. Other clinically relevant compounds of this class are nystatin or natamycin; however, both are applied orally or locally in contrast to the often intravenous application of amphotericin B. Polyenes are not only active against *C. albicans* but are commonly applied against various fungi, such as *Candida* ssp., *Aspergillus*, *Cryptococcus*, *Fusarium*, *Trichosporon*, etc.

Polyenes are amphiphilic molecules with a hydrophobic polyene side chain and a hydrophilic, hydroxyl-substituted, alkane chain to produce an overall cyclic molecule. (**Figure 3A**) Due to their structure, polyenes can intercalate between neighboring ergosterol molecules of the fungal membranes, sometimes in a 1:1 ratio. This intercalation leads to pore-formation, which destabilizes the membrane and produces leakage of the cell that results in detrimental, uncontrolled ion and chemical in- and efflux.

Polyenes have a high affinity to ergosterol, which is the main sterol in fungal cell membranes but not present in humans. Nevertheless, polyenes also display a low affinity to cholesterol, which can account for the reported side effects, such as nephrotoxicity. (Moen et al., 2009)

1.2.4.2 Fluoropyrimidines

This class of antifungal drugs represents structural base analogs of cytosine (**Figure 3B/C**) and inhibits DNA- and protein synthesis through their incorporation into DNA or RNA. 5-fluorocytosine (**Figure 3B**) is converted into 5-fluorouracil (5-FU, **Figure 3C**) by cytosine deaminase after entering the fungal cell. The 5-fluorouracil is first converted into 5-FU monophosphate and subsequently into 5-FU triphosphate or 5-fluorodeoxyuridine monophosphate, which then hinder protein synthesis or a key enzyme of DNA synthesis, the thymidylate synthase. (Vandeputte et al., 2012) In general, the activity of these nucleotide base analogs is not restricted to the kingdom of fungi (*Candia* and *Cryptococcus*) and it has been demonstrated that they are also active against certain protozoa. (Stiller et al., 1983)

The application of fluoropyrimidines is mainly restricted to combination-therapy, as the target fungi often quickly develop resistances against these drugs. Due to the side effects of the polyenes, the combination-therapy, which originally included fluoropyrimidines and polyenes, was adapted to

Introduction

include azoles as an additional therapeutic agent. The cytosine base analogs themselves normally show negligible adverse effects with rare exceptional cases. (Vandeputte et al., 2012)

1.2.4.3 Azoles

The azoles can be subdivided into two groups, that contain the active nitrogen either in an imidazole ring (**Figure 3D**), or in a triazole ring (**Figure 3E**). The free nitrogens of these compounds coordinate with the iron of the heme group of the important lanosterol 14 α -demethylase, which is encoded by *ERG11*. This enzyme is a critical player in the ergosterol biosynthetic pathway of many fungi and the inhibition of this enzyme leads to the accumulation of toxic methylated compounds. (Sanglard, 2002b) A positive side effect of azole treatment is the fact, that not enough ergosterol, the

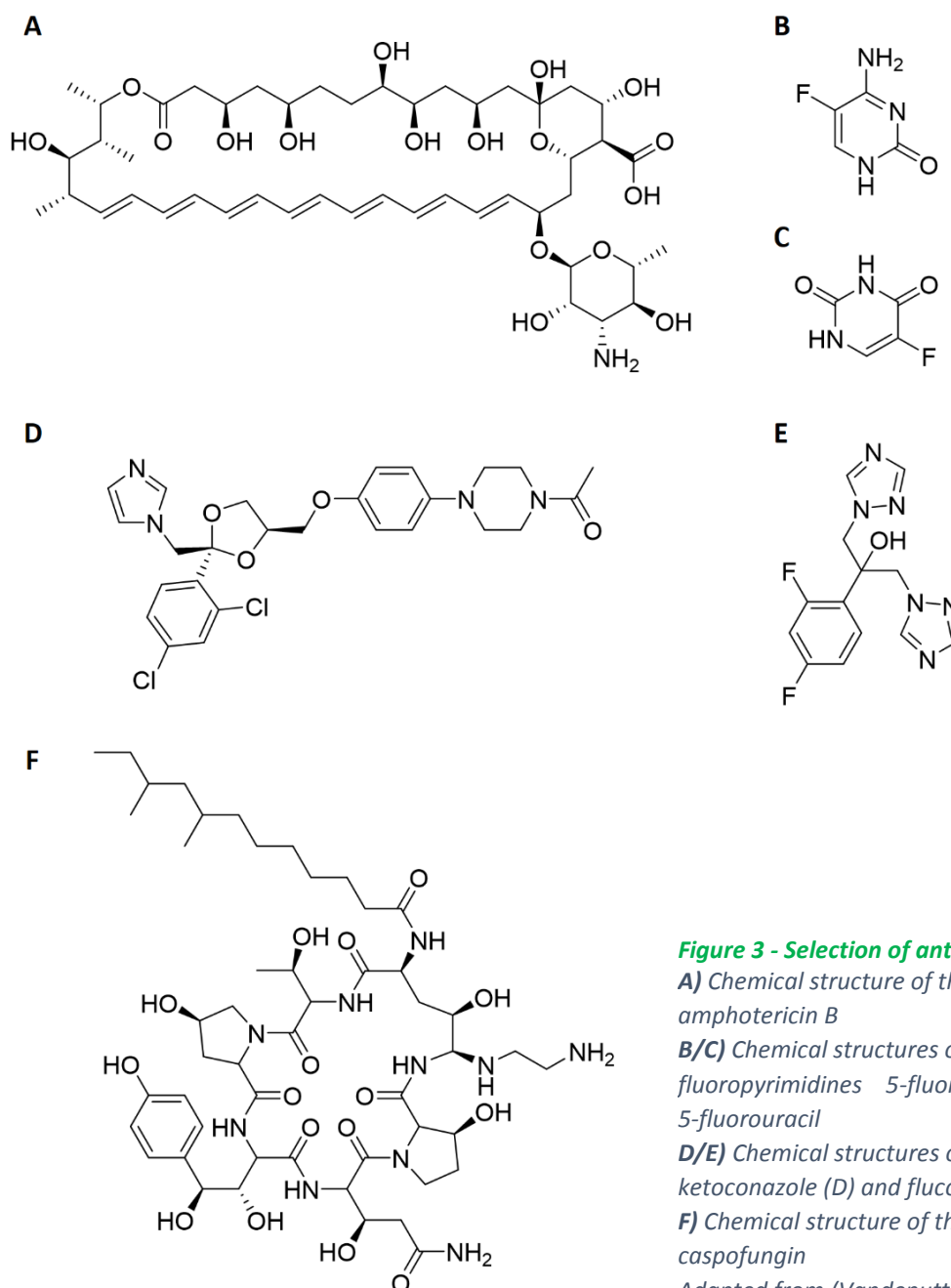


Figure 3 - Selection of antifungal agents:

A) Chemical structure of the polyene amphotericin B

B/C) Chemical structures of the fluoropyrimidines 5-fluorocytosine (B) and 5-fluorouracil

D/E) Chemical structures of the azoles ketoconazole (D) and fluconazole (E)

F) Chemical structure of the echinocandins caspofungin

Adapted from (Vandeputte et al., 2012)

Introduction

ubiquitous membrane brick, is produced, which influences membrane stability.

Several azoles are used in antifungal therapy, of which ketoconazole (D) and fluconazole (E) are depicted in **Figure 3**. The azoles represent the most commonly used antimycotics for the treatment of both superficial and systemic infections. Imidazole-based compounds are usually administered for superficial therapy, whereas systemic infections are mainly treated with triazole compounds. (Vandeputte et al., 2012) Both compounds are active against many different species, such as *Aspergillus*, *Cryptococcus*, *Penicillium*, *Trichosporum* and *Candida*.

A drawback of the azole therapies in the last decades is the increasing number of resistances, especially against the frequently applied fluconazole. These resistances also lead to cross-resistances with newly developed antimycotics based on a triazole scaffold. Another problem in the clinical use of fluconazole is the possibility of drug interactions with HIV medications or other chemotherapeutics leading to an overall increase in toxicity. (Vandeputte et al., 2012)

1.2.4.4 Echinocandins

The echinocandins are not only the most recent class of antifungal drugs to have been approved in the last 15 years, they are also the only available drugs that target the fungal cell wall. The fungal cell wall is an important target, as most of its components are not present in humans and the cell wall provides a physical barrier against the human immune system. The echinocandins are synthetic derivatives of cyclic lipopeptides (e.g. **Figure 3F**) and are synthesized by fungi. They inhibit $\beta(1-3)$ -glucan synthase in a non-competitive manner to prevent the production of glucans necessary for cell wall rigidity and integrity. (Perlin, 2007)

To date three echinocandins have been approved for antifungal therapy: caspofungin (**Figure 3F**), micafungin and anidulafungin, respectively. They are all applied intravenously with few reported side effects. Echinocandins are often applied in combination-therapy with azoles or amphotericin B, and synergistic effects have been observed. In addition, they provide a broad spectrum of activity against various fungal species with the exception to *Fusarium* or *Trichophyton* genera as well as *C. neoformans*. (Vandeputte et al., 2012)

1.2.4.5 Others

In addition to the above-mentioned groups, several inhibitors targeting the ergosterol biosynthesis are used medicinally. Ergosterol biosynthesis, is a valid target as its pathway includes several enzymes that are unique to fungi. However, the efficacy of the here-described compounds against *C. albicans* and *Candida* ssp. is limited. Squalene epoxidase, encoded by *ERG1*, is targeted by

Introduction

allylamines and thiocarbamates, such as terbinafine and tolnaftate. Two additional enzymes of the ergosterol biosynthesis pathway, sterol Δ^{14} reductase (*ERG24*) and Δ^{7-8} isomerase (*ERG2*), are inhibited by morpholines. (Niewerth and Korting, 2000) Due to their numerous adverse effects, the mentioned antimycotics are nearly exclusively applied against dermatophytes and resulting diseases, such as tinea capitis or tinea pedis. (Vandeputte et al., 2012)

1.2.5 Resistance mechanisms in *C. albicans*

1.2.5.1 Structure specific and independent resistance mechanisms

Two different kinds of resistance mechanisms have been identified in *C. albicans*. These resistance mechanisms can evolve either dependently or independently to the chemical structure of the above-described antimycotics and respectively produce a specific or a multidrug resistance. (Morschhauser, 2010a)

Resistance mechanisms have evolved for all previously described antimycotics. 5-Fluorocytosin resistance is mediated by mutations that hinder the uptake of antimycotics or that prevent the intracellular conversion of 5-FU to metabolites that would enter the cytosine metabolism pathways. (Polak and Scholer, 1975)

Several resistance mechanisms are connected to the important cell membrane component ergosterol. A rather rare mechanism of resistance is used against the polyenes, namely that the overall content of ergosterol in the membrane is decreased, which reduces the ability for intercalation of the polyenes with the ergosterol molecules, thereby impairing pore formation. On the other hand, point mutations in *ERG11* that prevent azole binding to lanosterol 14 α -demethylase or that increase the respective protein expression are well-described mechanisms of azole compound resistance. (Morschhauser, 2010a) Overexpression of the *ERG11* gene and other ergosterol biosynthesis genes, are for example caused by mutations in the regulating transcription factor Upc2. (Dunkel et al., 2008b) In addition to *ERG11* connected measures, alterations in the sterol biosynthesis pathways to avoid the accumulation of the toxic ergosterol biosynthesis side products have been observed. All these measures reduce the sensitivity of the cell to the azole compounds 30 to 2000-fold. (Douglas, 2003) Finally, not only the adopted resistance mechanisms against specific antimycotics are problematic, but cross-resistances to several drugs cause an additional challenge that frequently occurs for azole resistances and complicates the long-term applicability of newly-developed and established azole antimycotics. (Vandeputte et al., 2012)

Introduction

The more recently introduced echinocandins are already facing resistance processes, which are caused by point mutations in subunits of the $\beta(1-3)$ -glucan synthase, more specifically, the *FKS1* gene. Some mutations in this gene are causing cross-resistance against all clinically applied substances of the echinocandins. (Perlin, 2007)

In contrast to the substance specific alterations, a commonly observed resistance mechanism in *C. albicans* is the overexpression of drug unspecific efflux pumps. The enhanced drug efflux reduces the compound concentration in the fungal cell and thereby its detrimental effect(s). This mechanism of resistance is called multidrug resistance (MDR). (Morschhauser, 2010a) Two different pump types that are localized in the cytoplasmic membrane of the fungal cell have been described: the ATP-binding cassette (ABC) transporters and pumps belonging to the major facilitator superfamily (MFS). ABC transporters are dependent on ATP hydrolysis and the MFS pumps require a proton gradient to facilitate drug efflux or the export of endogenous metabolites. (Del Sorbo et al., 2000) Initial information on these efflux pumps has been gained through the investigation of *Saccharomyces cerevisiae*; the observations from these studies are possibly transferable to the resistance mechanisms in *C. albicans*. *Candida* drug resistance pumps 1 and 2 (Cdr1 and Cdr2) are representatives of the ABC transporters, whereas the multidrug resistance pump 1 (Mdr1) belongs to the MFS pumps. (Morschhauser, 2010a)

Sanglard *et al.* demonstrated in the 1990s that a reduced intracellular fluconazole concentration was connected to an increased Mdr1 expression as well as an increased Cdr1 and Cdr2 expression. (Sanglard et al., 1997; Sanglard et al., 1995) Although at a low level, Cdr1 expression is always observed in drug susceptible strains; however, neither Cdr2 nor Mdr1 are expressed in these strains. (Morschhauser, 2010a) Deletions of the Cdr-pump encoding genes in azole-resistant clinical isolates showed that Cdr2 is of lesser importance for drug resistance than Cdr1. Deletion of *CDR1* led to clearly increased drug susceptibility whereas the deletion of *CDR2* showed minor effects only. Hypersusceptibility was observed when both genes were deleted, which reinforces the importance of both pumps for MDR. (Tsao et al., 2009) Similarly to *CDR2*, *MDR1* is minimally expressed in drug-susceptible strains, but expression is significantly increased in resistant *C. albicans* isolates. The deletion of *MDR1* in fluconazole-resistant clinical isolates resulted in diminished fluconazole resistance. (Wirsching et al., 2000b)

Similar resistance mechanisms employing efflux pumps were reported for other *Candida* ssp. and are probably also applicable for *A. fumigatus*.

Introduction

1.2.5.2 Regulation of efflux pump expression

With the knowledge in mind that efflux pumps are important resistance factors, many studies were performed to identify regulating factors, such as regulating proteins or chemicals, and to elucidate the control mechanism of the observed pump overexpression. In general it was found that pump expression is triggered by distinct chemicals such as H₂O₂, benomyl (fungicide), estradiol (hormone) or other environmental signals. (de Micheli et al., 2002; Harry et al., 2005) In addition, investigations from de Micheli *et al.* and Wirsching *et al.* revealed that overexpression was caused by mutations located on the respective protein expression transcription factors but not by mutations in promoter regions (de Micheli et al., 2002; Wirsching et al., 2000a). Therefore, overexpression is mediated by *trans*-regulatory elements. This insight sparked further investigations of the involved *trans*-regulatory elements, but also included the identification of binding regions in the corresponding promoter (*cis*-regulatory elements).

Further studies identified areas in the promoter regions (*cis*-regulatory elements) of the transporters that mediate basal and/or hyperactive pump expression. The Binding of the transcription factors to the identified promoter regions is normally stress induced. The regulatory regions of *CDR1* and *CDR2* are significantly differing in sequence but nevertheless share a common part of the promoter, namely the drug-responsive element (DRE), which is responsible for compound-triggered expression upregulation. (de Micheli et al., 2002) The *MDR1* expression is regulated among others by promoter regions that respond to benomyl induced expression, including the so called benomyl response element (BRE). (Harry et al., 2005; Rognon et al., 2006) In addition, *cis*-regulatory elements responding to oxidative stress, such as the H₂O₂ response element (HRE), were identified. (Rognon et al., 2006)

Parallel to the identification of important promoter regions, *trans*-regulatory elements responsible for expression and overexpression of the respective transporter genes were identified. The basic leucine zipper domain (bZip) transcription factor Cap1 (*C. albicans* AP-1) binds to the HRE-binding region and regulates MDR1 overexpression, caused by oxidative stress. Cap1 might also be involved in MDR. (Alarco and Raymond, 1999; Schubert et al., 2011a) Another transcription factor that was identified to bind to the *MDR1* promoter and regulate Mdr1 expression is Mcm1, a MADS-Box transcription factor representative. This transcription factor likely has a regulatory role as it is necessary for Mdr1 overexpression in combination with another transcription factor named Mrr1 (multidrug resistance regulator 1). (Mogavero et al., 2011) Mrr1 is a zinc cluster protein with a specific, zinc-binding motif for two zinc ions, which is exclusively present in fungi. (MacPherson et al., 2006) Mrr1 is involved in the regulation of several MDR related genes, including the Mdr1 transporter. Mrr1 is responsible for benomyl and H₂O₂-induced pump expression. Importantly, the

Introduction

complete elimination of Mdr1 overexpression and Mdr1 mediated MDR was observed following gene deletions of *MRR1*. (Dunkel et al., 2008a; Morschhauser et al., 2007)

The two ABC transporters are regulated by the transcriptional activator of *CDR* genes (Tac1) transcription factor, which is also a zinc cluster protein. Tac1 binds to the described DRE regulatory element and deletions of the *tac1* gene in resistant clinical isolates resulted in the loss of Cdr1 and Cdr2 overexpression and led to the restoration of drug susceptibility. Thereby Tac1 is similar to Mrr1 as an important MDR element. (Coste et al., 2004)

In gene isolates of *MRR1* and *TAC1* of resistant clinical *C. albicans* strains several gain of function mutations could be identified. For many of these mutations homozygosity was observed, which indicates a clear advantage of the respective mutations for the organism, under drug selection pressure. It was shown that drug resistance was heavily increased when both alleles contained the mutation. In particular, heterozygous *tac1* mutants had only minor effects on Cdr1 and Cdr2 pump expression. The enormous advantage of homozygosity might be linked to the fact that the dimer is the functional entity of many zinc cluster proteins. (Coste et al., 2007; Coste et al., 2006; Dunkel et al., 2008a; MacPherson et al., 2006; Morschhauser et al., 2007)

The investigation of the activation and regulation processes of the identified transcription factors is still in progress; however, it has been shown that Mrr1 and Tac1 are able to induce their own expression. (Morschhauser, 2010a) Analyses of the transcription factor Yap1 from *S. cerevisiae* revealed that structural changes, induced by disulfide bond formation, are a response to oxidative stress and lead to target gene expression. (Yan et al., 1998) The reported gain of function mutations, which result in constitutively active transcription factors, might also cause structural rearrangements responsible for the activity. Therefore chemically induced or mutation related structural changes might be involved in the regulation of the transcriptional activity of the Mrr1 and Tac1 proteins.

1.2.6 Multidrug resistance regulator 1 (Mrr1)

The multidrug resistance regulator 1 (Mrr1) is a transcription factor deeply entangled in the fluconazole drug resistance processes of the fungus *C. albicans*. An Mrr1 homolog in *C. dubliniensis* was found to also be involved in azole resistance. (Schubert et al., 2008)

Mrr1 is, as alluded to above, an 1108 amino acid long transcription factor belonging to class III of the zinc cluster protein family, which is unique to fungi. A prominent and intensively-studied protein of this class is Gal4 from *S. cerevisiae*. (Marmorstein et al., 1992) Similar to Gal4, the typical Zn-binding motif could be identified in Mrr1, which is $CX_2CX_6CX_5CX_2CX_8C$, and is also called zinc finger. An

Introduction

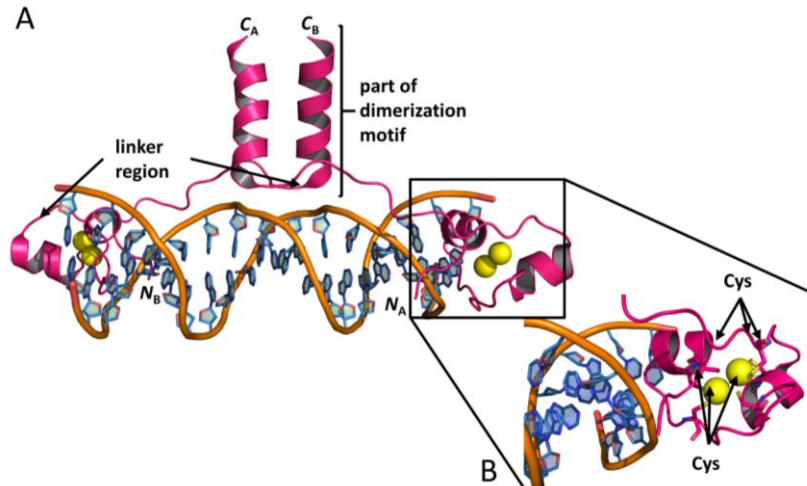


Figure 4 - Crystal structure of the DNA-binding domain (DBD) of Gal4:

A) The dimer of a part of the Gal4 DBD illustrates the suggested metal-binding in Mrr1. The protein (magenta) is shown in complex with DNA (orange and blue) and two Cd^{2+} -ions (yellow) per monomer. The Zn^{2+} was replaced by Cd^{2+} due to crystallographic reasons. **B)** Zoom in to the metal binding region with the six cysteines (black arrows) coordinating two ions (yellow). Adapted from pdb entry 1D66 and (Marmorstein et al., 1992)

interesting characteristic of the Zn-finger is its ability to bind two Zn^{2+} ions. (MacPherson et al., 2006) This is illustrated in the structure of the DNA binding domain (DBD) of Gal4 in complex with DNA (Figure 4). The structure contains a dimer of 57 amino acids per monomer, namely the N-terminal Gal4 residues E8 to L64, in complex with DNA. The depicted DBD is comprised of the zinc finger,

at the N-terminus, a linker and a dimerization motif at the C-terminus. Due to crystallographic constraints, the metal ions in the structure are cadmium ions instead of zinc ions. The metal-binding cysteines are embedded into two helices, which are C-terminally connected to the dimerization motif via a 15 amino acids spanning linker. Parts of the dimerization motif are visible in the structure and form an α -helix, which is proposed to form a coiled-coil structure. (MacPherson et al., 2006) The zoom in (Figure 4B) clearly shows the coordination of the metal ions by the six cysteines of the consensus binding region. Although the Zn-finger region is well conserved within the class III zinc cluster proteins and could easily be identified in Mrr1, the linker and dimerization motif are less conserved and could not be detected so far.

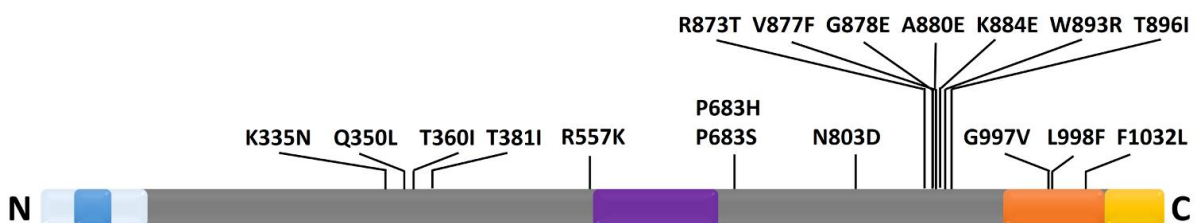


Figure 5 - Location of functional regions and gain of function mutations in Mrr1:

The Zn-finger domain (blue) is located at the N-terminus surrounded by a region of 106 amino acids (1 to 106, light blue) that are sufficient for the activation of the Mdr1 promoter. The middle homology region (MHR, amino acids 560 to 664, purple), an inhibitory domain (ID1: amino acids 951 to 1050, orange) and an activation domain (AD1: amino acids 1051 to 1108, yellow) are highlighted. Reported gain of function mutations and candidates from clinical isolates are also indicated. Adapted from (Schubert et al., 2011b)

Several gain of function mutations in Mrr1 play an important role for the overexpression of the Mdr1 pump in fluconazole resistant clinical isolates. (Dunkel et al., 2008a; Morschhauser et al., 2007)

Introduction

Under drug selection pressure, these mutations seem to have a highly beneficial impact for *C. albicans* as most were described to be homozygous for both alleles. (Dunkel et al., 2008a) In contrast, *C. albicans* strains, containing gain of function mutations, suffered from fitness defects under non-selective conditions. (Sasse et al., 2012) A functional dissection of Mrr1, performed in the group of Joachim Morschhäuser, led to the identification of several regulatory regions in the protein and some of the gain of function mutations correlate with these regions. Schubert *et al.* found that a region of only 106 amino acids at the *N*-terminus of the protein, surrounding the DBD, is sufficient to bind to the promoter of *MDR1*. When this region was shortened to 90 *N*-terminal amino acids no binding to the *MDR1* promoter was observed but some fluconazole resistance was triggered, which indicated the involvement of a different target. (Schubert et al., 2011a) As already suggested in earlier publications (MacPherson et al., 2006), Zn cluster domains often possess an activation domain (AD) at the *C*-terminus. This could be verified and the 58 *C*-terminal amino acids were identified as the AD1 of Mrr1. (Schubert et al., 2011b) In addition to AD1, an inhibitory domain (ID1) was found to be located within the stretch of amino acids spanning residues 951 to 1050. The deletion of 50 amino acids within ID1, either 951 to 1000 or 1001 to 1050, resulted in constitutive activity of the transcription factor. The deletion of AD1 and ID1, resulting in a protein consisting of amino acids 1 to 948, led to a transcription factor which was still inducible by benomyl or H₂O₂ and triggered subsequent Mdr1 expression. This hints to an additional activation domain located somewhere in the first 948 amino acids of the protein. (Schubert et al., 2011b) The middle homology domain (MHD), amino acids 560 to 664, is proposed to play a role in the activity control of the transcription factor. (Sanglard et al., 2009) The identified functional domains are shown in **Figure 5**.

18 gain of function mutations (**Figure 5**) were identified in clinical isolates with enhanced Mdr1 expression, and were noted to be of relevance. (Dunkel et al., 2008a; Eddouzi et al., 2013a; Morio et al., 2013; Morschhäuser et al., 2007) The gain of function mutation character of 15 of these mutations was experimentally confirmed. (Dunkel et al., 2008a; Eddouzi et al., 2013b) The mutations R557K, T896I and F1032L were included in the list of the 18 mutations due to their prevalence in at least three drug resistant clinical isolates, but so far, the experimental prove for being gain of function mutations is lacking. (Morio et al., 2013) Only a small number of these mutations are located within the identified regulatory domains. Specifically G997V, L998F and F1032L are located within ID1. These three point mutations clearly have a similar effect compared to the deletion mutants of the dissection study as they lead to constitutively active Mrr1. Interestingly, neither AD1 nor MHD contain any gain of function mutations. Deletion studies, where sequentially 50 amino acids of the protein were deleted, showed that most of the resulting proteins were inactive, likely caused by structural disintegration and protein instability. (Schubert et al., 2011b)

Introduction

In addition to the functional dissection experiments, studies were performed under inducing and hyperactive conditions that investigated the interplay between different transcription factors of Mdr1 which are summarized below. (Alarco and Raymond, 1999; Dunkel et al., 2008b; Heilmann et al., 2010; Rognon et al., 2006; Znaidi et al., 2009; Znaidi et al., 2008) In addition to Mrr1, Upc2, Mcm1 and Cap1 bind to the promoter region of Mdr1. Upc2 was found to be of no importance for the benomyl- or H₂O₂-induced activation of Mdr1; however, homozygosity of observed gain of function mutations in *UPC2* led to some *MDR1* expression. This level of *MDR1* expression does not contribute to fluconazole resistance whereas the three above described gain of function mutants of *UPC2* cause ERG11 overexpression and thereby contribute to fluconazole resistance. (Dunkel et al., 2008b; Heilmann et al., 2010; Hoot et al., 2011; Schubert et al., 2011a)

Cap1 is essential for oxidative stress response expression of Mdr1 and in this process Mrr1 is also involved. This type of Mdr1 overexpression might require an interplay of the two transcription factors. Cap1 also contributes to benomyl induced Mdr1 expression but is far less important than Mrr1. Hyperactive Mrr1 induced Mdr1 expression does not require Cap1 but Mcm1. (Mogavero et al., 2011) Gain of function mutations from clinical isolates were so far only reported in Mrr1 and have not been found in either of the two other transcription factors.

D	C	S	G	H	D
A	C	C	G	A	A
T		G		T	T
G				C	G

Figure 6 - Illustration of the DNA-binding motif of Mrr1:

D represents the bases A, T or G, *S* stands for the bases C or G and *H* represents the bases A, T or G. *C* or *G* stand for the respective base. Adapted from: (Schubert et al., 2011a)

As alluded to above, Mrr1 does not only bind to the promoter of Mdr1 but activates other unidentified genes. (Morschhauser et al., 2007) Some of the activated genes might be involved in the oxidative stress response of *C. albicans* induced by fluconazole. (Hoehamer et al., 2009) Additional open reading frames, in close proximity to Mrr1 binding sites have been identified and were partially assigned to a protein function. So far it remains unclear, which of all the identified candidates contribute to Mrr1-mediated fluconazole resistance. (Schubert et al., 2011a)

From other Zn cluster proteins, such as Gal4, it is known that they bind a repeating, sometimes reverse, CGG motif in the DNA and the so-called **DCSGHD** motif was identified **Figure 6**. (MacPherson et al., 2006; Schubert et al., 2011a) In this motif **D** stands for the bases A, T or G, **S** designates the bases C or G and **H** represents the bases A, T or C. It thus contains the important CGG motif. In the promoter of Mdr1 more than one region was identified that could bind to Mrr1, which supports the idea that no distinct region in the Mdr1 promoter is indispensable for constitutive overexpression of the pump. (Schubert et al., 2011a)

1.3 *Mycobacterium tuberculosis* – FadA5 involved in a new druggable pathway?

1.3.1 *Mycobacterium tuberculosis* and the disease

1.3.1.1 Tuberculosis

More than 60 years ago Selman A. Waksman received the Nobel Prize for the detection and isolation of the first antitubercular drug, streptomycin. (Woodruff, 2014) The great age of antibacterial discovery had just started and many people held tight on to the dream that tuberculosis (TB), also termed the 'Great White Plague' would be quickly overcome. Today we know that the so-called 'discovery void' of antibiotics, in combination with an increasing number of resistances, lead to *Mycobacterium tuberculosis* being a greater threat than ever. (Silver, 2011)

Therefore, the WHO is increasing and improving its surveillance of resistant *M. tuberculosis* strains and is following strict programs, such as the DOTS (directly observed treatment, short-course) or the Stop TB strategy. These are necessary measures as in the year 2012 8.6 million new TB cases and 1.3 million fatalities were reported. Although these numbers sound high, the WHO also reports a positive treatment success rate of 87% among new TB cases (2011). Interestingly, according to a distribution of the success rates by continents, Europe has the lowest curing quote with only 72%. In general TB is prevalent in Sub-Saharan Africa or Asia but the percentage of infections in countries of the former Soviet Union is constantly rising. (WHO, 2013)

The difficulties in treatment in Eastern Europe or Central Asia are not only caused by the drug-susceptible TB but also the increasing number of patients infected with multidrug (MDR) or extensively drug resistant (XDR) TB strains. It is estimated that 3.6% of all new infections, 20% of recurring infections and 13% of all fatalities are caused by MDR or XDR *M. tuberculosis* strains. A large problem is the low success rate (48%) of treating these infections which is most likely caused by the extensive and expensive treatment procedure that can last up to two years. (Gandhi et al., 2010; WHO, 2013)

In addition to these numbers, the WHO estimates that 1/3 of the world's population is currently infected with non-active but latent TB. Approximately 5 to 10% of people infected with latent TB can suffer from a break-out of the active disease, caused by impaired living conditions, aging, or in combination with other diseases such as HIV and diabetes-type-2. (WHO, 2013) Co-infections, for example with HIV, or suffering from other parallel diseases, such as diabetes, constitute a severe health threat. The combination HIV/TB is called a lethal combination, as one fourth of HIV-positive patients die by tuberculosis infections and the treatment with anti-TB and HIV drugs is often complicated. Treatment complications and therapeutic response problems also often occur when people are suffering from TB and diabetes. (Koul et al., 2011)

Introduction

Tuberculosis infections mostly affect the lungs (pulmonary TB) although they can also affect the brain, joints, bones or kidneys. Symptoms of tuberculosis are fever, coughing (with bloody sputum), loss of weight, and loss of appetite. (Ahmad, 2011; Kayser et al., 2001) The common diagnosis tools to identify the infection in time have hardly changed within the last decades. Normally TB is diagnosed by sputum smear investigations or, if available, by culturing methods supported by radiographs of the lung. Just recently, the Xpert® MTB/RIF test was released, which should enable a more rapid molecular determination of whether a drug-susceptible or a rifampicin resistant TB infection is prevalent. (WHO, 2014) Rifampicin is the most potent antitubercular drug (**Table 1**). Latent TB is difficult to detect; a tuberculin skin test within six to eight weeks after exposure to the bacteria or an interferon-gamma (IFN- γ) release assay can be applied. The tuberculin skin test lacks sensitivity and specificity whereas the IFN- γ test is more sensitive and specific. (Ahmad, 2011; Kayser et al., 2001)

1.3.1.2 *Mycobacterium tuberculosis*

Mycobacterium tuberculosis, the causative agent of tuberculosis, is a 2 to 4 μm long, non-motile, aerobic, gram-positive bacterium (**Figure 7**) that causes airborne infections with the above-described symptoms. (Kayser et al., 2001) Humans have suffered from tuberculosis infections for at least 10,000 to 11,000 years. This was proven by molecular and morphological methods investigating human bone samples from one of the first detected villages, located in the eastern Mediterranean. *M. tuberculosis* could be clearly identified with high genetic similarities between the ancient bacteria and the recent ones, which also underlines a long-term co-existence of humans and *M. tuberculosis*. (Hershkovitz et al., 2008) These bacteria belong to the order of *Actinomycetes* and have very long doubling times (20 h) in contrast to e.g. *M. smegmatis* with doubling times between 1 and 4 hours. They are therefore defined as slow growing mycobacteria. (Parish and Brown, 2008) Mycobacteria contain a genome with a high content of guanine and cytosine bases (GC), representing 65.6% of all DNA bases in *M. tuberculosis*. This fact often causes problems in recombinant protein expression of mycobacterial genes, for example *Escherichia coli* has an average GC content of only 50%. *E. coli* is often not able to recombinantly express soluble protein of mycobacterial origin. (Parish and Brown,



Figure 7 - *Mycobacterium tuberculosis*: This illustration depicts a three-dimensional (3D) computer-generated image of a cluster of rod-shaped drug-resistant *Mycobacterium tuberculosis* bacteria. The artistic recreation was based upon scanning electron micrographic imagery. Kindly provided by CDC (Brower, 2013)

Introduction

2008) Therefore *M. smegmatis*, a biosafety level II organism, is an established recombinant protein expression system for mycobacterial proteins. (Parish and Stoker, 1998)

Genomic studies of the *M. tuberculosis* genome revealed that many genes are assigned to lipid metabolic pathways. (Cole et al., 1998) One important aspect resulting from this genomic diversity is the very specific cell wall composition of *M. tuberculosis*, where long branched fatty acids, known as mycolic acids, are contributing to the complexity and impermeability of the cell wall. The cell



Figure 8 - *M. tuberculosis* colonies:

The rough, wax like colonies reveal the special morphology of M. tuberculosis. Kindly provided by CDC (Kubica, 1976)

skeleton is extremely lipid rich and hydrophobic and is often described as a wax-like 'skin' (Figure 8). The mycolic acids, together with other lipids, form the outermost layer of the cell wall and are interconnected and associated to the cell membrane by a diversity of sugar molecules or chains, such as arabinofuranoses, galactofuranoses or peptidoglycans. In addition, other lipids or proteins are part of the cell wall. (Brennan, 2003; Cole et al., 1998; Tonge, 2000)

Besides the fact that the lipids are required to establish this very complex cell wall, a special significance of the variety of lipid encoding genes is also given by the special environment of the bacteria in their human host.

1.3.1.3 Infection, granuloma formation and persistence

Normally infection with *M. tuberculosis* occurs by the inhalation of bacteria containing droplet nuclei exhaled by people suffering from active TB. Due to the small size of the nuclei, the bronchial defenses are avoided and the bacteria advance directly to the alveoli. (Ahmad, 2011)

It has been described, that mycobacteria are not only incorporated into alveolar macrophages but also into M cells, alveolar endothelial cells, pneumocytes or dendritic cells. The bacteria can survive and replicate within the cells and thereby can be disseminated by migrating cells. After two to eight weeks post-infection, an effective cell mediated immune response can be observed. This immune response normally inhibits replication and hinders further spread. The immune processes which are best studied and likely of the highest importance take place in the macrophages. During phagocytosis, the macrophages incorporate the mycobacteria in their phagosomes, which then fuse with lysosomes and become phagolysosomes to subsequently mature and destroy the bacteria. *M. tuberculosis* is able to prevent the maturation of the phagolysosomes. It is not clear if the inhibition of the phagolysosomes' maturation is essential for the bacteria to survive. If maturation occurs, the

Introduction

bacteria are still able to survive the hostile acidic environment that contains reactive oxygen and nitrogen species, because of their acid-fast cell walls and additional response mechanisms, such as the lysination of the cell wall. Additional factors of the human immune system, such as chemokines, inactivated monocytes or neutrophils that are summoned to the site of infection, are unable to kill the bacteria. Given that the immune system is not able to kill the bacteria, a capsulation process is initiated, where the infected macrophages are surrounded by several layers of healthy macrophages, lymphocytes, T-cells and other factors, which form the tubercular granuloma. Granuloma and the resulting fibrous or calcified lesions can be recognized on radiographs. (Smith, 2003; Tischler and McKinney, 2010; Ulrichs and Kaufmann, 2006) Although most of the bacteria are killed within this environment, some are able to survive over decades and can be reactivated in case of immunodeficiency. (Ahmad, 2011) It is still under debate how and under which conditions the bacteria survive. It is unknown if the bacteria are in a stagnating metabolic state, in which they do not die but also not replicate. This process could be latent or, if the same number of bacteria are dying and replicating and thereby forming units, chronic. (Smith, 2003) The latter scenario currently seems to be the more probable one, as complete gene clusters were found to be up-regulated during this so called latency. (Peyron et al., 2008) Latent TB might be more comparable to the chronic stage of the tuberculosis infection. It is known during this stage that the bacteria can change their metabolism from an aerobic carbohydrate dependent fashion to an anaerobic, lipid utilizing kind. (Barry et al., 2009) During chronic stages of infection, the bacteria are replicating only once in several days. It is both striking and horrifying that *M. tuberculosis* is not only able to survive in the center of the human immune response but may even take advantage of it. (Wipperman et al., 2014).

The above-mentioned lipid utilization leads back to the high number of lipid metabolizing enzymes in the *M. tuberculosis* genome, which might also have an important function not only with respect to the cell wall but also for the utilization of host lipids during these very long phases of persistence in the human body. (Barry et al., 2009) In this context, Cholesterol has been identified as an important nutrient for the bacteria during chronic stages of infection and will be further examined later. (Barry et al., 2009; Miner et al., 2009; Van der Geize et al., 2007)

1.3.2 Treatment of tuberculosis, anti-TB drugs and resistance

The treatment of tuberculosis is a complicated and long lasting task that has often failed in the past. *M. tuberculosis* is a difficult organism to treat with drugs as it comprises intrinsic barriers or resistance mechanisms that already exclude drugs or hinder therapy. (Sacchetti et al., 2008) *M. tuberculosis* naturally expresses β -lactamases, e.g. BlaC, which eliminates the ability to treat this organism with a common antibiotic, such as penicillin. However, clavulanate, a β -lactamase inhibitor,

Introduction

might enable the use of other β -lactams, such as amoxicillin or meropenem, in combination-therapy. (Poce et al., 2014) The low cell wall permeability for drugs or prodrugs is also a large problem in antibiotic therapy, as many compounds are not transferred through this envelope. Recently, the impact of not further defined efflux mechanisms was reported as a significant factor in drug resistance processes against tetracyclines, fluoroquinolones or aminoglycosides. (da Silva et al., 2011) Last, but not least, the persistence or latency ability of *M. tuberculosis* often hinders a complete clearance of the bacterial burden in the human body. (Balganesh et al., 2008; da Silva et al., 2011; Poce et al., 2014; Sacchetti et al., 2008)

Table 1 - First- and second-line anti TB drugs

	<i>name or compound class</i>	<i>discovered</i>	<i>target</i>	<i>inhibition</i>
first-line	Isoniazid (prodrug)	1952	<i>katG/inhA</i> ; catalase/enoyl reductase	mycolic acid synthesis
	Rifampicin	1966	<i>rpoB</i> ; β subunit of RNA polymerase	protein synthesis
	Ethambutol	1961	<i>embB</i> ; arabinosyl transferase	arabinogalactan synthesis
	Pyrazinamide (prodrug)	1952	<i>pncA</i> ; nicotinamidase/parazinamidase	depletion membrane energy
second-line	Fluoroquinolones	1963	<i>gyrA/B</i> ; DNA gyrase subunits A/B	DNA gyrase activity
	Ethionamide (prodrug)	1956	<i>etaA/ethA</i> ; flavin monooxygenase	mycolic acid synthesis
	<i>p</i> -amino salicylic acid	1948	<i>thyA</i> ; thymidylate synthase A	NF-kappa B (?)
	Capreomycin / Viomycin	1960	rRNA methyltransferase	protein synthesis
	Kanamycin / Amikacin	1957	16S rRNA	protein synthesis
	Streptomycin	1944	several; interfering with ribosome or rRNA	protein synthesis

(Almeida Da Silva and Palomino, 2011; da Silva et al., 2011; Poce et al., 2014; Zhang and Yew, 2009)

The first drugs administered against tuberculosis were discovered and isolated in the 1940's and were followed by a variety of structurally diverse drugs that were discovered in the following 20 years. (Almeida Da Silva and Palomino, 2011; Zhang and Yew, 2009) Although many of these medications were not established and approved for use against TB for years, all currently used drugs for the treatment of TB are more than 40 years old. (Poce et al., 2014) The common antitubercular medicines are classified as being either first-line or second-line drugs, with the first-line drugs being more effective and having less adverse side effects than the second line drugs (**Table 1**). A special role in the class of the first-line drugs is given to rifampicin and isoniazid. Drug-susceptible tuberculosis is treated over a period of six months and the therapy includes three to four of the first-line drugs applied during the first two months, followed by another four months of therapy with isoniazid and rifampicin. The treatment of MDR or XDR tuberculosis is much more complex. MDR strains are classified as being resistant against rifampicin and isoniazid and XDR strains are not only resistant against rifampicin, isoniazid, but also against the fluoroquinolones, and one of the following

Introduction

compounds: capromycin, kanamycin or amikacin. (Dorman and Chaisson, 2007) These strains require a more complex treatment over a longer period of time. This treatment can last up to two years and consists of a combination of four to five drugs sub-classified in five groups and chosen on the basis of efficacy, safety and costs. (Caminero et al., 2010)

Most of the listed anti TB drugs have one common drawback: the rapid development of drug resistance. Smith *et al.*, Balganesh *et al.* and Sacchetti *et al.* reviewed the molecular resistance mechanisms and causes for the appearance of drug resistance in *M. tuberculosis*. (Balganesh et al., 2008; Sacchetti et al., 2008; Smith et al., 2013) The predominant drug resistances were acquired following failed drug therapies. In the beginning of antibacterial TB therapy, mono-therapy was applied, which triggered fast emerging resistant strains. When this problem was identified, the therapies were adapted to the above-mentioned long lasting combination therapies. As these medications are expensive and patients either suffer from severe side effects or experience a regain of health before the end of the six-month treatment, therapy often was aborted without complete clearance of the bacteria. Aborted therapies in combination with mono-therapies are the main reasons for the emergence of MDR and XDR *M. tuberculosis* strains. Resistances in *M. tuberculosis* are caused by spontaneous mutations that occur with rather low frequencies of 10^{-6} to 10^{-8} per replication cycle. (Zhang and Yew, 2009) Mutations predominantly affect the drug targets or, in case of prodrugs, the activating enzymes. Other, more rarely observed, mechanisms of resistance in *M. tuberculosis* are chemical modifications of the applied drugs, enzymatic drug degradation or (increased) efflux of the antibiotics. Interestingly, horizontal gene transfer is a hardly-observed phenomenon in *M. tuberculosis*; therefore, strains acquire their resistance independently during failed drug therapy. (da Silva et al., 2011) Last, but not least, persistence under-down regulation of metabolic processes or change of the processes is also an opportunity to overcome drug pressure. (da Silva et al., 2011; Smith et al., 2013)

To counter the prevalent resistance problems of tuberculosis treatment, industry and academia were busy towards identifying new drug scaffolds and developing new drugs. These efforts now begin to yield successes, as some new drugs are currently in clinical phase trials or even have been approved for TB therapy. (Poce et al., 2014) Bedaquiline (trade name Sirturo) was the first drug that was specifically designed to target *M. tuberculosis* to be released in the last 40 years. Sirturo inhibits the ATP-synthase, which also represents a new point of action for TB antibiotics (**Table 2**). Sirturo has some severe side effects but was nevertheless approved for use in MDR or XDR therapy, similar to the nitroimidazole OPC-6783 (**Table 2**), a mycolic acid synthesis inhibitor. (Leibert et al., 2014) Importantly, both medications may also be able to target persisting mycobacteria and therefore also

Introduction

latent TB. Further drugs that are currently being investigated, or that are already used in MDR or XDR treatments, are listed in **Table 2**.

Table 2 - Selection of compounds in clinical trials or approved for MDR TB therapy

<i>compound class</i>	<i>Trade name</i>	<i>clinical trials / MDR</i>	<i>targets</i>	<i>latent TB?</i>
Bedaquiline	Sirturo* (TMC207)	phase III	inhibits ATP-synthase	maybe
Nitroimidazoles	PA-824 OPC-6783*	phase II phase III	active against resting and replicating bacteria	maybe
Ethylenediamines	SQ109	phase II	inhibition of Mmp13 (membrane integrity)	-
Benzothiazinones	BTZ043 PBTZ169	- -	inhibits cell wall arabinan synthesis	-
Fluoroquinolones	Levofloxacin* Moxifloxacin* Ofloxacin*	applied against MDR TB	<i>gyrA/B</i> ; DNA gyrase subunits A/B	-

* approved for conditional use in MDR therapy

(Poce et al., 2014; Stover et al., 2000; Villemagne et al., 2012)

The development of vaccines is a constantly researched topic and currently there are ten vaccines for preventive immunization and two additional vaccines for immune therapies are in the pipeline. (WHO, 2013) A recent study also investigated the impact of a combined parenteral and respiratory application of vaccines to enhance their potency and has demonstrated promising results. (Beverley et al., 2014)

Despite all the above-mentioned efforts and successes, the search for improved or new antibiotics must continue in order to reduce side effects during normal or MDR drug therapy and to target newly developing MDR or XDR strains as well as persisting mycobacteria. Most of the above-mentioned therapeutics target enzymes or processes that are required in phases of frequent cell replication, but very few agents are available that target bacteria in the chronic phase of infections, where they are replicating rarely or very slowly. Bacteria with low replication frequency might represent an interesting platform to study new drug targets. (Wipperman et al., 2014)

1.3.3 Approaching new ways – The cholesterol metabolism pathway

Several recent studies revealed the cholesterol metabolism pathway of *M. tuberculosis* to be a potential and so far unexploited drug target. (Miner et al., 2009; Ouellet et al., 2011) The studies also revealed that although cholesterol is dispensable during the infectious phase of tuberculosis, it is required during the chronic phase of infection. (Nesbitt et al., 2010; Pandey and Sasseti, 2008) Mycobacteria are not able to produce cholesterol and it is therefore necessary for the organism to import the compound. (Mohn et al., 2008) Interestingly, the lipid content of infected macrophages is

Introduction

clearly elevated in comparison to non-infected tissue and is predominantly comprised of phospholipids, triacylglycerides, cholesterol or cholesterol esters which are readily available for import into the bacteria. (Kim et al., 2010) As *M. tuberculosis* switches to a lipid based anaerobic metabolism upon entry into the chronic phase of infection, enzymes involved in lipid import and utilization might provide a new and promising drug scaffold.

To specifically target enzymes which utilize host cholesterol and to elucidate their potential as essential target, the mycobacterial genome had to be dissected. Enzyme studies of related families within the actinomycetae, such as *Rhodococcus sp.*, and the decoding of the *M. tuberculosis* genome significantly contributed to the identification of the genes involved in these processes. (Cole et al., 1998; Van der Geize et al., 2007) 250 out of 4,000 genes were annotated to be involved in fatty acid or lipid metabolism. The transcriptomal analysis of *M. tuberculosis* grown with or without cholesterol revealed 200 upregulated genes that were not completely identical to the above mentioned 250 genes. (Nesbitt et al., 2010) In addition, the Cho-region was defined, which includes homologs of 83 genes that were defined earlier as being involved in β -oxidation and biphenyl compound degradation in *rhodococci*. (Nesbitt et al., 2010; Van der Geize et al., 2007) The Cho-region is under control of KstR1 and KstR2, which belong to the Tet-repressor family. KstR1 controls the majority (74) of the Cho-region genes, which are associated with the degradation of the cholesterol rings A and B (**Figure 9**) as well as the side-chain. (Kendall et al., 2007) KstR2 is less well studied and it was suggested that it mainly controls genes involved in the catabolism of the cholesterol rings C and D. (Wipperman et al., 2014)

Cholesterol degradation can be subdivided into three processes that partially take place in parallel. These are the degradation of the rings A and B, and of the rings C and D as well as the side-chain degradation. After the initial oxidation and isomerization step, side-chain degradation and catabolism of the A and B rings are facilitated in parallel but follow different reaction kinetics. (Wipperman et al., 2014)

As mentioned above, very little is known regarding the degradation of the C and D rings, as genes identified in *rhodococci* are not conserved in *M. tuberculosis*. Even the extent of degradation of the two rings is unknown, but it is very likely that the participating enzymes require completely degraded side-chains to function correctly. Presumably, the two rings are degraded in an oxidative manner by the release of propionyl-CoA and acetyl-CoA as important metabolites. (Casabon et al., 2013; Wipperman et al., 2014)

Introduction

All enzymes described in the following passage are the identified and assigned *M. tuberculosis* proteins involved in the cholesterol degradation process. After the initial oxidation of the hydroxyl in position 3 (Figure 9) followed by an isomerization, both performed by the enzyme 3 β -Hsd, either degradation of the A and B rings or of the side-chain can progress. Ring degradation is initiated by the introduction of a double bond catalyzed by KstD. Subsequently, ring B is hydroxylated by KshA/B, which leads to the aromatization of ring A and ring opening of B. Hydroxylation by HsaA/B and the following catalysis by the deoxygenase HsaC results in the opening of ring A. 4,9-DSHA is then hydrolyzed by HsaD, which releases the precursors of pyruvate and propionate and a compound consisting of the rings C and D, where C is substituted with propionic acid (DOHNAA, Figure 9). The fate of these compounds is unknown but ligation with CoA by the enzyme FadD3 is likely. Some of the participating enzymes seem to be important for the bacteria because deletions lead to attenuated growth phenotypes of *M. tuberculosis* in infection models with macrophages or mice. (Wipperman et al., 2014)

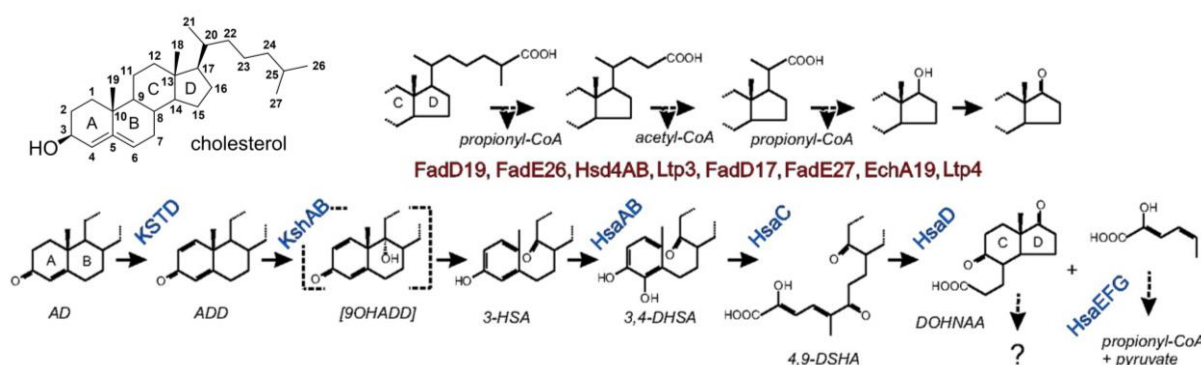


Figure 9 - Cholesterol degradation pathway of *Rhodococcus sp. RHA1*, *M. tuberculosis H37Rv*, and *M. bovis bacillus Calmette Guerin*.

The reaction steps of the ring opening and the side-chain degradation are shown. The dashed arrows indicate multiple enzymatic steps. 9OHADD is non-enzymatically hydrolyzed to yield 3-HSA (brackets). The common numbering of the cholesterol carbons is shown in the cholesterol structure on the upper left. AD, 4-androsten-3,17-dione; ADD, 1,4-androstadiene-3,17-dione; 9OHADD, 9-hydroxy-1,4-androstadiene-3,17-dione; 3-HSA, 3-hydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione; 3,4-DHSA, 3,4-dihydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione; 4,9-DSHA, 4,5-9,10-diseco-3-hydroxy-5,9,17-triox androsta-1(10),2-diene-4-carboxic acid; DOHNAA, 9,17-dioxo-1,2,3,4,10,19-hexanorandrostan-5-oic acid. Adapted from (Cowen, 2008; Van der Geize et al., 2007)

The side-chain degradation is initiated by the cytochrome cyp125, which oxidizes one terminal methyl group. After this initial step, the oxidized steroid is ligated to CoA by the enzyme FadD19. At this point, theoretically, three repeating steps of β -oxidation will be performed by an acyl-CoA dehydrogenase: an enoyl-CoA hydratase, a 3-hydroxy-acyl-CoA-dehydrogenase and a 3-ketoacyl-CoA thiolase. Due to the different chain lengths, the distinct steps are performed by different enzymes, which could not be assigned consistently. Enzymes of the FadE family catalyze the first step of the β -oxidation reaction, with possibly either FadE28 or FadE29 performing the final dehydrogenase step during degradation (Figure 10, compound 3 to AD). The enoyl-CoA hydratases EchA were identified

Introduction

to catalyze the second step and only EchA19 was so far identified to be involved in this reaction. For the third step, namely the formation of the β -keto compound, no distinct enzymes of the FadB family could be assigned. The final step of β -oxidation, called thiolysis, is performed by the FadA enzymes, for which FadA5 was identified. Deletion of this enzyme leads to an attenuated growth phenotype in the mouse model of infection. However, recent studies imply a potential mechanism to circumvent the last two steps of β -oxidation by employing an aldolase, an aldehydehydrogenase and a ligase. This process was described in *pseudomonas* bacteria but homologous enzymes in *M. tuberculosis* yet need to be identified. (Wipperman et al., 2014)

Finally, the question has to be asked, what mycobacteria actually gain by metabolizing cholesterol. One obvious answer is the production of valuable metabolites such as acetyl and propionyl-CoA. These metabolites are important as precursors of succinic acid or pyruvate or they are converted into methyl-branched lipids such as phthiocerol dimycocerosates (PDIM) or sulfolipids both of which are important for mycobacterial virulence. (Wipperman et al., 2014). The metabolites are also fed into the TCA cycle for energy purposes. In addition, the intermediate 4-androsten-3,17-dione (AD) is a prohormone, which might be important in the host immune response.

1.3.4 The β -ketoacyl-CoA thiolase FadA5

1.3.4.1 Role in *M. tuberculosis* infection

In vivo and *in vitro* experiments with *M. tuberculosis* strains, grown in mouse lungs or human macrophages, showed that the gene expression of the *fadA5* gene was upregulated. (Dubnau et al., 2005; Dubnau et al., 2002) In addition, *fadA5* belongs to the genes in the Cho-region, which are upregulated by the presence of cholesterol and suppressed by the KstR1 regulator. (Nesbitt et al., 2010)

These results triggered an in-depth investigation of the thiolase FadA5. The following studies, performed by Nesbitt *et al.*, validated that FadA5 functions as a thiolytic enzyme. The conversion of acetoacetyl-CoA and CoA into acetyl-CoA, although at low rates, confirmed that FadA5 is functioning as a β -ketoacyl-CoA thiolase. The low turnover rates showed that acetoacetyl-CoA is likely not the natural substrate of FadA5. (Nesbitt et al., 2010)

Growth of *M. tuberculosis* on cholesterol as a sole carbon source showed that a Δ *fadA5* variant was not able to grow, whereas WT bacteria and a *fadA5*-complemented strain showed comparable growth patterns. This, together with a mouse model of infection, where a *fadA5* WT, Δ *fadA5* variant and a *fadA5*-complemented *M. tuberculosis* strain were used, revealed an attenuated disease

Introduction

phenotype when the *fadA5* gene was missing. This phenotype was not apparent until week five post infection; up to this point the bacterial burden in the mouse lungs was comparable in all three strains. In contrast, at week eight post infection, the Δ *fadA5* strain showed a 10-fold reduction in colony forming units in comparison to the two other strains. Previous studies have already shown that the immune system reacts by the latest at week eight post-infection. (Ahmad, 2011). As soon as the infection enters the chronic or persistence phase, the FadA5 enzyme is required for proper growth of *M. tuberculosis* in the host. (Nesbitt et al., 2010) This confirms the theory that cholesterol metabolism is negligible during the initial phase of infection and later becomes crucial in the chronic phase of infection. (Ouellet et al., 2011) Thus, *fadA5* is important for *M. tuberculosis* survival *in vivo* and may be a potential drug target. (Nesbitt et al., 2010)

Since FadA5 plays an important role in cholesterol metabolism, the location of action became of interest. Metabolic labeling with ^{14}C revealed the presence of the metabolites androsten-4-en-3,17-dione (AD) or androsta-1,4-dien-3,17-dione (ADD) (Figure 9 and Figure 10) in the culture supernatant of the *fadA5* WT strain, which were missing in the *fadA5* deficient strain. These metabolites suggest a role of FadA5 in the side-chain degradation of cholesterol, although van der Geize *et al.* proposed a degradative activity in the DOHNAA degradation. (Nesbitt et al., 2010; Van der Geize et al., 2007) Following kinetic studies showed that FadA5 has a 400-fold increased specificity for the 3,22-dioxo-chole-4-ene-24-oyl-CoA (Figure 10, compound 2) as a substrate compared to acetoacetyl-CoA. This underlines FadA5's role in cholesterol side-chain degradation. (Schaefer et al.)

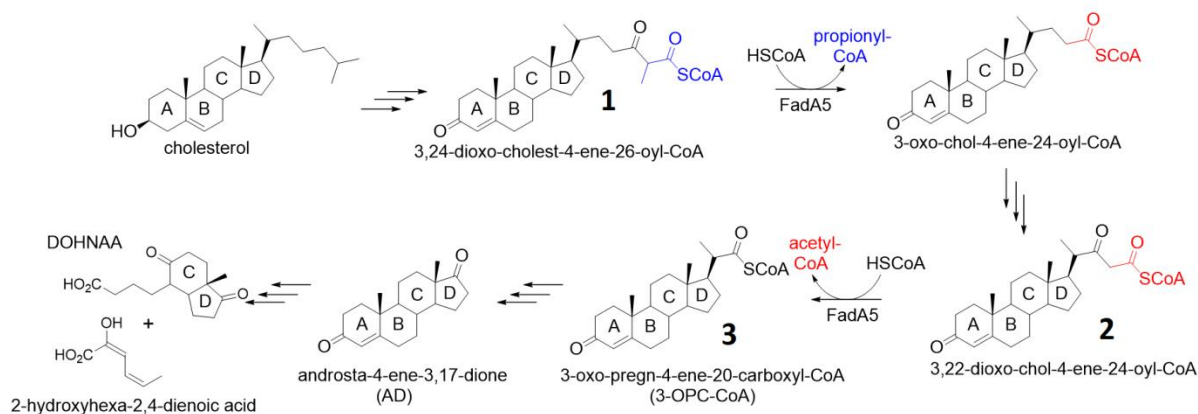


Figure 10 - Proposed role of FadA5 in the side-chain degradation of cholesterol:

FadA5 catalyzes the C-C cleavage step of the β -oxidation and therefore the cleavage of β -keto-thioesters into acetyl- (red) or propionyl- (blue) CoA and the steroid-CoA product. (Schaefer et al.)

Interestingly, a deletion of *fadA5* and the growth of the deficient *M. tuberculosis* strain on cholesterol did not lead to the production of toxic intermediates as observed for *hsaC* or *igr*-operon deficient strains. (Chang et al., 2009; Yam et al., 2009)

Introduction

The question remains, however, how *M. tuberculosis* benefits from cholesterol metabolism. As explained above, acetyl-CoA and propionyl-CoA are important metabolites but the study of Nesbitt *et al.* also presents AD or ADD as important compounds based on androsterone. Both metabolites were secreted into the medium during growth of *M. tuberculosis* and, as they are prohormones, they may play a role in the host-bacterium-interplay especially after the host immune response was initiated. (Nesbitt *et al.*, 2010)

1.3.4.2 Thiolase activity and reaction of FadA5

The catalytic moiety of thiolases consists of a cysteine residue that acts as nucleophile, and a general acid and base, histidine and another cysteine. (Nesbitt *et al.*, 2010) Degradation is the thermodynamically favored catalytic direction of thiolases, but the performed C-C cleavage is reversible and therefore thiolases for both biochemical directions exist. (Thompson *et al.*, 1989) The most detailed structural and functional analysis of a thiolase was performed for a biosynthetic thiolase from the bacterium *Zoogloea ramigera*. (Kursula *et al.*, 2002; Merilainen *et al.*, 2008; Modis and Wierenga, 1999, 2000) The results obtained for the biosynthetic pathway can also be transferred to the degradative enzymes and enable initial suggestions for the catalytic steps of FadA5.

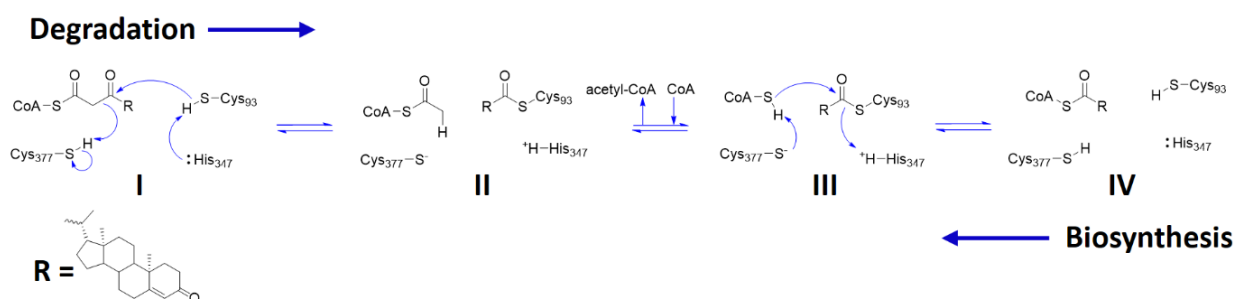


Figure 11 - Proposed thiolase reaction mechanisms of FadA5:

The large blue arrows indicate the two reaction directions of the thiolases, the degradation is described from left to right and the biosynthesis from right to left. The roman numerals stand for the enzyme states during thiolysis and the small blue arrows depict the electron transfer during the degradative thiolase reaction. (Schaefer *et al.*)

In the degradative pathway, a β -ketoacyl-CoA is cleaved into acetyl-CoA and a two-carbon shortened acyl-CoA. (Gilbert, 1981; Gilbert *et al.*, 1981) In the case of FadA5, at least in the cholesterol metabolism pathway, this means the turnover of a β -keto-steroid-CoA into a shortened steroid-CoA under the release of acetyl-CoA (Figure 11). The initial step of the reaction is the nucleophilic addition of one catalytic cysteine at the β -keto-ester of the substrate. Subsequently an acylated enzyme intermediate is formed under the release of the first product acetyl-CoA (Figure 11, state I \rightarrow II). The acetyl-CoA is exchanged for a CoA (Figure 11, state II \rightarrow III), which nucleophilically attacks the acyl-cysteine intermediate that releases the free cysteine. In this step, the acyl-CoA product is

Introduction

formed (Figure 11, state III \rightarrow IV). To prove that the described reaction is catalyzed by FadA5, the turnover of the 3,22-dioxo-chol-4-ene-24-oyl-CoA (Figure 10, compound 2) into 3-oxo-pregn-4-ene-20-carboxyl-CoA (3-OPC-CoA, Figure 10, compound 3) under the release of acetyl-CoA was shown. (Schaefer et al.) Catalysis can be described as a ping-pong mechanism, as the first product (acetyl-CoA) has to be released before the binding of the second substrate (CoA). Prior to the confirmed catalysis, side-chain degradation requires the release of propionyl-CoA also by β -oxidation (Figure 10, compound 1). If this step is also performed by FadA5 is not clear but due to the normally observed broader substrate specificity of degradative enzymes it seems very likely that this step can also be catalyzed by FadA5.

1.3.4.3 What about human thiolase homologs?

As FadA5 might serve as a potential drug target to inhibit *M. tuberculosis* in the chronic stage of infection or even latent tuberculosis infections, homologies to human enzymes are very critical. In the human genome six different thiolase enzymes have been identified, with different isoforms. (Anbazhagan et al., 2014) The sequence identity between FadA5 and the human thiolases is between 24% and 37%. Three of the enzymes, namely the human CT-thiolase (CT) (Kursula et al., 2005), located in the cytosol, the T2-thiolase (T2) (Haapalainen et al., 2007) located in the mitochondria and the AB-thiolase (AB, PDB entry 2iik) located in the peroxisomes have been structurally characterized. The mitochondrial T1-thiolase (T1) (Gilbert et al., 1981), the human mitochondrial trifunctional enzyme thiolase (TFE) (Fould et al., 2010) and the peroxisomal SCP2-thiolase (SCP2) (Antononkov et al., 1997) were so far partially characterized towards their function. CT is assumed to be solely a biosynthetic enzyme whereas the T2 enzyme is able to perform the biosynthetic and the degradative reaction. Biosynthetic enzymes are described to possess high substrate specificity in comparison to the degradative enzymes. (Kursula et al., 2005) The other human thiolases were characterized to be solely degradative enzymes. (Anbazhagan et al., 2014) The human enzymes interact with various substrates, ranging from acetyl-CoA to long chain, branched and unbranched 3-keto-CoA compounds or, in the case of SCP2, even with bile acid synthesis intermediates; however, all thiolase substrates require a covalent bond to CoA. (Anbazhagan et al., 2014)

A recent phylogenetic analysis of the human enzymes in comparison to thiolases from *M. tuberculosis* and *M. smegmatis* was performed by aligning 130 thiolases from the afore-mentioned organisms and additional organisms to support the analysis confidence. This analysis resulted in 12 sequence classes with characteristic sequence fingerprints for the catalytic residues and loops. The complete results will not be discussed in this introduction, but valuable information was obtained concerning FadA5. FadA5 was identified to belong to the trifunctional enzyme like thiolases (TFEL)

and to be more specific to type 1. As the name already implies, these enzymes do not share a branch with any human enzyme but they are related to the human TFE. Based on this study, the T1, T2 and CT enzymes are the most different from FadA5 as they assume a completely different and secluded region in the phylogenetic tree. As described, human TFE thiolases are the closest relatives to FadA5, more distantly related are the AB enzymes but they are still more closely related to FadA5 than SCP2 thiolases, which again differ to a large extent. (Anbazhagan et al., 2014)

1.4 Research objective

The research objective for this study was divided into two independent parts, both of which were aimed to provide a basis for drug development.

The *C. albicans* protein Mrr1 is a transcription factor heavily involved in anti-fluconazole resistance of the fungus and several mutations are known in this transcription factor that trigger drug resistance. One major focus of this thesis was therefore to determine the structure of Mrr1 to support the microbiological and functional studies, in the organism of origin, conducted in the lab of Prof. Dr. Joachim Morschhäuser. For this purpose recombinantly expressed protein variants should be generated and analyzed towards their solubility and applicability for biochemical studies and crystallization experiments. To overcome solubility issues it was also attempted to express the protein in *P. pastoris* and different purification strategies were pursued. The overall goal was the successful structural characterization of the Mrr1 protein and one of the truncation variants, to compare the fold of the wild type protein with the most likely differently folded gain of function variant. This could provide important information with respect to the activation mechanism and thereby expression upregulation processes of the enzymes under control of this transcription factor, as the Mdr1 transporter.

The FadA5 protein might prove to be an interesting drug target concerning the chronic phase of *M. tuberculosis* infections. Biochemical studies were performed prior to and in parallel this thesis in the group of Prof. Dr. Nicole Sampson, a close collaborator on this project. Therefore, the main focus of this thesis was an initial structural characterization of FadA5 in its apo form and in complex with different CoA-ligands. Crystallization experiments were performed with FadA5 WT and an active site variant C93S. The structure of the apo enzyme and complex structures with acetyl-CoA, CoA and 3-oxo-pregn-4-ene-20-carboxyl-CoA (3-OPC-CoA, **Figure 10**, compound 3), a steroid product of the enzyme, should provide important insights along the reaction pathway and reveal interactions between the protein and the ligands as well as the location of the steroid-binding pocket. Combined the gained information should finally be used to evaluate the druggability of FadA5 in the context of human thiolases and if reasonable for initial lead compound design.

2 Materials

2.1 Chemicals

If not stated differently all used chemicals, in analytical grade or better, were purchased from Applichem (Darmstadt), Carl Roth (Karlsruhe), Fluka (Neu-Ulm), Sigma-Aldrich (Seelze) and Hampton Research (Laguna, USA). The chemicals used for protein crystallography were of the highest available purity.

2.2 Equipment

Table 3 - Technical Equipment

	<i>Specification</i>	<i>Company</i>
Autoclave	Systec V-150	Systec
Balances	XS 105 Dual Range	Mettler Toledo
	XS 6002S Dual Range	Mettler Toledo
Biological Safety Cabinet	Hera Safe KS 12	Kendro
Camera	AxioCam MRC	Zeiss
CD-Spectrophotometer	J-810	Jasco
Centrifuges	5415 D / 5415R / 5417R / 5810R	Eppendorf
	Heraeus Multifuge X3R Fiber Lite	Thermo Scientific
	F12-8x50C	
	Biofuge 13R No. 3757	Heraeus Kendro
	Avanti J-26S XP	Beckmann Coulter
	Avanti J-HC	Beckmann Coulter
Columns for affinity chromatography (Batch)	Econo-Column 1.5 x 20 cm	Bio-Rad
	Econo-Column 0.7 x 20 cm	Bio-Rad
Columns for FPLC Systems	HisTrap FF crude (5 ml)	
	MonoQ 5/50 GL	
	MonoQ 10/100 GL	
	HiLoad 16/60 Superdex 200 pg	GE healthcare
	HiLoad 26/60 Superdex 200 pg	
	Superdex 200 10/300 GL	
	Superdex 200 Increase 10/300 GL	
	Ni-MAC (1 ml)	Novagen
Cryo loop	Crystal Cap™ hT (Spine) (different sizes)	Hampton Research
Crystallization robot	Honey Bee 963	Zinsser Analytic
DLS (Dynamic light scattering)	DynaPro Dynamic Light Scattering	Wyatt Technology Corporation
Fast protein liquid chromatography systems (FPLC)	ÄKTA avant 25	GE healthcare
	ÄKTApurifier 10	
	ÄKTA pure M1	
Electrophoresis	Mini-Protean Tetra Cell	Bio-Rad
	Mini-Sub Cell gT	
	Sub Cell gT	
	PowerPac hC	
	PowerPac Basi	
Electroporation system	GenePulser MX Cell System	Bio-Rad

Materials

Gel-drying device	Heto-Dry gD-1	Heto Lab Equipment
Homogenizer	Cell Disruptor M-110P Labsonic P French Pressure Cell FA-078	Microfluidics B. Braun Thermo Spectronic
Imager	Universal hood II Electronic UV-Transilluminator	Bio-Rad Ultra Lum
Incubator	B15 (small)	Heraeus
ITC	VP-ITC	Microcal (Malvern)
Light source	KL 2500 LCD	Zeiss
Magnetic stirrer	MR 3002	Heidolph
MALS (<u>M</u> ulti <u>a</u> ngle <u>l</u> ight <u>s</u> cattering)	DAWN hELEOS 8 ⁺ Optilab T-rEX	Wyatt Technology
Microscope	Stemi 2000-C	Zeiss
PCR Cycler	Real time (MX3005P) Mastercycler ep gradient S	Stratagene Eppendorf
pH-Meter		Schott
Pipettes	Pipet-Lite LTS Easypet Pipetus	Mettler Toledo / Rainin Eppendorf Hirschmann Laborgeräte
Pipetting robot	Lissy 2002	Zinsser Analytic
Rotors	JLA-25.50 JLA-10.500 JLA-16.250 JA-8.1000 JS-5.0	Beckmann Coulter
Scanner	Odyssey	Li-Cor Biosciences
Sealing robot	RoboSeal	Zinsser Analytic
Shaker incubator	KS 15 Certomat U Innova 4300 Incubator Shaker ISF-1-W ISF-1-X LT-X Thermomix comfort	Edmund Bühler B. Braun New Brunswick Science Kühner Kühner Kühner Eppendorf
Spectro photometer	Bio-Photometer NanoDrop ND1000	Eppendorf Peqlab
Thermoblock	Rotilabo Block-Heater h250	Carl Roth
Vortexer	Vortex genie 2	Scientific Industries
X-ray detector	R-Axis IV++ R-Axis hTC	Rigaku
X-ray generator	Ultra X18 MicroMax-007 hF	Rigaku

Table 4 - Consumables

<i>Product</i>	<i>Supplier</i>
24-well Crystallization plate	Crystalgel
96-well Crystallization plate	Greiner Bio-One
96-well PCR plate	Carl Roth
Blotting membrane PUDF Rotilabor (0.45 µM)	Carl Roth
Baysilone silicone paste	GE Bayer
Cap, silicone (Silicosene type), different sizes	Hirschmann Laborgeräte
<u>Concentrator MWCO 30000</u>	Sartorius Stedim Biotech
Vivaspin 20	

Materials

<u>Concentrator MWCO 50000</u>	
Vivaspin 20	Sartorius Stedim Biotech
Vivaspin 6	Sartorius Stedim Biotech
Vivaspin 500	Sartorius Stedim Biotech
Millipore Amicon Ultra-15	Merck Millipore
Millipore Amicon Ultra-4	Merck Millipore
Cover slides (siliconized, $\varnothing = 22$ mM)	Jena Biosciences
<u>Cuvette</u>	
Bio-Photometer	Carl Roth
Electroporator (0.2 cm)	Bio-Rad
Quarz glass Suprasil (1.5 mM)	Hartenstein
<u>Filter</u>	
Minisart, sterile for syringe (0.22 μ M, $\varnothing = 50$ mM)	Carl Roth
Whatman 0.2 μ M	Carl Roth
Glas beads (Zirconia / silica beads; 0.5 mm)	Carl Roth
Gloves (nitril)	Star Lab
Greiner tube (15, 50 ml)	Greiner Bio-One
Needle Sterican (1.1 x 50 mM)	Carl Roth
PCR Reaction tube (0.5 ml)	Eppendorf
Petri dishes	Carl Roth
Pipette tips (10, 200, 1000, 5000 μ l)	Rainin
Pipettes (sterile; 5, 10, 50 ml)	Greiner Bio-One
Rapid-Flow Sterile Filter (150, 250, 500, 1000 ml; 0.2 μ M)	Nalgene
Reaction tubes (1.5 and 2.0 ml)	Eppendorf
<u>Resin</u>	
Co-Talon	Clontech
Protino Ni-IDA	Macherey-Nagel
Ni-NTA Agarose	Qiagen
Protino Ni-TED	Macherey-Nagel
Sealing tape Crystal Clear	Hampton Research
Syringe (1, 2, 5, 10, 20, 50 ml)	B. Braun

2.3 Enzymes, Kits and Additives

The following enzymes and additives were used during cloning, cell disruption or for the removal of the N-terminal His₆-tag during purification.

Table 5 - List of enzymes and additives

<i>Enzyme / Additive</i>	<i>Final concentration</i>	<i>Supplier</i>
Antarctic Phosphatase	0.175 U/ μ l	New England Biolabs
BSA (Bovine serum albumin)	Differing concentrations	New England Biolabs
cOmplete – EDTA-free protease inhibitor mix	1 tablet / 20 g (cell pellet)	Roche
DNase I	360.9 U/ μ l	Invitrogen
DNase	Differing concentrations	Applichem
dNTPs		New England Biolabs
Dpnl	0.35 U/ μ l	New England Biolabs
Kpnl	0.35 U/ μ l	New England Biolabs
Lysozym	20000 U/mg (cell pellet)	Carl Roth
NcoI-HF	0.5 U/ μ l	New England Biolabs
Phusion high-Fidelity DNA polymerase	0.04 U/ μ l	Thermo Scientific

Materials

PMSF – phenylmethylsulfonyl fluoride	1 mg/g (cell pellet)	Fluka
Quick T4 DNA Ligase		New England Biolabs
RecA	1 ng/μl	New England Biolabs
SpeI-HF	20 U/μl	New England Biolabs
T4 DNA polymerase	Differing concentrations	New England Biolabs
Taq DNA polymerase	0.05 U/μl	New England Biolabs
TEV (Tobacco Etch Virus) protease	Depending on the protease concentration	Monika Kuhn – Structure Biology

Table 6 - List of commercially available Kits used during this study

Kit	Application	Supplier
NucleoSpin gel and PCR Clean-up	DNA isolation / purification from agarose gels	Macherey-Nagel
NucleoSpin Plasmid Kit	DNA isolation from <i>E. coli</i> cells	Macherey-Nagel
NucleoBond Xtra Midi	DNA isolation from <i>E. coli</i> cells	Macherey-Nagel
PichiaPink™ Expression Strain Kit	<i>P. pastoris</i> expression strains	Invitrogen
PichiaPink™ Secreted Protein Vector Kit	<i>P. pastoris</i> vector for secreted protein expression	Invitrogen
PichiaPink™ Vector Kit	<i>P. pastoris</i> vector for non-secreted protein expression	Invitrogen
<i>S. c.</i> EasyComp. Transformation Kit	Kit to prepare chemical competent yeast cells for transformation	Invitrogen

2.4 Bacteria and Yeast culture

2.4.1 Bacterial strains, yeast strains and plasmids

In this thesis the following bacterial strains and yeast strains were used for plasmid generation and amplification as well as for protein expression.

Table 7 - List of bacterial strains and yeast strains

Species	Strain	Genotype	Supplier
<i>E. coli</i> (Plasmid amplification)	DH5α	<i>F- φ80dlacZΔM15 Δ(lacZYA-argF)U169 deoR recA1 endA1 hsdR17(rk-, mk+) phoA supE44 λ-thi-1 gyrA96 relA1</i>	Clontech Laboratories
<i>E. coli</i> (Plasmid amplification)	XL1-Blue	<i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacIqZΔM15 Tn10 (Tetr)].</i>	Stratagene
<i>E. coli</i>	BL21 (DE3)	<i>B F⁻dcm ompT hsdS(r_B⁻m_B⁻) galλ(DE3)</i>	New England Biolabs
<i>E. coli</i> (Protein expression)	BL21 (DE3) arctic express	<i>F⁻ ompT hsdS(r_B⁻ m_B⁻) dcm Tetr gal λ(DE3) endA Hte [cpn10cpn60 Gentr]</i>	New England Biolabs
<i>E. coli</i> (Protein expression)	BL21(DE3)- pLysS	<i>B F⁻ dcm ompT hsdS(r_B⁻ m_B⁻) gal λ(DE3) [pLysS Camr]</i>	Novagen
<i>E. coli</i> (Protein expression)	BL21(DE3)-RIL	<i>argU (AGA, AGG), ileY (AUA), leuW (CUA), Cam_r</i>	Stratagene
<i>E. coli</i> (Protein expression)	Rosetta (DE3)	<i>lactose permease (lacY) mutant, deficient in lon and ompT proteases; contains plasmid encoding argU, argW, glyT, lleX, leuW, metT, proL, thrT, thrU, and tyrU</i>	Merck Millipore

Materials

<i>E. coli</i> (Protein expression)	BL21 Star (DE3)	<i>F - ompT hsdS B (r B - m B -) gal dcm rne131</i>	Invitrogen
<i>M. smegmatis</i> (Protein expression)	mc ² 155	Efficient-plasmid-transformation (ept) mutant strain from the wild type ATCC 607 (Snapper et al., 1990)	ATCC (American Type Culture Collection; Manassas, VA, USA)
<i>P. pastoris</i> (Protein expression)	PichiaPink strain 1	<i>ade2</i>	Invitrogen
<i>P. pastoris</i> (Protein expression)	PichiaPink strain 2	<i>ade2, pep4</i>	Invitrogen
<i>P. pastoris</i> (Protein expression)	PichiaPink strain 3	<i>ade2, prb1</i>	Invitrogen
<i>P. pastoris</i> (Protein expression)	PichiaPink strain 4	<i>ade2, prb1, pep4</i>	Invitrogen

Table 8 contains the analyzed protein constructs, it also provides information about the used vectors, the protein sequences and the expression organism. All constructs contain a hexahistidine-tag (His₆-tag) either at the *N*- (pSD31, pBadm11, pPink α HC, pPinkHC, pPinkLC) or *C*-terminal (pETm11) end of the protein. Gene expression in the pBadm11 vector in *Escherichia coli* is controlled by the araBAD operon where protein expression is induced by L(+)-arabinose. The *E. coli* vector pETm11 is under the control of the T7 RNA polymerase promoter that is controlled by the LacI repressor; protein expression is induced by addition of the lactose-analog IPTG (Isopropyl- β -D-1-thiogalactopyranoside). Protein expression in *Mycobacterium smegmatis* is under the control of the acetamidase operon where the expression is induced by addition of acetamide. In the methylotrophic yeast *Pichia pastoris* protein expression is controlled by the alcohol oxidase gene *aox1*, and expression is induced by the addition of methanol. The source of origin for the FadA5 constructs (**Table 8**: 1-3) is *fadA5* (Rv3546) from *Mycobacterium tuberculosis* h37Rv. The source of origin of the Mrr1 constructs (**Table 8**: 4-20) is *mrr1* from *Candida albicans* SC5314. The vectors were obtained from EMBL (pBadm11, pETm11) or from invitrogen (pPink α -HC, pPink-HC, pPink-LC).

Table 8 - List of expression vectors

No.	Construct Name	Vector	Insert	Expression organism
1	pSD31 FadA5 WT *	pSD31	FadA5 wild type	<i>M. smegmatis</i>
2	pSD31 FadA5 C93A *	pSD31	FadA5 C93A variant	<i>M. smegmatis</i>
3	pSD31 FadA5 C93S *	pSD31	FadA5 C93S variant	<i>M. smegmatis</i>
4	pBadm11 FL **	pBadm11	Mrr1 Full Length protein (aa 1 – 1108)	<i>E. coli</i>
5	pBadm11 AR **	pBadm11	Mrr1 (aa 298 – 1033)	<i>E. coli</i>
6	pBadm11 '250' **	pBadm11	Mrr1 (aa 249 – 1108)	<i>E. coli</i>
7	pPink α HC FL **	pPink α -HC	Mrr1 Full Length protein (aa 1 – 1108)	<i>P. pastoris</i>
8	pPink α HC AR **	pPink α -HC	Mrr1 (aa 298 – 1033)	<i>P. pastoris</i>
9	pPink α HC '250' **	pPink α -HC	Mrr1 (aa 249 – 1108) WT	<i>P. pastoris</i>

Materials

10	pPinkHC FL **	pPink-HC	Mrr1 Full Length protein (aa 1 – 1108)	<i>P. pastoris</i>
11	pPinkLC FL **	pPink-LC	Mrr1 Full Length protein (aa 1 – 1108)	<i>P. pastoris</i>
12	pPinkHC ARlong **	pPink-HC	Mrr1 (aa 298 – 1033)	<i>P. pastoris</i>
13	pPinkLC ARlong **	pPink-LC	Mrr1 (aa 298 – 1033)	<i>P. pastoris</i>
14	pPinkHC '250' **	pPink-HC	Mrr1 (aa 249 – 1108) WT	<i>P. pastoris</i>
15	pPinkLC '250' **	pPink-LC	Mrr1 (aa 249 – 1108) WT	<i>P. pastoris</i>
16	pBadm11 '250' P683S **	pBadm11	Mrr1 (aa 249 – 1108) P683S variant	<i>E. coli</i>
17	pBadm11 '250' G997V **	pBadm11	Mrr1 (aa 249 – 1108) G9978V variant	<i>E. coli</i>
18	pBadm11 '250' K335N **	pBadm11	Mrr1 (aa 249 – 1108) K335N variant	<i>E. coli</i>
19	pBadm11 '250' G878E **	pBadm11	Mrr1 (aa 249 – 1108) G878E variant	<i>E. coli</i>
20	pETm11 '250' CT **	pETm11	Mrr1 (aa 249 – 1108) WT	<i>E. coli</i>

* Kindly provided by Prof. Dr. Nicole S. Sampson – Stony Brook University, NY, USA (Nesbitt et al., 2010)
 ** This work

2.4.2 Agarose gel electrophoresis

The following solutions were used for the analysis of DNA or DNA fragments via gel electrophoresis. Imaging of the DNA bands was performed with the Bio-Rad Universal hood II.

Table 9 - Components for DNA agarose gel electrophoresis

Solution	Ingredients
5x DNA sample buffer	0.01 % (w/v) bromophenol blue 50% (v/v) glycerol 1x TE buffer (10 mM Tris-HCl pH 8.0, 1 mM Ethylenediaminetetraacetic acid (EDTA))
1% Agarose gel	1% (w/v) agarose 1% (v/v) ethidium bromide / 1% (v/v) Midori green Advance (Nippon genetics) 1x TE buffer
DNA size standard	GeneRuler 1 kb DNA Ladder (2.5 µl per gel) (Thermo Scientific)

2.4.3 Antibiotics, media and additives

Antibiotics were purchased from Carl Roth (Karlsruhe), they were dissolved in ddH₂O and filtered through a sterile syringe filter.

Table 10 - List of antibiotic solutions and concentrations used

Antibiotic	Stock	Final concentration
Ampicillin	100* / 200** mg/ml	100* / 200** µg/ml
Chloramphenicol	34 mg/ml	34 µg/ml
Cycloheximid	10 mg/ml	10 µg/ml
Hygromycin	50 mg/ml	50 µg/ml
Kanamycin	50 mg/ml	50 µg/ml

* *E. coli* / ** *M. smegmatis*

Materials

Additives that could not be autoclaved were filtered through a sterile filter prior to the addition to the medium. The growth medium was autoclaved before use and was complemented, if necessary, with the non-autoclavable additives.

Table 11 - List of the used media and additives

Organism	Medium / Additive	Supplier / Information
<i>Escherichia coli</i>	SOC – S uper O ptimal Broth medium 20 mM glucose, 2% (w/v) peptone, 5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 20 mM MgCl ₂	
<i>Escherichia coli</i>	LB agar	Carl Roth
	LB liquid medium (Lennox)	Carl Roth
<i>Mycobacterium smegmatis</i>	7H9 liquid medium (Middlebrook)	Fluka
	7H10 solid medium (Middlebrook)	Fluka
<i>Pichia pastoris</i>	Glycerol (Final concentration: 0.5%)	Carl Roth
	BMMY - B uffered M ethanol-complex M edium for Y east 1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate, pH 6.0, 1.34% (w/v) YNB (Yeast Nitrogen Base), 0.00004% (w/v) biotin, 0.5% (v/v) methanol	(Invitrogen, 2010)
	BMGY - B uffered glycerol-complex M edium for Y east Recipe as above but with 1% (v/v) glycerol instead of methanol	
	PAD Agar – P ichia- A denine- D ropout Agar 13.4 g/L Yeast Nitrogen Base with Ammonium Sulfate without amino acids, 1.25 g/L Adenine Amino Acid Mix (CSM-ADE from MP Bio), 2% (w/v) agar, 5mg/L biotin, 2% (v/v) glucose	
	YP Agar 1% (w/v) yeast extract, 2% (w/v) peptone, 2% (v/v) glucose, 2% (w/v) agar	
	YPD 1% (w/v) yeast extract, 2% (w/v) peptone, 2% (v/v) glucose	

2.5 Protein purification

2.5.1 Buffers and solutions

Table 12 - List of buffers for protein purification and analysis

Protein / Organism	No	Purification technique	Buffer	Composition	Annotations
FadA5 WT and variants <i>Mycobacterium smegmatis</i>	1	Affinity chromatography Co-Talon resin	Binding buffer	50 mM Tris-HCl pH 7.4 300 mM NaCl 10 mM imidazole 10% (v/v) glycerol	---
	2		Elution buffer	50 mM Tris-HCl pH 7.4 300 mM NaCl 150 mM imidazole 10% (v/v) glycerol	---
	3	Anion-exchange chromatography MonoQ 10/100 GL	MonoQ buffer low salt	50 mM Tris-HCl pH 7.6 50 mM NaCl	sterile /degased
	4		MonoQ buffer high salt	50 mM Tris-HCl pH 7.6 1.0 M NaCl	sterile /degased
	5	Size exclusion chromatography HiLoad 26/60 Superdex 200 pg	Superdex buffer	20 mM Bicine-NaOH pH 8.5 250 mM NaCl	sterile /degased
	5o		Superdex buffer	60 mM Bicine-NaOH pH	sterile /degased

Materials

			(old)	8.5 250 mM NaCl	
Mrr1 Full length (FL) and Armadillo Repeat (AR) constructs <i>Escherichia coli</i>	6	Affinity chromatography Co-Talon, Ni-IDA, Ni-NTA, Ni-TED resins	Binding buffer	100 mM Tris-HCl pH 8.0, 500 mM NaCl 5 mM imidazole	
	7 I - V		Wash buffers	Binding buffer with 15, 30, 45, 60, 75 mM imidazole	
	8		Wash buffer special	Binding buffer with 10 mM imidazole 2 mM ATP 10 mM MgCl ₂	
	9 I / II		Elution buffers	Binding buffer with 150 / 250 mM imidazole	
Mrr1 250 variant with N-terminal His ₆ -tag <i>Escherichia coli</i>	22	Affinity chromatography Ni-NTA resin	Binding buffer	40 mM NaPhosphate pH 8.0 500 mM KCl 10 mM imidazole	
	23 I - VI		Wash buffers	Binding buffer with 25, 35, 45, 50, 75, 100 mM imidazole	
	24 I / II		Elution buffers	Binding buffer with 150, 250 mM imidazole	
Mrr1 250 with C-terminal His ₆ -tag <i>E. coli</i>	42	Affinity chromatography Ni-NTA resin	Binding buffer	20 mM Tris-HCl pH 8.0 300 mM KCl 5 mM imidazole	
	43 I - III		Wash buffers	Binding buffer with 25, 50, 75, 100 mM imidazole	
	44 I/II		Elution buffers	Binding buffer with 100, 250 mM imidazole	
Mrr1 250 variant <i>Escherichia coli</i>	29	Anion-exchange chromatography MonoQ (MQ) 5/50 GL	MQ lowsalt buffer	50 mM Bicine-NaOH pH 8.5 50 mM KCl	sterile /degased
	30		MQ highsalt buffer	50 mM Bicine-NaOH pH 8.5 1 M KCl	sterile /degased
Mrr1 <i>Escherichia coli</i>	FL AR 250	Size-exclusion (SE) chromatography Superdex 200 columns different sizes	SE buffer	100 mM Tris-HCl pH 8.0 300 mM NaCl	sterile /degased
	250		SE buffer	20 mM Bicine-NaOH pH 8.0 300 mM KCl	sterile /degased
Mrr1 250 variants <i>E. coli</i>	38	TEV cleavage	TEV buffer	50 mM Tris-HCl pH 8.0 0.5 mM EDTA 1 mM DTT	
Mrr1 <i>Pichia pastoris</i>	39	Affinity chromatography Ni-NTA resin (Expression test)	Breaking buffer	50 mM NaPhosphate pH 7.4 1 mM PMSF 1 mM EDTA 5% (v/v) glycerol	Recipe Invitrogen PichiaPink manual
Mrr1 250 variants <i>E. coli</i>	40	CD spectroscopy	CD buffer	20 mM Bicine-NaOH pH 8.0 300 mM NaF	
FadA5 variants	41	Isothermal titration calorimetry (ITC)	ITC buffer	50 mM MES pH 6.7 250 mM NaCl	

Materials

2.5.2 Protein gel electrophoresis

Table 13 - Buffers and Solutions for SDS-PAGE

<i>Solution</i>	<i>Ingredients</i>
Coomassie G250 safe stain	36 mM HCl, 94 μ M Coomassie-Brilliant Blue g-250
Coomassie R250 stain	50% (v/v) ethanol, 10% (v/v) acetic acid, 0.1% (w/v) Coomassie-Brilliant Blue R-250
SDS-PAGE gel Destaining solution	10% (v/v) ethanol, 5% (v/v) acetic acid
1x SDS-PAGE Running buffer	25 mM Tris, 192 mM glycine, 0.1% (w/v) SDS
5x SDS-PAGE Sample buffer	250 mM Tris-HCl pH 6.8, 0.5% (w/v) bromophenol blue, 500 mM dithiothreitol (DTT), 10% (w/v) SDS, 50% (v/v) glycerol
15% Separating gel solution	15% (v/v) Rotiphorese gel 30 (37.5:1), 167 mM Tris-HCl pH 8.8, 0.1% (w/v) SDS, 0.1% (v/v) APS, 0.1% (v/v) TEMED
5% Stacking gel solution	5% (v/v) Rotiphorese gel 30 (37.5:1), 125 mM Tris-HCl pH 6.8, 0.1% (w/v) SDS, 0.1% (v/v) APS, 0.1% (v/v) TEMED
Protein size standards	PageRuler Prestained Protein Ladder (Thermo Scientific) PageRuler Plus Prestained Protein Ladder (Thermo Scientific)

2.5.3 Western blotting

Table 14 - Solutions for Western Blotting

<i>Solution</i>	<i>Ingredients</i>
Blotting buffer	25 mM Tris, 1.92 mM glycine, 20% (v/v) methanol, 0.1% (w/v) sodium dodecyl sulfate (SDS)
Ponceau red staining solution	0.1% (w/v) Ponceau S, 5% (v/v) acetic acid
10x TBS buffer	195 mM Tris, 1.5 M NaCl, pH adjusted to 7.6 by HCl
Blocking solution	1x TBS, 0.05% (v/v) Tween 20, 3 g milk powder
TBST buffer	1x TBS, 0.05% (v/v) Tween 20
anti-His antibody	in blocking solution (Cell signaling technology – Merck Millipore)
HRP antibody	in blocking solution (Cell signaling technology – Merck Millipore)
ECL, enhanced chemiluminescent (ECL) substrates for detection of horseradish peroxidase (HRP)	Thermo Scientific

2.6 Crystallization screens and compounds

The commercially available crystallization screens used in this study were produced utilizing the Lizzy 2002 pipetting robot (Zinsser Analytic) according to the manufacturer's protocols. The self-designed screens were also produced with the Lizzy 2002 pipetting robot.

Table 15 - Crystallization Screens

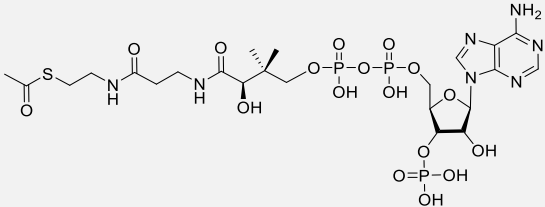
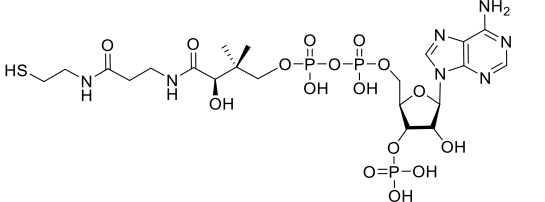
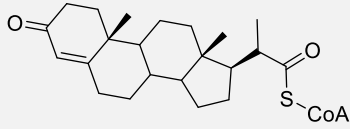
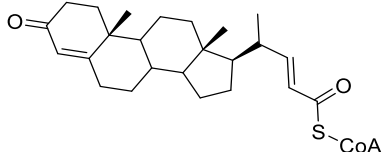
<i>No.</i>	<i>Screen</i>	<i>Supplier</i>
1	Additive Screen hR2-138	Hampton Research
2	Crystal Screen I/II	Hampton Research
3	Index Screen	Hampton Research

Materials

4	Midas	Molecular Dimensions
5	Nextal PEGs Suite	Qiagen
6	Nextal pHClear Suite	Qiagen
7	Nucleix Suite	Qiagen
8	Silver Bullets Screen	Hampton Research
9	Topaz OptiMix I	Fluidigm
10	Topaz OptiMix II	Fluidigm
11	Topaz OptiMix III	Fluidigm
12	Topaz OptiMix V	Fluidigm
13	Topaz OptiMix-PEG	Fluidigm
14	Wizard Screen I/II	Emerald BioSystems
15	FadA5 screen	Self-designed (Table S 4)
16	Thiolase screen	Self-designed (Table S 5)

The following compounds were purchased from Sigma Aldrich (Seelze) or synthesized and kindly provided by Professor Nicole Sampson's laboratory ([Table 16](#)).

Table 16 - Crystallization Compounds

Compound	Information
	acetyl Coenzyme A sodium salt Sigma Aldrich (A2056)
	Coenzyme A sodium salt hydrate Sigma Aldrich (C3144)
	3-oxo-pregn-4-ene-20-carboxyl-CoA (3-OPC-CoA) M _w = 1089.98 Da (Schaefer et al.)
	3-oxo-cholesterol-4,22-diene-24-oyl-CoA (3-OCD-CoA) M _w = 1116.02 Da

2.7 Software and Databases

Table 17 - Computer software and Databases

<i>Software</i>	<i>Reference</i>
Aimless	(Evans, 2011)
ApE – A plasmid Editor	http://biologylabs.utah.edu/jorgensen/wayned/ape/
Arp/Warp	(Morris et al., 2003)
ASTRA 6 (MALLS)	Wyatt Technology
BLAST	http://blast.ncbi.nlm.nih.gov/Blast.cgi
CCP4 suite 6.3.0	(1994; Winn et al., 2011)
ClustalW	(Thompson et al., 1994)
Coot 0.7	(Emsley and Cowtan, 2004)
Dssp / Dssp2pdb	(Kabsch and Sander, 1983) / (James Stroud, 2002)
EndNote X7	endnote.com / Thomson Reuters
ESPrpt	(Gouet et al., 2003)
ExPASy Proteomics Server	http://www.expasy.ch/
HotspotsX	(Neudert and Klebe, 2011)
iMosflm 1.0.7	(Battye et al., 2011)
iTasser Server	(Roy et al., 2010; Zhang, 2008)
Matthews	(1994)
MolProbity Server	(Chen et al., 2010)
National Center for Biotechnology information (NCBI)	http://www.ncbi.nlm.nih.gov/
Phaser	(McCoy, 2007)
PDB - RCSB Protein Data Bank / PDBe	(Velankar et al., 2012)
PDBeFold	(Krissinel and Henrick, 2004)
PDBePISA	(Krissinel and Henrick, 2007)
Phenix 1.9-1692	(Adams et al., 2010)
Phyre ² Server	(Kelley and Sternberg, 2009)
Pointless	(Evans, 2011)
Porter	(Pollastri and McLysaght, 2005)
PRODRG server	(Schuttelkopf and van Aalten, 2004)
ProtParam	(Gasteiger et al., 2003)
PyMOL 1.6.0.0	(2010)
Refmac 5.5.0102	(Murshudov et al., 1997; Winn et al., 2003)
Scala	(Evans, 2006)
SCRATCH protein predictor	(Cheng et al., 2005)
TLSMD server	(Painter and Merritt, 2006)
XDS package	(Kabsch, 2010)

3 Methods

3.1 Molecular biology and microbiology techniques

3.1.1 Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR), originally developed by Saiki *et al.*, allows the amplification of distinct DNA fragments with the help of a thermostable DNA polymerase, a primer pair and deoxyribonucleotide triphosphates (dNTPs). (Saiki et al., 1988) At the beginning of the PCR the DNA double strand is melted (denaturation), so that the specific primer can then anneal with the insert or vector DNA which is the starting point of the reaction (annealing). The steps of denaturation, annealing and elongation are called a PCR cycle, which is repeated between 20 and 40 times during the entire process. In the elongation process the chosen DNA polymerase synthesizes complementary DNA strands. PCR can be used to amplify DNA for the purpose of cloning (3.1.6.1, 3.1.6.2), by generating defined DNA fragments; for site-directed mutagenesis (3.1.6.3), to generate DNA constructs with distinct point mutations; or for the evaluation of successful construct insertion into plasmids.

In the present work the Phusion high-Fidelity DNA polymerase from Thermo Scientific (Table 5) with a very low error rate and short protocol times was used for cloning and site-directed mutagenesis whereas the *Taq* DNA polymerase from New England Biolabs was used for the evaluation processes (Colony PCR). The primer pairs that were used for cloning, site-directed mutagenesis and evaluation processes in this work are listed in Table S 1. For protein expression in *E. coli* the *E. coli*-expression optimized Mrr1 Full length (FL) sequence (Geneart, Regensburg) and the truncation variants Armadillo Repeat (AR) and '250' (Table 8) were cloned into pBADM11. The Mrr1 point mutants were generated via site-directed mutagenesis using pBADM11 '250' as a template. In addition the '250' variant was cloned into pETM11 just containing the C-terminal His₆-tag. For protein expression in *P. pastoris* the Mrr1 FL, '250' and AR constructs were cloned into the pPink-HC, pPink-LC and pPink α HC vectors (Table 8). All constructs were verified by sequencing (Seqlab, Göttingen). The primers for the PCR were purchased from biomers.net GmbH (Ulm) or Sigma Aldrich (Seelze).

The pipetting scheme and the cycle information are listed in Table 18 and Table 19. For cloning and site-directed mutagenesis two primer-separated reactions were set up and the PCR protocol was performed for 8 cycles. Then the samples, primer 1 and primer 2, were combined and the PCR was performed for another 25 cycles. The result was analyzed by DNA agarose gel electrophoresis (3.1.2) and the reaction mixture was purified with the NucleoSpin gel and PCR Clean-up Kit (Macherey-Nagel) to obtain pure amplified DNA without contaminants (3.1.3).

Methods

Table 18 - PCR cycle and pipetting scheme for cloning or site-directed mutagenesis

Cloning / Site-directed mutagenesis		Pipetting scheme		
Initialization	98 °C, 30 s	DNA Template	200 ng	1.0 µl (each)
Denaturation	98 °C, 10 s	Primer	0.8 µM	5 µl (primer 1) 5 µl (primer 2)
Annealing	T _m °C*, 30 s	dNTP	0.4 mM	1.0 µl (each)
Elongation	72 °C, 30 s/kb	Phusion buffer HF / GC	1x	5.0 µl (each)
Final Extension	72 °C, 5 min	Phusion HF DNA polymerase	0.02 U	0.5 µl (each)
Cooling	4 °C, ∞	ddH ₂ O		12.5 µl (each)
* T _m = melting temperature primer				50 µl

Table 19 - PCR cycle and pipetting scheme for the colony PCR

Colony PCR				Pipetting scheme		
		<i>E. coli</i>	<i>P. pastoris</i>			
Initialization	98 °C	5 min	2 min	DNA Template		1 colony
Denaturation	95 °C	30 s	1 min	Primer pair	0.4 µM	0.4 µl (each)
Annealing	T _m °C*	30 s/kb	1 min	dNTP	0.2 mM	0.4 µl
Elongation	72 °C	30 s	1 min	ThermoPol buffer	1x	2.0 µl
Final Extension	72 °C	10 min	7 min	Taq DNA polymerase	0.01 U	0.2 µl
Cooling	4 °C		∞	ddH ₂ O		16.6 µl
* T _m = melting temperature primer						20 µl

3.1.2 DNA agarose gel electrophoresis

The results of the PCR or restriction digestion were analyzed by agarose gel electrophoresis. The applied DNA samples are separated on an agarose gel matrix according to their negative charge, which is proportional to the DNA length and thereby the DNA size. Smaller DNA fragments are moving faster to the anode through the agarose matrix than larger fragments. Thereby a separation of the sample by DNA length is achieved. (Sambrook and Russell, 2001)

For agarose gel electrophoresis 1% (w/v) agarose was melted in 1x TE buffer, poured into a gel caster and supplied with Midori green for DNA staining (Table 9). The DNA samples were prepared by adding 3 µl of 5x DNA sample buffer to 12 µl of the according PCR or restriction digestion sample. 1x TE was used as running buffer and 2.5 µl of the 1 kb DNA ladder was applied as a standard. DNA was separated electrophoretically, according to molecular size at a constant voltage of 120 V for 35 minutes. For subsequent analysis the agarose gel was exposed to UV-light in the Universal hood II (Bio-Rad) and recorded with the gel Doc XR System.

3.1.3 DNA purification

After the analysis by agarose gel electrophoresis the chosen reaction mixtures have to be cleaned, so that the DNA fragments, generated by PCR, are free of the former reaction partners or the initial DNA used for amplification. To purify the PCR mixture the NucleoSpin gel and PCR Clean-up Kit

Methods

(Macherey-Nagel) was used. In case of extensive contamination, the clean-up procedure can be extended by a gel extraction process in combination with the PCR Clean-up Kit.

The cleaning process was performed at room temperature. After amplification, the DNA was mixed with NTI buffer (1 volume DNA sample + 2 volumes NTI buffer), pipetted into the provided spin column and centrifuged for 10 s at 16,100 g. The flow through of the column was reloaded onto the column twice and centrifuged in the same manner. Afterwards the flow through was discarded and the silica membrane of the column was washed twice with 700 µl NT3 buffer. After washing, the silica membrane was dried at 16,100 g for 1 minute and the DNA was eluted with 20-30 µl of preheated (70 °C) NE buffer by centrifugation (16,100 g, 1 minute). The DNA was used immediately or stored at -20 °C until further use.

3.1.4 Transformation and genetic integration

Transformation is a process where exogenous genetic material, DNA, is selectively taken up by an organism, through the cell wall. (Dorland, 2007)

For plasmid amplification or recombinant protein expression exogenous DNA has to be taken up by the host organism. In this work *E. coli* strains DH5α or XL1 blue were used for plasmid amplification of all constructs and for cloning or mutagenesis approaches. As protein expression was performed in three different organisms, transformation of plasmid DNA into *E. coli* and *M. smegmatis* as well as genetic integration into the genome of the yeast *P. pastoris* was performed.

3.1.4.1 Transformation into *E. coli* cells

As *E. coli* cells are not competent for DNA uptake by nature, they have to be treated with ice-cold CaCl₂-solution in their exponential growth phase to be made chemo-competent. By this treatment the cell wall becomes permeable for the entry of plasmid DNA into the cell. The permeability of the cell walls is further increased by either heat shock treatment or electroporation shortly after the addition of DNA to the cell suspension. Even though the cells are competent and the cell walls should be permeable, the uptake of DNA is still a quite random event with success rates of around 0.1%. (Chen and Dubnau, 2004) This issue and the fact that additional DNA is of disadvantage for the organism makes it necessary to introduce selective antibiotic resistance markers, which are included in the plasmid DNA. The cell culture is therefore supplemented with an antibiotic and will thus only permit the growth of bacteria which have taken up the additional plasmid containing the antibiotic resistance gene and the gene of interest.

Methods

The *E. coli* strains used during this work are listed in **Table 7**. Prior to transformation the chemo-competent cells were thawed on ice and then 100 ng of plasmid DNA (for plasmid amplification) or 20 – 400 ng of the cloning sample were added. The mixture was incubated for at least 10 or 20 minutes (cloning) on ice followed by the heat shock at 42 °C for 75 s. After at least 2 minutes incubation on ice 400 µl LB medium or 400 µl SOC medium (for cloning reactions) were added. The cells were incubated while shaking (600 rpm) for 45 minutes at 37 °C. In case of cloning reactions, the cells were carefully pelleted at 2000 g for 1 minute, 250 µl of the supernatant were discarded and the remaining medium was used for resuspension. The resuspended cells were plated onto an LB agar plate, supplemented with the selection antibiotics, and were incubated overnight at 37 °C to permit colony growth. For plasmid amplification or protein expression 200 µl of the cell suspension were directly plated onto an LB agar plate without centrifugation.

3.1.4.2 Transformation into *M. smegmatis* cells - electroporation

The mycobacterial cell wall is especially stable (1.3.1.2) and transformation is only effective if it is performed by a physical transformation method, namely electroporation. (Brennan, 2003; Parish and Brown, 2008)



Figure 12 - *M. smegmatis* colonies

An effective electroporation process requires electro-competent *M. smegmatis* cells. The mc²155 cells purchased from ATCC (Manassas, VA, USA) were prepared according to the manufacturer's protocol and were rendered competent during their exponential growth phase, utilizing an ice-cold 10% glycerol solution according to the procedure described in *Mycobacteria* Protocols (Parish and Brown, 2008). Mycobacterial plasmids also contain antibiotic selection markers to monitor successful plasmid entry.

Prior to transformation, an aliquot (200 µl) of the electro-competent mc²155 *M. smegmatis* cells was thawed on ice and then 100 ng of plasmid DNA (**Table 8**) were added. The mixture was incubated on ice for 10 minutes, transferred into a 0.2 cm electroporation cuvette (Bio-Rad) and then electroporated (2.5 kV, 10 µF, 600 Ω). The suspension was supplied with 600 µl pre-cooled 7H9-medium to chill on ice for 10 minutes and then incubated for 4 h at 37 °C, while shaking (200 rpm).

The following work was performed under sterile conditions in a biological safety cabinet (Hera Safe KS12, Kendro). The bacteria were resuspended and the complete sample was plated onto 7H10 solid medium, which was supplemented with the selection marker hygromycin (50 µg/ml), with

Methods

cycloheximid (10 µg/ml), against yeast or fungi contaminations, and ampicillin (200 µg/ml), against bacterial contaminations. The plates were wrapped into a layer of aluminum and incubated at 37 °C for 4-5 d until the colonies reached a diameter of at least 2 mm (**Figure 12**). (Parish and Brown, 2008; Singh and Reyrat, 2009) The grown colonies were used for the inoculation of the starting culture (3.1.7.2).

3.1.4.3 Genetic integration into *P. pastoris*

The cloning processes and plasmid amplifications of the yeast constructs were performed in *E. coli* as described in 3.1.4.1. In contrast to the above described techniques, the pichia vectors were genetically integrated into the *Pichia* genome and not only transformed. The advantage of the integration is the generation of a stable *Pichia* strain containing the gene of interest, in variable copy numbers, in the genome. Once a positive integrant is identified, glycerol stocks of the positive strain can be prepared and used for all following expressions. Yeast vectors for genetic integration do not only contain an antibiotic selection marker, as *AmpR*, for cloning and amplification processes in *E. coli*, but these vectors contain an additional, non-antibiotic selection function for the expression in yeast. Some common genes that are used for selective expression processes in yeast are *ura3* (orotidine-5'-phosphate decarboxylase), *lys2* (α -amino adipate reductase), *ade1* (phosphoribosylamino-imidazole-succinocarboxamide synthetase) and *ade2* (phosphoribosylamino-imidazole-carboxylase), the *gal1* promoter or the *lacZ* (β -galactosidase) reporter gene. All those genes have in common that mutations, deficiencies or insertions permit selection based on distinct media or based on the colony color. In the system used in this thesis, selection was performed by the complementation of the *ade2* deficient strains with the *ade2* gene on the plasmid, which was genetically integrated together with the gene of interest. Strains, deficient in *ade2*, suffer from adenine auxotrophy, which leads to the accumulation of purine precursors in the vacuole and causes a red or pink coloring of the colonies. Colonies that display a creamy white coloring were successfully complemented with the *ade2* gene and therefore contain the gene of interest. (Invitrogen, 2010; Mülhardt, 2009)

For the genetic integration of plasmids into *P. pastoris* either chemo-competent or electro-competent cells are necessary. During this work, exclusively chemo-competent cells prepared with the *S. c.* EasyComp. Transformation Kit (Invitrogen), according to the manufacturer's protocol, were used.

Prior to transformation the plasmid DNA had to be digested with the restriction enzyme *SpeI* that cuts the DNA within the *trp2* locus of the plasmid and enables genetic integration into the *trp2* locus of the pichia genome. For the digestion procedure 10 µg of plasmid DNA were added to a reaction mixture containing *SpeI* (1x NEB4 buffer, 2 µg BSA, 20 U *SpeI*). The total volume of the mixture,

Methods

including DNA, was adjusted with ddH₂O to 20 µl. After incubation at 37 °C for 2.5 h the enzyme was heat inactivated for 20 minutes at 80 °C. The digested plasmids were subsequently analyzed via gel electrophoresis (3.1.2), cleaned by DNA purification (3.1.3) and the final DNA concentration was assessed spectrophotometrically at a wavelength of 260 nm using the NanoDrop ND 1000 spectrophotometer.

The competent pichia cell aliquots (100 µl) were thawed at room temperature and then further aliquoted into 50 µl portions. 5-10 µg of the digested and cleaned DNA were added to the cells, which were then supplemented with 1 ml of solution II (*S. c.* EasyComp. Transformation Kit). The cells were vortexed and incubated at 30 °C for 1 h under shaking (350 rpm), which was interrupted every 15 minutes by additional vortexing. After incubation, the cells were heat shocked for 10 minutes at 42 °C and subsequently the mixture was split into 2 microcentrifuge tubes and supplemented with 1 ml YPD medium each. After 3 h incubation at 30 °C under stirring (350 rpm) the cells were pelleted at 3,000 g for 5 minutes. The supernatant was discarded and after addition of 500 µl solution III (*S. c.* EasyComp. Transformation Kit) the cells were combined and pelleted as described before. The supernatant was again discarded, the cells were resuspended with newly added solution III (150 µl) and plated onto PAD agar selection plates (**Table 11**). The plates were incubated for 3-5 d at 30 °C to enable colony growth. The positive, white transformants were re-streaked on PAD agar plates and incubated for another 2 days at 30 °C to exclude false positives. The plates were sealed with parafilm for storage at 4 °C until further use.

3.1.5 Plasmid amplification and isolation

Plasmid DNA has to be amplified for differing purposes. On the one hand sufficient DNA has to be available to ensure the correctness of the cloning approach by a DNA double digestion test or sequencing and on the other hand DNA has to be amplified to maintain a sufficient stock of DNA for transformation and subsequent expression experiments and as template for cloning approaches. For the purpose of amplification specially tailored *E. coli* strains are used, that for example prevent the degradation of non-methylated DNA. (Taylor et al., 1993)

For plasmid amplification approx. 100 ng of DNA (**Table 8**) were transformed into DH5α or XL1blue cells (3.1.4.1) and plated onto LB agar plates. One colony was chosen from the plate to inoculate 5 ml or 100 ml LB medium, supplemented with 100 µg/ml ampicillin. The cultures were grown overnight under stirring (200 rpm) at 37 °C and were harvested by centrifugation. The amplified plasmid DNA was extracted using the NucleoSpin Plasmid Kit or the NucleoBond Xtra Midi Kit (Macherey-Nagel), dependent on the culture volume, following the manufacturer's protocol. In short, the cell pellets were resuspended, lysed and separated from the cell debris through filtration which led to loading of

Methods

the DNA onto a silica membrane that was thoroughly washed subsequently. After elution of the DNA from the membrane, the DNA concentration was determined spectrophotometrically at 260 nm with the NanoDrop ND 1000. When the DNA was purified with the Midi Kit, the elution fraction of the silica membrane was precipitated with isopropanol and washed with ethanol to remove contained salt. After the ethanol wash step, the DNA was pelleted by centrifugation at 16,100 g, for 5 minutes. The ethanol supernatant was discarded and the pelleted DNA was then resuspended in water and the final DNA concentration was assessed with the NanoDrop. The DNA was stored at -20 °C for later use.

3.1.6 Cloning Techniques

Molecular cloning or just cloning is a method to manipulate DNA (plasmid or genomic DNA) of a host organism to incorporate a foreign gene of interest. (Sambrook and Russell, 2001) Cloning is the first step to yield recombinant protein expression in host organisms for scientific or industrial use. With the help of various and constantly increasing numbers of cloning methods, the foreign DNA is tailored into a plasmid or vector that later enables transport of the gene of interest into the host organism, as described in the transformation or genetic integration chapters above (3.1.4). (Mülhardt, 2009)

3.1.6.1 Restriction digestion and ligation

By restriction digestion the vector and the gene of interest are digested with the help of one or two endonucleases to create overlapping sticky ends in the vector and the gene (**Figure 13A**). Since these ends are complimentary to each other (see **Figure 13A**) the overlapping ends can be fused with the help of a ligase, thus finally generating an elongated vector without any nicks in the DNA (**Figure 13A**). After verification of the construct's correctness, the vector can then be transformed or genetically integrated into the host organism for protein expression.

Restriction digestion and ligation was used during this work to generate the Mrr1 pBADM11 '250' construct, as other cloning methods such as SLIC (3.1.6.2) failed. In advance to the restriction digestion the DNA for the '250' construct had to be amplified via PCR using the pBADM11 FL construct as template and to thereby obtain the truncated Mrr1 version. Amplification, verification and purification of DNA was performed as described in chapters 3.1.1 to 3.1.3. For the amplification of the Mrr1 '250' construct, a primer pair (**Table S 1**) was used that contained specific restriction sites for the endonucleases NcoI-HF and KpnI that were thereby added to either ends of the gene of interest. The NcoI restriction site was introduced 5' to the insert and the KpnI restriction site 3' to the

Methods

insert. The concurrent insertion of the restriction sites during amplification enables restriction digestion of the insert and the vector, which also contains the same restriction sites. After the length of the insert was validated by gel electrophoresis, the pBadm11 vector and the newly created insert were digested with the help of NcoI-HF and KpnI. For this purpose 500 ng of the purified insert DNA were added to a reaction mixture containing 0.25 U of each restriction enzyme, as well as 2 µg BSA, in a total volume of 20 µl 1x NEB2 buffer in ddH₂O. A similar reaction mixture was set up containing 800 ng pBadm11 vector DNA. The digestion samples were incubated at 37 °C for 2 h and afterwards cleaned with the NucleoSpin gel and PCR Clean-up Kit. The efficiency of the vector digestion was checked by gel electrophoresis and DNA concentrations were determined spectrophotometrically with the NanoDrop ND 1000 at 260 nm.

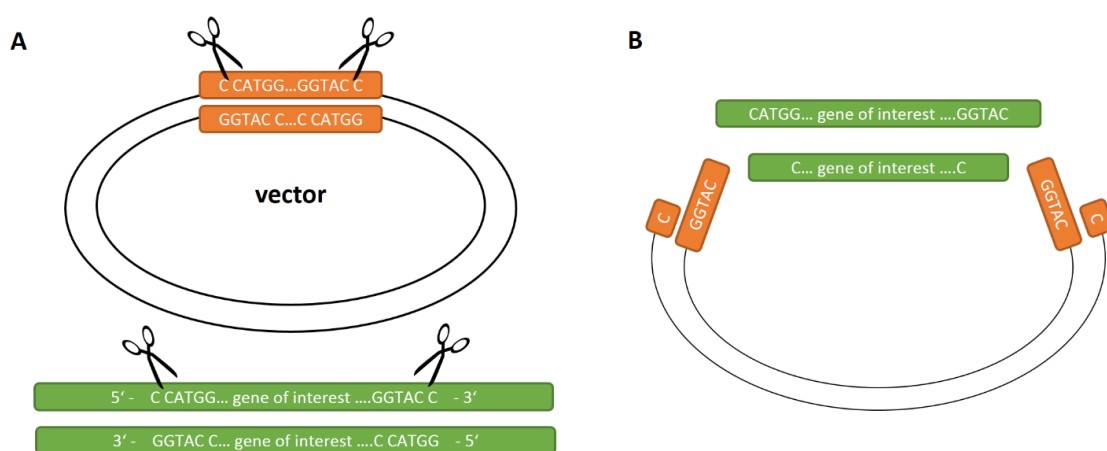


Figure 13 - Principle of cloning by restriction digestion and ligation:

A) The vector (black and orange) contains complementary digestion sites to the DNA (green) containing the gene of interest. The restriction sites are indicated by the scissors. **B)** After digestion, the gene of interest and vector are ligated via their complimentary digestion sites. Adapted from: (Berg et al., 2007)

The digested pBadm11 vector was further treated with Antarctic phosphatase (Table 5) to remove the 5' phosphates of the DNA, which prevents recircularization of the cloning vectors in the absence of the insert. 250 ng of digested and purified vector DNA were incubated for 1 h at 37 °C with 0.12 U of Antarctic phosphatase in 1x NEBcutsmart buffer adjusted with ddH₂O to a total volume of 20 µl. After heat inactivation of the phosphatase for 5 minutes at 65 °C, the reaction mixture was cleaned with the Clean-up Kit and was used in the ligation reaction.

The purified restriction products were ligated employing Quick T4 DNA ligase (Table 5) and the appropriate Quick ligation reaction buffer. A total of 100-200 ng DNA were used for the ligation reaction, with a molar ratio of 1:4 with respect to the vector and the insert DNA. The DNA was incubated with 1.5 µl Quick T4 DNA ligase in 1x Quick ligation reaction buffer, adjusted to a total volume of 20 µl with ddH₂O, for 20 minutes at room temperature. After incubation the complete ligation mixture was transformed into DH5α cells as described in 3.1.4.1. The grown colonies were

Methods

checked by colony PCR (3.1.6.4) to verify the cloning approach. Chosen positive clones were amplified (3.1.5) and finally verified by sequencing via SeqLab (Göttingen).

3.1.6.2 Sequence and ligation independent cloning (SLIC)

SLIC is a cloning method that does not employ any restriction enzymes or ligation reactions but is using *in vitro* homologous recombination of overlapping DNA fragments. (Li and Elledge, 2007) SLIC uses 'natural' recombination of artificially, by PCR produced fragments via their overlapping

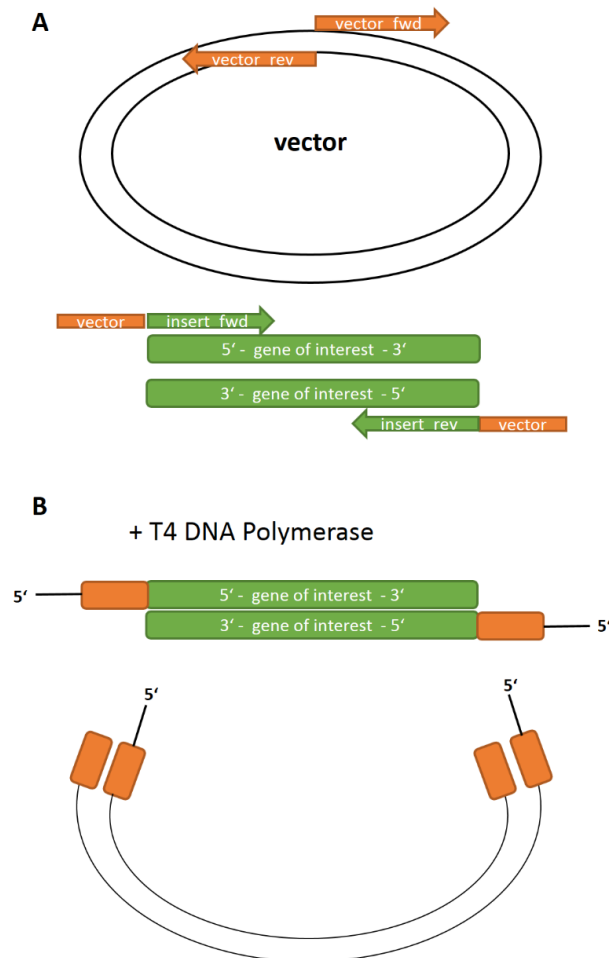


Figure 14 - SLIC:

A) Generation of overlapping fragments by PCR with overlapping primer for vector (orange) and insert (green). **B)** The purified PCR products are excised at the 3' ends by T4 DNA polymerase to generate 5' overhangs. The overhangs are subsequently recombined with RecA or by the recombination system in *E. coli*. Adapted from: (Li and Elledge, 2007)

homologous ends. The success rate of this technique is also very high if more than two fragments are recombined. The first step of SLIC is the generation of fragments with overlapping ends (Figure 14A) by PCR. Subsequently those fragments are treated with T4 DNA polymerase, without NTP's, which provokes an exonuclease activity of the polymerase and leads to single stranded 5' overhangs (Figure 14B). The fragments with those single stranded DNA ends are either recombined with the help of RecA and subsequently transformed into *E. coli* or they are directly transformed and the endogenous repair machinery of the bacteria is performing the recombination reaction.

The fragments necessary for the SLIC process were generated by PCR (3.1.1), with the according primer pairs (Table S 1) and were analyzed by gel electrophoresis. Prior to purification the fragments were digested with DpnI to remove the methylated template DNA. The complete PCR mixture was supplied with 0.012 U DpnI and 1x NEB4 buffer and was adjusted to a total volume of 60 µl with ddH₂O.

The mixture was incubated overnight at 37 °C and was subsequently purified with the NucleoSpin gel and PCR Clean-up Kit. The DNA concentration was determined spectrophotometrically with the NanoDrop ND 1000.

Methods

In the first step of the SLIC reaction 1 µg of vector and insert were separately treated with 0.06 U of T4 DNA polymerase, supplemented with 2 µg BSA in 1x NEB2 buffer in a total volume of 20 µl. The reaction mixture was incubated at room temperature for 30 minutes and the exonuclease activity was then stopped by the addition of 2 µl dCTP (deoxyribo cytosine triphosphate, 10 mM) and chilling on ice for ten minutes. In the annealing step of the SLIC process varying insert to vector ratios, from 1:1 to 7:1, were used. In addition to the DNA, the annealing mixtures contained 0.1 ng RecA and 1x T4 Ligase buffer I adjusted to a total volume of 20 µl with ddH₂O. After incubation for 30 minutes at room temperature, every annealing mixture was transformed into DH5α cells as described in 3.1.4.1. Colonies were checked by colony PCR (3.1.6.4) to verify the cloning approach. Chosen positive clones were amplified (3.1.5) and verified by sequencing via Seqlab (Göttingen).

3.1.6.3 Site-directed mutagenesis

Site-directed mutagenesis is a technique to insert specific point mutations into the gene of interest. There are different reasons for the introduction of point mutations for example, the generation of less reactive or completely inactive enzymes. If the catalytic residues are unknown, it is possible to pinpoint them, by a systematic mutation analysis of likely important residues through the introduction of point mutations and the subsequent analysis of enzyme activity. A third application is the insertion of known disease-related mutations into a recombinant gene to investigate the influence on the molecular or structural level.

Site-directed mutagenesis in this work was performed with the help of PCR. In PCR a complementary primer pair is used for the insertion of the desired point mutation. In this primer pair the codon for a specific amino acid is exchanged against the codon of another, specifically chosen, amino acid. The variants that were generated in this work are based on two research articles (Eddouzi et al., 2013b; Schubert et al., 2011b) and were selected out of 15 disease related Mrr1 mutations. Those mutations were identified as gain of function mutations and were detected in clinical isolates of fluconazole resistant *C. albicans* strains (1.2.6). The K335N, P683S, G878E and G997V mutations were individually introduced into the Mrr1 '250' truncated construct. The primer pairs to generate these mutations are listed in **Table S 1** and the PCR was performed as described in section 3.1.1. After PCR, the reaction mixture was digested with DpnI, as described in 3.1.6.2, to remove the template DNA and thus to reduce the number of colonies which would potentially contain the original construct. Following DpnI digestion the PCR products were purified with the NucleoSpin gel and PCR Clean-up Kit (Macherey-Nagel) and subsequently the complete Clean-up eluate was transformed into DH5α cells (3.1.4.1). Selected colonies were used to inoculate 5 ml LB cultures for DNA plasmid amplification (3.1.5) and the presence of the mutations was verified via sequencing analysis (Seqlab, Göttingen).

Methods

3.1.6.4 Colony PCR – *E. coli*

Colony PCR is used as an initial method to identify bacterial colonies which contain the insert of interest after a cloning experiment. With the help of a distinct primer pair, DNA fragments of a defined length can be amplified. One primer is annealing at the DNA upstream of the gene of interest and the second primer either somewhere in the middle of the gene of interest or downstream of it. The following analysis of the PCR by DNA gel electrophoresis permits the visualization of the lengths of the amplified DNA and thereby permits the comparison with expected theoretical DNA lengths for the chosen primer pair. The supposable positive clones are finally verified by sequencing.

Clones, chosen for colony PCR, were streaked into the PCR mixture described in 3.1.1 and then the pipette tip was additionally restreaked onto an LB plate to maintain the probably positive clone for further verification. Colonies, identified as positive in the colony PCR experiment were used for plasmid DNA amplification (3.1.5) and the isolated DNA was subsequently further analyzed by sequencing.

3.1.6.5 Colony PCR – *P. pastoris*

Colony PCR for yeast colonies was similarly performed as described above but the colonies required a specific treatment in advance to the PCR (**Table 19**, *P. pastoris*). The colony was picked with a sterile pipette tip and thoroughly scraped into 75 µl of sterile 1x TE buffer in a PCR tube. To break the cell wall, the reaction mixture was applied to five microwave cycles at the highest voltage for 3.5, 2, 1.5, 1 and 0.5 minutes followed by harsh vortexing in between the cycles. The samples were subsequently frozen in liquid nitrogen for 10 minutes or overnight at -80 °C and thawed at 95 °C for 2 minutes. The mixture was then centrifuged (2,000 g) for 15 minutes, at room temperature and 5 µl of the supernatant were used for the PCR 3.1.1. Colonies that were identified to contain the gene of interest were chosen for subsequent expression tests.

3.1.7 Protein expression

Due to expression properties and insolubility issues three different expression organisms were used during this work, *E. coli*, *M. smegmatis* and *P. pastoris*.

3.1.7.1 *E. coli*

E. coli was used as expression host for different constructs of the Mrr1 protein (**4 to 6** and **16 to 20** **Table 8**). In advance to a large scale expression of the according protein, expression tests with different *E. coli* strains were performed in a 100 ml scale rather than the 2 l scale as described below for the regular cultures.

For recombinant protein expression the expression vectors (**Table 8**) were transformed into the according *E. coli* expression strain (**Table 7**) as described in 3.1.4.1. Mrr1 FL was expressed in BL21 (DE3), BL21 (DE3) arctic express or Rosetta (DE3) cells. For the expression of Mrr1 AR or Mrr1 '250' BL21(DE3)-pLysS cells were used. LB agar plates and liquid medium were supplemented with the respective antibiotics (**Table 10**) for the applied strain-plasmid combination to prevent contamination with other bacteria and to maintain the selection pressure.

200 ml LB liquid medium were inoculated with one of the colonies obtained after transformation and the cells were grown at 200 rpm and 37 °C overnight. 25 ml, or 30 ml respectively (BL21(DE3)-pLysS cells), of the preculture were used to inoculate 2 l LB liquid medium in 5 l Erlenmeyer flasks for a large scale expression. The cultures were grown at 37 °C until the exponential growth phase, with an OD₆₀₀ of 0.6, was reached. Protein expression was induced by addition of 0.06% (v/v) L(+)-arabinose, for pBadm11-vectors, or 1 mM IPTG, for the pETm11-vector. The expression was conducted for 7 h at 30 °C (BL21(DE3)-pLysS cells) or overnight at 16 °C (BL21 (DE3) or Rosetta (DE3)). An exception were the BL21 (DE3) arctic express cells where in addition to the gene of interest a bacterial chaperone was co-expressed to increase solubility. Induction was achieved by the addition of 0.06% (v/v) L(+)-arabinose and 1 mM IPTG after the culture was cooled down to 11 °C. Protein expression was conducted for 22 h at 11 °C. Bacteria were harvested by centrifugation, at 4 °C, 5,000 g for 15 minutes, and stored in aliquots of 20 – 30 g at -80 °C.

3.1.7.2 *M. smegmatis*

To prevent contamination of the *M. smegmatis* cultures with faster growing bacteria or yeast, inoculation of the preculture and the main culture were performed in a biological safety cabinet (Hera Safe KS12, Kendro) with sterile equipment. All FadA5 constructs were expressed in *M.*

Methods

smegmatis similar as described by Nesbitt et al.. 7H9 and 7H10 media were supplemented with 200 µg/ml ampicillin, hygromycin (50 µg/ml) and cycloheximid (10 µg/ml), to enable the selection of the gene of interest and to prevent contamination, as well as 0.5% (v/v) glycerol. For recombinant protein expression the pSD31 FadA5 vectors (Table 8) were transformed into the *M. smegmatis* mc²155 expression strain (Table 7) by electroporation (3.1.4.2). A single colony selected on 7H10 solid medium was used to inoculate the 100 ml 7H9 liquid medium preculture. The preculture was grown for 3 – 4 days at 37 °C and 200 rpm until it was a turbid, flaky and dense cell suspension. 2 l 7H9 liquid medium in a 5 l Erlenmeyer flask were inoculated with 20 ml of the thoroughly resuspended preculture and were grown for around 30 h at 37 °C while shaking (200 rpm). When the culture was dense and flaky, protein expression was induced by the addition of 0.2% (w/v) acetamide. The culture was grown for another 40 h at 37 °C while shaking (200 rpm) and subsequently harvested by 30 minutes centrifugation at 6,600 g at 4 °C. The cell pellet was stored in aliquots of 30 – 40 g at -80 °C.

3.1.7.3 *P. pastoris*

The Mrr1 constructs were also cloned into the PichiaPink expression vectors, pPink α -HC, pPink-HC and pPink-LC, as expression of the Mrr1 FL and Mrr1 AR in *E. coli* did not yield protein suitable for biochemical investigation or structural characterization. The yeast *P. pastoris* is a fungus and is as the Mrr1 organism of origin, *Candida albicans*, an eukaryote. Therefore increased soluble protein expression and possible post-translation modifications might increase the amount of functional protein that could subsequently be used for its characterization.

The following adapted protocol (Invitrogen) was applied to pursue the expression of the different constructs (Table 8) in the used expression strains (Table 7). All culture tubes, flasks and the lids as well as the pipette tips, were autoclaved before use. To allow a maximum of aeration either loose fitting metal caps or silicone caps, with a special pore area (silicone caps Table 4), were employed.

For recombinant protein expression the PichiaPink vectors (Table 8) were genetically integrated into the *P. pastoris* expression strains (Table 7) as described in 3.1.4.3. In addition to the tested colonies, negative controls, growth medium only controls and pink colonies or not induced cultures were grown in parallel, to monitor eventual contamination or to determine the intrinsic protein expression background. One colony was thoroughly brought into 5 ml of BMGY medium (Table 11), in a 15 ml sterile culture tube, by rubbing the colony, picked with an inoculation loop, at the glass housing of the tube into the liquid. This preculture was grown for 15 to 17 h while shaking (200 rpm) at 30 °C until the OD₆₀₀ of the cell suspension was between 1.5 and 2. The cells were harvested in 15 ml

Methods

conical Greiner tubes at 1,500 g for 5 minutes at room temperature. The supernatant was discarded and the cells were resuspended in 2 ml of BMMY medium (Table 11) to induce protein expression with 0.5% (v/v) methanol which is present in the BMMY medium. The cells were either grown in 15 ml culture tubes or transferred to 100 ml Erlenmeyer flasks with 20 ml of BMMY medium. Growth was continued for another 72 or 96 h at 30 °C while shaking (200 rpm). 100 µl (2 ml culture) or 1 ml (20 ml culture) samples were extracted at different time points for subsequent analysis of the protein expression pattern. The test samples were centrifuged at 16,100 g for 5 minutes at room temperature. Cells and supernatant, for secreted pPink α -HC expression only, were separately shock frozen in liquid nitrogen and stored at -80 °C until the samples were analyzed by SDS-PAGE. The cultures were supplemented with 2% (v/v) methanol 24 h after induction. Whereas the 20 ml culture remained unaltered after the second induction, the 2 ml culture had to be split into two culture tubes and was supplemented with 1 ml BMY medium to prevent culture growth to too high density. The cells were finally harvested after 72 or 96 h by centrifugation (16,100 g, 5 minutes, room temperature), frozen in liquid nitrogen and stored at -80 °C.

3.2 Protein purification and characterization

3.2.1 Protein purification

To succeed in crystallization it is very important to obtain protein samples of the highest purity and homogeneity. The purity and stability can be verified by methods such as SDS-PAGE gel electrophoresis, dynamic light scattering and multiangle laser light scattering. (Chayen and Saridakis, 2008; Zulauf and D'Arcy, 1992) The purification of the proteins in this work was achieved in two, three or four chromatography steps.

FadA5 was purified, similarly to an already established protocol (Schaefer, 2010), in a three-step purification by immobilized metal affinity chromatography, anion-exchange chromatography and size-exclusion chromatography. The initial step via metal affinity chromatography facilitated the isolation of the protein from the cell lysate. The second step separated the protein of interest from similar sized impurities and the final, polishing step removed remaining contaminants and aggregated protein. The His₆-tag was not removed during the purification protocol. All steps were performed at 4 °C if not stated otherwise.

The Mrr1 variants were purified in two, three or four chromatography steps by immobilized metal affinity chromatography and size-exclusion chromatography or in addition an anion-exchange chromatography step. As neither expression nor purification protocols for the variants were available, different affinity resins and buffer combinations were tested. In addition, in some cases the

Methods

His₆-tag was removed by cleavage with the TEV (Tobacco Etch Virus) protease followed by a reverse-Ni-NTA chromatography step. All steps were performed at 4 °C if not stated otherwise.

3.2.1.1 Cell lysis

3.2.1.1.1 *FadA5* from *M. smegmatis*

For cell lysis, around 30 g of cell pellet were thawed on ice and resuspended in 250 ml binding buffer (1, Table 12) to which 1.5 ml lysozyme (50 mg/ml), 6 µl DNase and 30 mg PMSF were added. The cell suspension was subdivided in eight aliquots (30 ml) and was applied to three lysis cycles with the French Press. After the first lysis cycle the lysate was again supplemented with 1 ml lysozyme (50 mg/ml). To clear the lysate and remove cell debris, the lysate was centrifuged for 1 h at 4 °C (33,000 g, Avanti J-26 XP Beckmann Coulter centrifuge with rotor JLA-16.250). The supernatant was subsequently applied to metal-affinity chromatography (3.2.1.2).

3.2.1.1.2 *Mrr1* variants from *E. coli*

For small cultures in which the expression of the target protein was analyzed, the cell pellets were thawed on ice, supplemented with 1 ml binding buffer (6, Table 12) and applied to cell lysis by sonication (Labsonic P, B. Braun). The cells were lysed on ice by two cycles (amplitude: 90%, cycle: 0.9) of 15 s with a break of at least two minutes in between. The lysate was cleared by centrifugation at 4 °C for 10 minutes (16,100 g, 5415R centrifuge, Eppendorf) and was subsequently analyzed by metal-affinity chromatography (3.2.1.2).

15 g pellet of the *Mrr1* Full length (FL) construct were resuspended on ice with binding buffer (6, Table 12), supplemented with 3 µl DNase (Invitrogen) and one tablet of the cComplete – EDTA-free protease inhibitor mix (Roche). The cells were lysed by two passages through the cell disruptor (M-110P, Microfluidics) at 1.3 kbar. The lysate was supplemented with 15 mg PMSF and was cleared by centrifugation for 1 h at 4 °C (45,000 g, Avanti J-26 XP Beckmann Coulter centrifuge with rotor JA-25.50) and was subsequently applied to metal-affinity chromatography (3.2.1.2).

Due to differences in the expression strain (BL21(DE3)-pLysS), the cell lysis for the *Mrr1* Armadillo repeat (AR) and '250' was performed differently. The pLysS-cells are intrinsically producing lysozyme and are difficult to resuspend with binding buffer (6, 22, Table 12) as they appear slimy and viscous in the resuspension solution. Direct cell lysis with the cell disruptor was therefore not possible and homogenizing cycles with the sonicator had to be performed in advance. Before lysis, the cells, 15 g from the AR expression and 25 to 30 g from the '250' expression, were resuspended on ice and were supplemented with 20 or 40 µl DNase (Invitrogen) and one tablet cComplete – EDTA-free protease inhibitor mix (Roche). The cells were lysed, on ice, by two cycles (amplitude: 90%, cycle: 0.9) of 2 to

Methods

2.5 minutes with a break of at least two minutes in between. After this initial homogenization step the cell suspension was twice passed through the cell disruptor (M-110P, Microfluidics) at 1.3 kbar. After addition of PMSF, the lysate was cleared by centrifugation for 1 h at 4 °C (45,000 g, Avanti J-26 XP Beckmann Coulter centrifuge with rotor JA-25.50) and was subsequently applied to metal-affinity chromatography (3.2.1.2).

3.2.1.1.3 *Mrr1* variants in *P. pastoris*

The cell lysis for the expression tests of all Mrr1 constructs in *P. pastoris* was performed in the same way. The frozen cell pellets were quickly thawed at room temperature, then placed on ice and supplemented with 200 µl of breaking buffer (38, Table 12). An equal volume of glass beads (0.5 mm) was added to the suspension and the mixture was applied to eight vortex cycles (highest intensity, vortex genie 2, scientific industries) of 30 s with a break of 30 s on ice in between. The lysate was cleared from cell debris by centrifugation at 4 °C for 10 minutes (16,100 g, 5415R centrifuge, Eppendorf) and subsequently analyzed by metal-affinity chromatography (3.2.1.2).

3.2.1.2 Metal-affinity chromatography

The basis of all affinity chromatography methods is either the interaction of intrinsic protein regions or of customized tags connected to the protein, which can interact with immobilized metal ions or specific ligands that are fused to beads. DNA-binding proteins can for example naturally interact with an immobilized heparin matrix, which is imitating the polyanionic structure of the nucleic acid. In addition, a lot of different affinity tags are known, the most common one is the hexahistidine tag (His₆-tag), but tags as the glutathione-S-transferase (GST-tag) or streptavidin tag (Strep-tag) are also

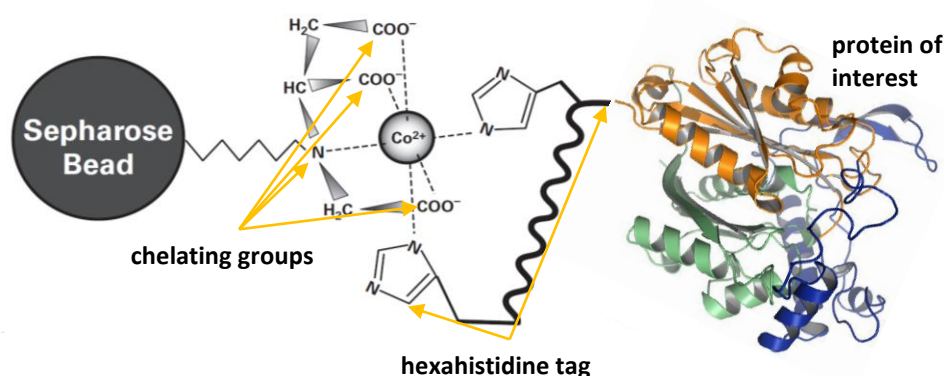


Figure 15 - Protein-binding to immobilized Co²⁺ ions (schematic illustration):

The protein of interest is captured by the interaction of the free coordination sites of the ion with the hexahistidine tag; adapted from (Clontech, 2009; Roy et al., 2010; Sanglard, 2002a)

Methods

frequently applied. In this work, affinity chromatography was solely performed with a His₆-tag but different interacting media were analyzed. For the affinity medium Ni²⁺ or Co²⁺ ions were immobilized by a chelator to silica, sepharose or agarose beads (**Figure 15**). Iminodiacetic acid (IDA), nitrilotriacetic acid (NTA), tris(carboxymethyl)ethylene diamine (TED) or 2-(carboxymethyl amino) succinic acid (Co-Talon) are used as chelating agents and the resin is named accordingly. The chelating groups occupy four to five of the six coordination sites provided by the divalent cations and the remaining binding site(s) can interact with the His₆-tag of the protein of interest. Through this interaction the protein of interest is captured and separated from the endogenous proteins of the expression organism. Unspecific protein binding can occur and is dependent on the chelating group but the addition of 5 to 10 mM imidazole to the binding buffer reduces this unspecific background. The above mentioned resins are used in self-packed columns with gravity flow. In addition pre-packed affinity columns from GE Healthcare (HisTrap FF crude) or Novagen (Ni-MAC) in combination with an FPLC (Fast protein liquid chromatography) system were used. (2006; 2008)

Table 20 - Established metal-affinity chromatography procedures

Protein	<i>FadA5</i>	<i>Mrr1 FL (AR)</i>	<i>Mrr1 '250'</i>
Resin	Co-Talon	Ni-TED	Ni-NTA
Column volume	1 ml / 10 g pellet	1 ml / 5 g pellet	1 ml / 15 g pellet
Buffer combination	1, 2 (Table 12)	6, 7, 9 (Table 12)	22, 23, 24 (Table 12)
Column equilibration	5 CV binding buffer		
Sample application	Load once over resin (1.5 h)	Load twice over resin (4 h)	Load overnight over resin
Wash unbound sample	10 CV binding buffer		
Wash steps	W I 2.5 CV buffer 1 W II 1.5 CV buffer 1	W I 2 CV buffer 7 I W II 2 CV buffer 7 II W III 2 CV buffer 7 III W IV 2 CV buffer 7 IV W V 2 CV buffer 7 V W VI 2 CV buffer 7 VI	W I 2 CV buffer 23 I W II 4 CV buffer 23 II W III 4 CV buffer 23 III W IV 2 CV buffer 23 IV W V 2 CV buffer 23 IV W VI 2 CV buffer 23 V
Elution steps	E I 1 CV buffer 2 E II 2 CV buffer 2 (incubation) E III 1 CV buffer 2 E IV 1 CV buffer 2 E V 1 CV buffer 2 E VI 6 CV buffer 2	E I 2 CV buffer 9 I (incubation) E II 4 CV buffer 9 I E III 2 CV buffer 9 I E IV 4 CV buffer 9 II E V 4 CV buffer 9 II	E I 2 CV buffer 24 I (incubation) E II 4 CV buffer 24 I E III 2 CV buffer 24 I E IV 4 CV buffer 24 II E V 8 CV buffer 24 II

In the following only the established purification procedures for the proteins are described (**Table 20**). Prior to the experiment, the required amount of loose resin was filled into Econo chromatography columns (Bio-Rad) and was equilibrated with the according binding buffer. For the *Mrr1* Armadillo repeat (AR) construct no established protocol exists, the test purifications were performed with all available loose resins according to the protocol described for Full length (FL) *Mrr1* with Ni-TED resin. The cleared lysates were loaded onto the respective column and subsequently the unbound proteins were eluted with 10 column volumes (CV) of the according binding buffer. Several washing steps with buffers containing increasing imidazole concentrations were performed to

Methods

remove unspecifically bound protein. After incubation times of 10 to 15 minutes with elution buffer, the protein of interest was eluted in several elution steps in a total volume of 16 to 20 CV. The fraction size was dependent on the pellet mass used which also determined the column volume and therefore differed throughout all purifications. All fractions, including the cell pellet, were analyzed by SDS-PAGE gel electrophoresis (3.2.2.1).

To validate protein expression tests, a shortened fast-track metal-affinity chromatography was performed. For this purpose the cleared lysate or culture supernatant (secreting pichia expression) was mixed with 50 to 100 μ l of Ni-NTA resin in buffer **6** or **39** (+ 5 mM imidazole) and incubated for 2 h at 4 °C while slightly shaking. The mixture was centrifuged (12,500 g, 5415R centrifuge, Eppendorf) for 10 minutes at 4 °C, the supernatant was discarded and 500 μ l of fresh buffer **6/39** was added. The suspension was centrifuged a second time, the supernatant was separated and the resin was again resuspended with 40 μ l of fresh buffer **6/39**. The initial cell pellet was resuspended in 8 M urea and a sample was taken for SDS-PAGE analysis. In addition, samples from the 500 μ l and 40 μ l resuspension were prepared for analysis. The expression test was validated by SDS-PAGE gel electrophoresis (3.2.2.1).

3.2.1.3 Anion-exchange chromatography

The principle of ion-exchange chromatography is the separation of proteins dependent on their net charge. For this purpose the isoelectric point (pI), i.e. the specific pH value where the net charge is zero, of proteins is important. Is the pH value of the protein's buffer below the pI, then the protein is positively charged and vice versa. To apply separation by ion-exchange chromatography a buffer with a $\text{pH} = \text{pI} \pm 1.5$ is chosen. To prevent problems that might occur when the protein is exposed to extreme pH values, the respective charged ligands, anion- or cations, are chosen accordingly. As the theoretical pI values of all proteins in this work was around 6, anion-exchange chromatography was chosen, since pH values of 7.5 and higher seemed reasonable. In anion-exchange chromatography, negatively charged proteins are interacting with the positively charged column resin, where positively charged ligands are coupled to cellulose or agarose beads. Through this interaction the protein of interest is separated from positively charged or uncharged impurities. In addition, a separation of the negatively charged proteins is achieved, as the binding strength is dependent on their respective pI. The proteins are eluted by a salt gradient, from low salt to high salt. The salt ions, which interact more strongly with the matrix thereby displace the bound proteins. More strongly bound proteins elute later, than weakly bound ones. (Berg et al., 2007)

Methods

In this work, anion-exchange chromatography was performed with columns of the MonoQ series (GE Healthcare) utilizing different Äkta FPLC systems. The protein sample, diluted with low salt buffer, was applied with the Äkta system's sample pump. For the Mrr1 '250' protein variants different gradient profiles were tested and the established profiles are listed, together with the buffers in **Table 21**. Protein elution was monitored by the absorption curves at 280 nm and 260 nm. After the chromatography step, the quality of the purification was analyzed by SDS-PAGE gel electrophoresis (3.2.2.1) and the samples were concentrated for the following purification steps.

Table 21 - Anion-exchange chromatography protocols

Protein	<i>FadA5</i>	<i>Mrr1 '250'</i> <i>N-terminal His₆-tag</i>	<i>Mrr1 '250'</i> <i>C-terminal His₆-tag</i>
pl theoretic *	5.78	6.17	6.23
experimental **	6.0	-	-
Column	MonoQ 10/100 GL (GE Healthcare)		MonoQ 5/50 GL (GE Healthcare)
Column cleaning before the run	2 CV 2 M NaCl, 4 CV 1 M NaOH, 2 CV 2 M NaCl, 2 CV ddH ₂ O		
Column volume	8 ml		1 ml
Buffer combination	3, 4 (Table 12)		29, 30 (Table 12)
Column equilibration	3 CV buffer 3		5 CV buffer 29 with 1.5% 30 (= 65 mM KCl)
Sample application	Load with FPLC (Äkta systems)		
Gradient	Linear and step gradients	Linear and step gradients	Linear gradients
Wash unbound sample	8 CV buffer 3	10 CV buffer 29 with 1.5% 30 (= 65 mM KCl)	12 CV buffer 29 with 1.5% 30 (= 65 mM KCl)
Fraction size	2 ml		1 ml
Gradient 1	3 CV (0 - 22% buffer 4)	3 CV (1.5 - 22% buffer 30)	3 CV (1.5 - 15% buffer 30)
Step	5 CV (22% buffer 4)	4 CV (22% buffer 30)	--
Gradient 2	15 CV (22 - 40% buffer 4)	10 CV (22 - 45% buffer 30)	30 CV (22 - 45% buffer 30)
Gradient 3	6 CV (40 - 100% buffer 4)	6 CV (40 - 100% buffer 30)	10 CV (40 - 100% buffer 30)
* theoretic pI calculated with ProtParam ExPasy tool (Gasteiger et al., 2003)			
** experimental pI calculated with PROPKA (Li et al., 2005; Olsson et al., 2011) on the basis of the apo structure			

3.2.1.4 TEV-cleavage reaction

The gene constructs in the pBadm11 vector contained a TEV cleavage site, which permits the removal of the His₆-tag with the help of the TEV protease.

For the cleavage reaction the protein of interest was concentrated to volumes suitable for dialysis and the absorption of the protein at 280 nm was determined spectrophotometrically. This absorption value at 280 nm is important, as the TEV protease is added in a ratio 1:100 of the absorption value at 280 nm. After the addition of the protease in the correct ratio, the protein-

Methods

protease sample is filled into a dialysis cassette and applied to overnight dialysis at 4 °C in TEV buffer (39, Table 12). To separate the cleaved protein from the un-cleaved protein a reverse-Ni-NTA purification was performed. For this purpose, the protein sample was diluted 1:9 in binding buffer 22 and was applied to equilibrated Ni-NTA resin. The flow through was collected, then the resin was washed with 2 CV of wash buffer 23 I and IV and remaining proteins were eluted with 2 CV elution buffer 24. All fractions were analyzed by SDS-PAGE to determine the success of the cleavage reaction. The cleaved protein should be found in the flow through fraction, as it cannot bind to the resin any longer. Fractions containing the protein of interest were pooled and concentrated for the final purification step.

3.2.1.5 Size-exclusion chromatography

As the name of this technique already implies, the separation process is dependent on the protein size. The highly cross-linked agarose matrix with covalently bound dextran (\varnothing pore diameter 34 μm) provides a separation basis for proteins of different sizes. Larger sized proteins will not intrude into the matrix pores whereas medium or small sized proteins will. The smaller the proteins are, the longer they need to pass through the column material and vice versa for large proteins.

The elution profiles of the size-exclusion chromatography, monitored at 280 and 260 nm, are an important indicator for the quality of the protein. The shape of the peaks and the elution volume provide insights into the homogeneity of the protein or can indicate protein aggregation. This information is important for all subsequent experiments as, for example, aggregated proteins most likely do not represent the correct functional state of the protein.

In this work, several size-exclusion columns from GE Healthcare were used and are listed in Table 3. In preparation to the chromatography step, the columns were equilibrated with at least one column volume (CV) of the according size-exclusion buffer (Table 12) and the protein sample was concentrated to yield 5% or less of the column volume. The sample was centrifuged for at least 10 minutes at 16,100 g and 4°C to remove precipitated proteins or dust particles. The sample was applied to the column with a capillary loop in the according size. Isocratic elution over 1 to 1.3 CV was performed with flow rates of 1 ml/min (HiLoad columns) or 0.5 ml/min (analytical columns) in the respective buffer and fractions were collected. The chromatography step was validated by SDS-PAGE and the fractions containing pure protein were



Figure 16 - Principle of size exclusion chromatography: Large proteins (red) do not intrude into the matrix, whereas smaller (yellow) and middle (orange) sized proteins can enter the pores and thereby elute later. (GEHealthcare)

Methods

pooled and concentrated. The protein concentration was measured spectrophotometrically with the NanoDrop at 280 nm and the protein was flash frozen and stored at -80 °C for further use.

3.2.2 Protein characterization

3.2.2.1 SDS-PAGE gel electrophoresis

The sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is a method that is commonly used to validate protein purity and amounts after each purification step or to verify protein (over)expression in an expression test. (Berg et al., 2007) After the actual SDS-PAGE is performed, the protein bands are stained with coomassie blue staining solution.

In this work only polyacrylamide gels with a 5% stacking gel and a 15% separating gel were used. The composition of the gels, as well as the used sample and running buffers and solutions used for gel staining and destaining are listed in **Table 13**.

12 µl of the protein sample were mixed with 3 µl of 5x sample buffer and the mixture was denatured for 5 minutes at 95 °C. 10 µl of the sample were loaded on the polyacrylamide gel; 2.5 µl protein marker was loaded as well for subsequent size determination. Electrophoresis was performed at 250 V for 45 minutes in SDS-PAGE running buffer.

Staining was performed by two different methods. For R250 coomassie staining, the gel was incubated for 15 minutes in the staining solution, washed with destaining solution and destained (approx. 30 minutes) until the background was clear but the protein bands were still visible. The gel was then transferred into water.

By staining with the G250 safety stain, the gel initially was boiled three times in water. This process removes the SDS completely from the gel and the gel was subsequently stained with G250 coomassie safe stain, by boiling the gel in the staining solution. Subsequently the gel was incubated for another 10 minutes in the staining solution until the protein bands were readily visible. The gel was then transferred into water.

3.2.2.2 Western Blotting

Western Blotting uses specific antibodies to detect and verify smaller amounts of protein, which are not detectable by coomassie staining. The basis of this technique is the utilization of two antibodies, a primary antibody, specifically interacting with the protein of interest and the secondary antibody that binds to the primary antibody and is fluorescently labeled. If no antibody against the gene of

Methods

interest is available, primary antibodies against protein affinity tags are commonly used. (Berg et al., 2007)

The initial step of Western Blotting was the SDS-PAGE without the staining and destaining processes. After gel electrophoresis was performed at 200 V for 60 minutes, the gel was transferred to the blotting membrane in the blotting chamber. This was a process that had to be performed cautiously, as the correct orientation and assembly had to be arranged to allow bubble free transfer of the proteins from the gel onto the nitro cellulose membrane and to prevent transfer into the wrong direction. The blot was assembled in blotting buffer (Table 14) with pre-hydration of the membranes. The blot was composed of a rough filter, a whatman filter paper, the gel, the nitrocellulose membrane, another whatman filter and a rough filter, in direction from the anode to the cathode. Blotting was performed in blotting buffer for 75 minutes at 300 mA on ice.

All used buffers, solutions or antibodies are listed in Table 14. The membrane was cautiously removed, slightly dried and stained for approx. one minute with ponceau red solution. Spare staining solution was removed by rinsing the membrane with distilled water and then the (bubble free) transfer from the gel to the membrane was checked. When the transfer was successful, the blot was blocked with blocking solution for 1 h at room temperature while shaking and followed by the addition of the primary antibody (1:2000; α His) and incubation overnight at 4 °C under constant rotation. The solution was then decanted and the blot was washed three times (10, 20, 30 minutes) with TBST buffer at 4 °C while shaking. Subsequently the blot was incubated with the secondary antibody (1:10000; HRP) for 1 h at 4 °C while rotating, followed by three washing cycles of 30 minutes with TBST buffer at 4 °C while shaking.

The membrane was then thoroughly removed, slightly dried and the ECL solution was applied to enable luminescence. The luminescence was detected with the Electronic UV-Transilluminator (Ultra Lum) at different time scales to permit an optimum signal to noise ratio. In advance, the position of the protein size marker was recorded to enable the identification of the signals at the correct protein size.

3.2.2.3 Thermofluor assay

The thermofluor assay is used to analyze protein stability dependent on the buffer, the pH, additives or cofactors. Optimized or increased protein stability can be inevitable during the purification process, in the crystallization experiment or for functional assays. The high throughput set up was described by Ericsson et al. in which the temperature dependent denaturation of the protein of interest is monitored in the presence of different buffer combinations and additives. (Ericsson et al.,

Methods

2006) In a real-time PCR cycler (here Mx3005p, Stratagene) the temperature of the samples is constantly increased. This increase leads to the denaturation of the protein sample and thus to the unfolding of the macromolecule. Since hydrophobic residues are buried in the core of soluble proteins when they are correctly folded, they cannot interact with the fluorescence dye Sypro Orange (Invitrogen) when they are still intact. In an aqueous surrounding the dye's fluorescence is quenched, but in the increasing hydrophobic environment of the unfolding protein the fluorescence emission at 575 nm is increasing. During the thermofluor experiment the unfolding of the protein is thus measured by the fluorescence increase dependent on the increase in temperature. In principle it can be assumed that the more stable a protein is, the higher is its melting temperature T_m . The T_m can be determined at the inflection point of the fluorescence curve.

In this work, thermofluor assays were performed for the two Mrr1 variants Full length and '250'. The optimum buffer condition for FadA5 was already determined in a previous work (Schaefer, 2010). Initially the standard thermofluor buffer screen (**Table S 3**) was used and based on these results individual buffer screens, further analyzing different variables were designed. These screens also contained additives from the crystallization Additive and Silver Bullets screens (**Table 15**). 1 μ l protein (1.2 to 2.7 mg/ml) was mixed with 1 μ l 2.5% (v/v) Sypro Orange and a buffer solution with a chosen additive at different concentrations. The reaction mixture was adjusted to 20 or 25 μ l through the addition of the tested buffer. Reference and control reactions were also set up that contained the protein in the current buffer, in water or were free of protein. The mixtures, in a 96-well PCR plate, were heated from 25 °C to 95 °C in steps of 1 °C per minute.

3.2.2.4 Circular-dichroism (CD) spectroscopy

CD spectroscopy is a valuable tool to gain initial information on the secondary structure of the protein of interest in a comparably short time. It measures differences in the absorption, of peptide bonds, between right-handed and left-handed polarized light, which occurs due to structural asymmetry in the far ultraviolet (UV) light region between 170 to 250 nm. As the different secondary structure motifs, as α -helices or β -strands as well as disordered or poly-proline regions, have different positive and negative absorption patterns in the far UV region they all contribute to the CD spectrum. A CD spectrum does not provide information about which motifs are located where in the protein but they provide information about the general contribution of all the motifs in the protein. With the help of CD spectroscopy not only information about the secondary structure of a protein is obtained but also changes in its structure by the introduction of mutations or deletions are easily monitored. (Dodero et al., 2011)

Methods

For CD measurements the proteins were dialyzed into a special buffer (40), containing sodium fluoride instead of potassium chloride (Table 12). The proteins were used in initial concentrations of 2 mg/ml, but were further diluted to ensure that the detector voltage would not exceed 500 V during the experiment. Due to the dilution issue the applied protein concentrations varied from 0.15 to 0.5 mg/ml.

3.2.2.5 Dynamic light scattering (DLS)

Dynamic light scattering (DLS) is a spectroscopic method, with laser light, at wavelengths between 400 and 800 nm, for the determination of the particle size distribution in solution. As all molecules are in motion at temperatures above 0 K (Brownian motion), those movements and their velocity can be monitored on the basis of the light scattering they cause. The velocity of the movements is dependent on the particle size and therefore larger particles show a different scattering behavior than smaller particles. With the assumption that all particles are of globular shape, which leads to a bias of the results, the molecular weight of the particles can be estimated. This easy to handle and fast method permits the determination of a protein sample's polydispersity to validate its applicability for crystallization. In addition information about the protein's oligomeric state is gained and the amount of protein aggregation can be documented. (Banachowicz, 2006; Bergfors, 1999)

Prior to the experiment, the protein sample was centrifuged for 30 minutes at 16,100 g and 4 °C to remove aggregated protein or dust particles. 40 µl of a protein sample (> 0.6 mg/ml) were pipetted into a quartz cuvette (Table 4) and was measured several times, for different time lengths, with the DynaPro Titan DLS instrument (Wyatt Technology). The method parameters are listed in Table 22. After the experiment the data were evaluated with the Dynamics V6 software (Wyatt Technology).

Table 22 - DLS parameter setting

Solvent	Water	Acquisition time	10 s / 20 s
Laser Power	25%	Number of acquisitions	10 / 20
Sample	globular	Temperature	4 °C / 20 °C

3.2.2.6 Size-exclusion chromatography coupled multiangle laser light scattering (SEC-MALLS)

Multiangle laser light scattering (MALLS) is a so called absolute technique, where the molecular mass of protein samples can be determined without any molar mass standard or other calibration methods. As the name of the method already indicates, the particles' light scattering has to be measured from different, distinct angles relative to the exciting laser beam. This is necessary as proteins are rarely particles with a globular shape. The intensity of light that is scattered by a

Methods

molecule is directly proportional to the molar mass and together with the specific refractive index increment, also detected during the experiment, it allows the calculation of the molar mass with an accuracy of around 5%. (Erinc Sahin, 2012; Technology, 2014a, b; Wyatt, 1993)

MALLS can be coupled or measured in-line with a size-exclusion chromatography step, so that the protein sample is first applied to chromatography, thereby undergoes separation by mass and is then directly forwarded into the light scattering detector (DAWN hELEOS 8⁺), which is coupled to the refractive index (RI) detector (Optilab T-rEX).

In practice, 200 μ l of the centrifuged protein sample (0.9 to 5 mg/ml) was applied to a Superdex 200 10/300 GL column (GE Healthcare) which was run with a flow rate of 0.5 ml/min by an ÄktaPurifier system. Prior to applying the sample, the column was extensively equilibrated with thoroughly filtered, freshly made size-exclusion buffer (5, 37, Table 12) stored in extensively flushed glass bottles. These preparations were necessary to minimize the scattering background of dust or other particles in the buffer to a minimum. The scattering noise should vary just at the 4th decimal place. As the MALLS and RI detectors were coupled in-line to the Äkta system, the sample was directly applied to those detectors after it came through the UV detector of the Äkta system. The results of the light scattering experiment were analyzed by the ASTRA 6 software, supplied with the instrument.

3.2.2.7 Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry is a method to determine the binding profile between two interaction partners, for example protein-protein or protein-ligand interactions. In a single ITC experiment, the enthalpy (ΔH) of binding, the binding constant (K_d) and the stoichiometry (N) of the reaction can be deduced from a heat difference. This difference is measured between a reference cell, containing only buffer and a sample cell containing protein in the same buffer, to which the other protein or the ligand are titrated. The free energy binding constants (ΔG) and the entropy (ΔS) of binding can be calculated from the binding constant. (Doyle, 1997; Pierce et al., 1999)

ITC experiments were performed with the FadA5 wild type protein and the FadA5 variant C93A. The interacting ligands were the products of the thiolase reaction, 3-oxo-pregn-4-ene-20-carboxyl-CoA (3-OPC-CoA, compound 2 Figure 10, Table 16) and Coenzyme A. For the ITC experiment a buffer with a small enthalpy of ionization was chosen, as buffers with high enthalpies might disturb the measurement but for buffers with low enthalpies this effect can be neglected. The enzyme was therefore extensively dialyzed against the ITC buffer (50 mM MES pH 6.7, 250 mM NaCl, 41 Table 12) overnight at 4 °C. Reference buffer, protein sample and ligand were degassed prior to the experiment. The experiments were performed with the VP-ITC instrument (MicroCal, Malvern

Methods

Instruments Ltd, Malvern, UK) and the according instrument and experimental settings are listed in **Table 23**. Buffer-ligand to buffer-ligand titrations were performed, to subtract the heat production in this process from the actual protein-ligand experiment prior to curve fitting. For curve fitting, the first injection was deleted, a reaction process in first order was assumed and the values for the apparent K_d , N , ΔH and ΔS were calculated with the ORIGIN 7 software. Measurements for the WT and C93A variant were performed several times.

Table 23 - Experimental parameters for the ITC experiment

Total number of injections	21	Injection Parameters	
Temperature	25 °C	Injection 1	
Reference Power	12 μ Cal/s	Volume	4 μ l
Delay time	120 s	Duration	5.7 s
Syringe concentration	0.420 mM - 0.450 mM	Spacing	320 s
Cell concentration	0.022 mM – 0.026 mM	Filter period	2 s
Stirring speed	260 / s	Injections 2 - 21	
Feedback mode	high	Volume	15.5 μ l
ITC Equilibration	Fast / Auto	Duration	22.1 s
		Spacing	840 s
		Filter period	2 s

3.3 Protein crystallization

The main goal of this work was the structural characterization of the mycobacterial protein FadA5, in the presence of its ligands, and the structural characterization of the fungal transcription factor Mrr1. A crucial step in this process is the crystallization of the protein of interest. Only if crystallization is successful, detailed molecular insights into the protein structure can be gained by X-ray crystallography.

3.3.1 Vapor diffusion method

Several methods are known to achieve protein crystallization, as micro-batch, micro-batch dialysis, free-interface diffusion or vapor diffusion. In general, to achieve protein crystallization, the protein has to be brought into a supersaturated state, which then forces the protein to go out of solution. This happens either disordered or highly ordered, the latter process is called nucleation and the other, disfavored process, is called precipitation. Precipitation also happens, if the protein in solution is already too highly concentrated at the beginning of the experiment. The processes happening during crystallization are described in **Figure 17**, where the arrow labelled with (ii) represents the vapor diffusion method. Only the vapor diffusion method was applied in this work, and will therefore

Methods

be introduced here. When the supersaturated protein forms nuclei, the protein concentration in the solution is reduced and crystal growth can take place until the undersaturated protein concentration is reached again. Vapor diffusion is achieved in a closed system, where a protein-precipitant mixture in a drop, generally in a 1 to 1 ratio, is either hanging on a glass cover (hanging-drop method) or sitting on an elevated 'bench' (sitting-drop method). In addition to the mixture a larger volume of precipitant solution, in a reservoir, is present. As the precipitant concentration in the drop is much lower than in the reservoir an equilibration process takes place that reduces the droplet size, by diffusion of water and other volatile solvents from the drop to the precipitant. This decrease in drop volume leads to an elevation of the protein concentration and can reach supersaturation, which might lead to nucleation followed by crystallization. (Chayen and Saridakis, 2008; Rupp, 2010)

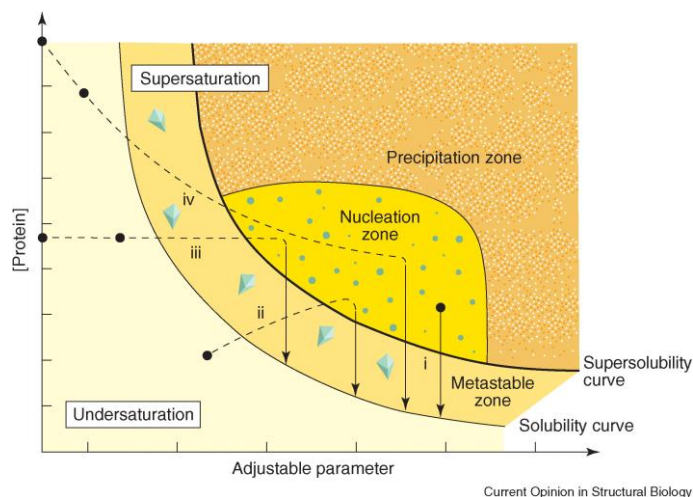


Figure 17 - Phase diagram of protein crystallization (Schematic illustration):

The phase transitions for differing methods are depicted; (i) micro-batch, (ii) **vapor diffusion** (iii) dialysis, (iv) FID. (Chayen, 2004) page 579, Figure 1 with permission from Elsevier

3.3.2 Sample preparation

A protein sample suitable for crystallization should be as homogeneous and monodisperse as possible and the protein should be in a buffer that stabilizes it. In advance to the crystallization process, dispersity and stability of the protein were investigated by DLS, SEC-MALLS and thermofluor analysis. Another critical component of the process is the protein concentration, which is crucial, as too low concentrations would not allow to achieve supersaturation and too high concentrations could result in immediate precipitation without crystallization. The optimum protein concentration cannot be calculated and has to be determined empirically. In general, crystallization trials are started at protein concentrations around 10 mg/ml and the amount of precipitate in 'indicator' screens as Crystal Screen I/II or the Index Screen can be used to fine-tune the protein concentration.

In this work FadA5 WT and its variants were crystallized in the size-exclusion buffer (20 mM Bicine-NaOH pH 8.5, 250 mM NaCl) at concentrations around 12 mg/ml. Prior to crystallization the proteins were slowly thawed on ice and centrifuged (16,100 g) for 30 minutes at 4 °C to remove precipitated protein or dust particles. The protein solution was then transferred into a new tube and was ready for crystallization.

Methods

FadA5 was also applied to co-crystallization experiments with different compounds, such as acetyl-CoA, 3-oxo-pregn-4-ene-20-carboxyl-CoA (3-OPC-CoA) or 3-oxo-cholesterol-4,22-diene-24-oyl-CoA (3-OCD-CoA). Acetyl-CoA was dissolved in water at concentrations of up to 50 mg/ml. The steroid compounds (3-OPC-CoA and 3-OCD-CoA) were less soluble and maximum concentrations of 10 mg/ml were reached in the protein buffer. Co-crystallization was performed with a 5 to 10-fold molar excess (1.4 mM to 2.8 mM) of the ligand relative to the protein, and the ligand was added to the protein solution prior to crystallization. Acetyl-CoA was incubated for 1 h at 4 °C with the protein prior to crystallization. As the steroid compounds were not as soluble, the according amount for a 5 to 10-fold molar excess was added to the protein solution. By this, the protein was diluted too much and so the mixture had to be concentrated by a Speed-Vac (Eppendorf) until the original sample volume was approximately restored. The concentrated samples were transferred into a new tube and were ready for crystallization

The Mrr1 '250' truncation variant was crystallized unmodified or with the introduced K335N, P683S and G997V mutations. Protein samples with concentrations between 10 mg/ml and 16 mg/ml in different buffers were centrifuged (16,100 g) for 30 minutes at 4 °C to remove precipitated protein or dust particles. The protein solution was then transferred into a new tube and was ready for crystallization

3.3.3 Initial screening

At the beginning of every crystallization approach initial screening with various purchasable screens was performed (Screens 2 - 7, 9 - 14 [Table 15](#)). With the help of these screens it is possible to test more than 1,000 different crystallization conditions, with various buffers, pH ranges, different precipitants, alcohols and salts, in a comparably short time. (Bergfors, 1999; Chayen and Saridakis, 2008)

Dependent on the amount of available protein, the initial crystallization trials were performed with a selection of the twelve available screens in 96-well format utilizing the Honey Bee crystallization robot (Zinsser analytic). In these initial screen, 40 µl of precipitant were filled into the reservoir and 0.3 µl of protein or protein-ligand mixture were pipetted to 0.3 µl of precipitant on a 'bench'. The screens were sealed by the sealing robot RoboSeal (HJ-Bioanalytik) with adhesive foil, to ensure a closed crystallization atmosphere. The screens were incubated at 20 °C or 4 °C and were evaluated at several time points with increasing intervals at the Stemi 2000-C microscope (Zeiss).

Methods

Depending on the ratio between precipitate containing and clear drops after one day in the Crystal Screen and the Index Screen, it was evaluated if the protein concentration should be increased or lowered for following set ups.

3.3.4 Hit optimization

In case that initial screening led to successful crystallization, this “hit” condition was chosen for further fine screening. This means that the content of the precipitant from the hit condition was varied with respect to its concentration, its pH or through the addition of additives, e.g. from the Additive or Silver Bullets screens (1, 8, [Table 15](#)). Other possibilities are the variation of the protein condition, either by applying higher or lower protein concentrations or by changing the protein to precipitant ratio.

The common setups for crystallization were 0.5 μ l protein plus 0.5 μ l precipitant and 40 μ l of reservoir solution in a 96-well format, or 1 μ l protein plus 1 μ l precipitant with 1 ml reservoir solution in a 24-well format.

The optimization trials for FadA5 were mainly performed in the 96-well format as set ups with 24-well plates often did not yield crystals. For FadA5 two individual screens the FadA5 Screen and the Thiolase Screen ([Table S 4](#) / [Table S 5](#)) were designed and used for hit optimization. The optimization screens were incubated at 20 °C and were evaluated at several time points with increasing intervals at the Stemi 2000-C microscope (Zeiss).

Mrr1 optimization was performed in the 24-well and the 96-well format with the above described setup. Optimization setups were incubated at different temperatures, 4 °C and 20 °C, and different initial protein buffers were tested. In addition, Mrr1 was crystallized with and without a His₆-tag. Evaluation of the setups were performed at several time points with increasing intervals.

3.4 X-ray crystallography

After crystallization has been successfully achieved the next crucial step on the way to the structural characterization of the macromolecule is x-ray crystallography. This step requires a lot of computational power and the application of computational methods. If not otherwise stated, the references for this chapter are from Bernhard Rupp's book 'Biomolecular Crystallography' (Rupp, 2010) or Gale Rhodes 'Crystallography made crystal clear' (Rhodes, 2006).

3.4.1 X-ray diffraction and important factors

X-rays, with wavelengths (λ) between 10^{-8} m (100 Å) and 10^{-12} m (0.01 Å), are necessary to analyze structures on the molecular level. Bond lengths in organic molecules lie between 1 Å and 1.6 Å and probes in the same order of magnitude are necessary to detect and resolve these. A protein crystal lattice is of parallel nature and is important for the diffraction experiment because the electrons of atoms or molecules, arranged in such a lattice, are excited by intruding x-rays, which leads to their oscillation. When the excited, oscillating electrons return into their ground state, they emit radiation or they scatter rays with the original wavelength of the intruding x-ray but in a random direction. Formally the x-rays are reflected at the crystal lattice, as the incidence angle is the same as the exit angle. This process of scattering is of course not only occurring at one electron but at every electron of the crystal lattice, which leads to an uncountable number of scattered waves. Most of these waves are interfering in a destructive way and add up to zero but if the interference is constructive, fulfilling Bragg's law, the emitted waves can be detected and are called reflections. Bragg's law was described by Sir William Henry and Sir William Lawrence Bragg and states that constructive interference can occur, when the additional distance ($2d \cdot \sin\theta$) the x-rays have to overcome between two lattice planes is equal to an integer multiple (n) of the wavelength of the radiation (λ).

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

Equation 1 - Bragg's law: $n =$ integer number, $\lambda =$ wavelength, $d =$ distance between the lattice planes, $\theta =$ incidence / exit angle of the radiation

By the detection of the diffracting waves during the x-ray crystallography experiment, intensities I_{hkl} and positions of the reflections are reported. The positions are defined by their deviation from the beam center and are described by the coordinates h, k, l , also called the Miller indices.

The aim of an experimental structure solution is an electron density map of the protein, deduced from the electronic scattering, which is used for the generation of a three-dimensional model of the protein. Two important factors in this process are the electron density ρ_{xyz} and the structure factors F_{hkl} (Equation 2) that sum up the scattering contributions of every atom in the unit cell. The structure

Methods

factors are described in a Fourier transform by atomic scattering factors f_j , which specify the atomic character of the scattering atoms, and by the relative phase angles emanating from each atom. These are dependent on the direction of scattering, h, k, l , and the position of the atoms j in relation to the origin (x_j, y_j, z_j) .

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

Equation 2 - Calculation of the structure factors: h, k, l = Miller indices, f_j = atomic scattering factors, x, y, z = coordinates in the reciprocal lattice

The relation between the structure factors F_{hkl} and the electron density ρ_{xyz} is described by the following Fourier transform (**Equation 3**).

$$\rho_{xyz} = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{-2\pi i(hx + ky + lz)}$$

Equation 3 - Calculation of the electron density: V = volume of the unit cell

By examination of **Equation 3** it becomes evident that the structure factors F_{hkl} have to be known for the calculation of the electron density ρ_{xyz} . As described above the collected scattering data do only provide intensities for each reflection but not structure factors. The detected intensities I_{hkl} (**Equation 4**) are proportional to the square of the absolute value of the structure factor amplitudes F_{hkl} , which are a part of the complete structure factor but are lacking the corresponding phases φ_{hkl} (**Equation 5**).

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

Equation 4 - Correlation between the intensities I_{hkl} and the structure factor amplitudes F_{hkl} .

$$F_{hkl} = |F_{hkl} \cdot e^{i\varphi_{hkl}}|$$

Equation 5 - Alternative calculation of the structure factors F_{hkl} : F_{hkl} = structure factor amplitudes, φ_{hkl} = structure factor phase

The inability to measure the phases is called the phase problem of crystallography and several methods have been determined to overcome this problem. During this work the phase problem was solved by molecular replacement (MR) which will be described later.

Methods

3.4.2 Data collection and processing

The selection of crystals for data collection was performed with a microscope (Stemi 2000-C microscope, Zeiss). In advance to the crystal harvesting or fishing process a suitable cryoprotectant solution had to be mixed. These solutions are necessary as frost protection that prevents ice crystal formation in the protein crystal and the surrounding solution. Additives that are commonly used as cryoprotectants are glycerol, polyethylene glycol (PEG), ethylene glycol (EG), different sugars or 2-methyl-2,4-pentanediol (MPD) in concentrations between 15% and 30% (v/v). The cryoprotectant solution is designed by the increase of the original precipitant by 2% to 3% (v/v) and the addition of up to 30% (v/v) of the cryoprotectant to the original crystallization condition. Crystallization conditions with high amounts of salt can be cryoprotected by a maximum increase of the salt content although this often leads to detrimental salt crystallization. For crystals containing ligands the cryosolution was also supplemented with a distinct amount of the ligand. (McFerrin and Snell, 2002)

The well, containing the chosen crystals, was opened and 0.5 μ l to 1 μ l of reservoir solution (mother liquor), depending on the drop size, was added to the drop, to prevent drying out. The crystal was then fished with a nylon loop, in the appropriate size, transferred into the cryoprotectant solution and incubated for a few seconds or up to 10 minutes in the solution. After incubation, the crystal was flash frozen in liquid nitrogen and stored in suitable containers until the measurement was performed.

Initial test pictures and validation of the harvested crystals was performed with the in-house x-ray source (MicroMax-007, Rigaku), so that a selected collection of promising crystals could be sent to the synchrotron. For collecting diffraction images, the crystal was mounted on the goniometer and centrally positioned in the x-ray beam. The primary beam is absorbed by a beam stop, so that it cannot be detected by the detector and outshine the comparably weak reflection spots. The data collection takes place at -180 °C in a cold nitrogen stream. During this study different experimental set ups were used. The in-house x-rays were produced by a rotating copper anode that is emitting radiation at a fixed wavelength of 1.5418 Å, and the diffraction patterns were detected with an imaging plate detector (R-Axis IV++ and R-Axis HTC, Rigaku). However, data collection was mainly performed at synchrotron facilities (Berliner Elektronen Speicherring Gesellschaft für Synchrotronstrahlung, **BESSY II**, Berlin; European Synchrotron radiation Facility, **ESRF**, Grenoble), where synchrotron radiation with wavelengths of 0.9184 Å (BESSY) and 0.9537 Å (ESRF) was used and diffraction data were recorded on CCD (charge-coupled device) or hybrid photon counting (e.g. PILATUS) detectors.

To collect diffraction data it is necessary to rotate the crystal during the exposure with x-rays around the phi axis. The amount of rotation to obtain clearly separated diffraction spots ranges from 0.1 ° to

Methods

1°. It is dependent on the quality of the diffraction spots, the unit cell dimensions and the crystal mosaicity. To obtain useful diffraction data the single spots have to be well defined and separated, this separation can be improved by reducing the rotation angle. The minimal angle range for the data collection process, as well as the total necessary rotation of the crystal, dependent on the crystal symmetry, was determined by the program iMosflm (Leslie, 1992) in advance to the experiment with the help of recorded test images.

After data collection the diffraction images were indexed and integrated with the help of programs as XDS (Kabsch, 2010) / XDSapp (Krug et al., 2012) or iMosflm (Battye et al., 2011), where first information about mosaicity, symmetry and unit cell dimensions were gained. In addition the program Pointless (Evans, 2006) from the CCP4 package was used to identify or verify the assigned space group. Subsequently the datasets were scaled with either Scala (Evans, 2006) or Aimless (Evans, 2011), which in addition provided information concerning data quality, signal-to-noise ratio, redundancy and completeness. Dependent on different factors such as $\langle I/\sigma \rangle$, R_{merge} , overall and in the highest-resolution shell, if applicable R_{pim} , and the Pearson correlation coefficient between random half-datasets $CC_{1/2}$ it was decided to which resolution the dataset could be used or where it should be truncated. The R_{merge} is defined by Equation 6 and should not exceed 60% to 80%, for the highest resolution shell, and at the same time the empirical signal-to-noise ratio should not drop below 2.0. (Evans, 2011) The overall R_{merge} should not exceed 20%, because the data are defined as of questionable quality then. (McRee and David, 1999) The R_{merge} is only used for not highly redundant data; in case of redundant data, the so called R_{pim} is applied in addition for the validation. The R_{pim} is defined similarly as the R_{merge} but employs an additional term taking the redundancy and thereby the higher accuracy of each reflection into account. The described limits are also true for the R_{pim} .

$$R_{merge} = \frac{\sum_{\mathbf{h}} \sum_{i=1}^N |I_{(\mathbf{h})i} - \bar{I}_{(\mathbf{h})}|}{\sum_{\mathbf{h}} \sum_{i=1}^N I_{(\mathbf{h})i}}$$

Equation 6 - Definition of the R_{merge} value: N = number of reflections \mathbf{h} , $I_{(\mathbf{h})i}$ = intensity of the reflection, $\bar{I}_{(\mathbf{h})}$ = averaged intensity of each reflection

A more statistically meaningful value is $CC_{1/2}$, which is also calculated by the programs XDS and Aimless. Diederichs and Karplus define the limit of $CC_{1/2}$ as 20%, but Aimless recommends a cutoff at $CC_{1/2} = 50\%$. For the calculation of $CC_{1/2}$, the unmerged data is divided into two parts, each containing a random half set of the measurements of each unique reflection. Then, the Pearson correlation coefficient CC (Equation 7) was calculated between the average intensities of each subset. This quantity is then denoted $CC_{1/2}$. (Diederichs and Karplus, 1997; Karplus and Diederichs, 2012)

Methods

$$CC = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{(\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2)^{1/2}}$$

Equation 7 - Pearson correlation coefficient CC: $x, y =$ paired variables;

3.4.3 Phase determination and molecular replacement (MR)

To apply molecular replacement (MR) for solving the phase problem a structural model with reasonably high sequence identity is needed. The literature describes a sequence identity of at least 30% as necessary. In this work, MR was only performed for the FadA5 protein, where the protein data bank (pdb) entry 1ulq, a thiolase from *Thermus thermophilus*, with the highest available sequence identity (43%) to FadA5 was chosen as model. After this initial structure determination the new FadA5 structure was used as search model for the following structure solutions by MR.

Before the actual structure solution, the number n of FadA5 molecules in the asymmetric unit had to be determined. This was calculated with the program MATTHEWS_COEF (1994), belonging to the CCP4 program package. The Matthew's coefficient V_M (**Equation 8**) describes the crystal volume of protein with a specific mass and takes space group symmetry, the volume of the unit cell and the molecular weight of the protein into account. The reasonable number of molecules in the asymmetric unit is determined. In addition to n , the solvent content of the asymmetric unit is calculated. The values for V_M are normally between 1.72 Å³/Da and 3.53 Å³/Da and the solvent content is between 27% and 78%. (Matthews, 1968)

$$V_M = \frac{V_{unit\ cell}}{M_r \cdot z \cdot n}$$

Equation 8 - Calculation of the Matthew's coefficient: $V_{unit\ cell} =$ volume of the unit cell; $M_r =$ molecular mass of a protein monomer; $z =$ number of symmetry operators of the space group

After the number of molecules per asymmetric unit was determined, MR was performed with the program Phaser (McCoy, 2007). A crucial function for solving the phase problem is the Patterson function, which is calculated for the search model and for the measured data. The Patterson function P_{uvw} (**Equation 9**) is phase-independent and the square of the experimentally determined structure factor amplitudes $|F_{hkl}|^2$ is sufficient to calculate P_{uvw} . By rotational and translational search processes Phaser tries to determine the optimal agreement between Patterson functions derived from either the experimental data or the search model. This is done in two separate steps as each search has to be performed in three dimensions and a 6-dimensional search would be too complex. The result of MR is an initial electron density that is occupied by the protein sequence of the search model and has to be adapted to the protein of interest by model building.

$$P_{uvw} = \frac{1}{V} \sum_{hkl} |F_{hkl}|^2 \cdot \cos[2\pi(hu + kv + lw)]$$

Equation 9 - Patterson function P_{uvw} : u, v, w = coordinates in the Patterson space; V = volume of the unit cell; h, k, l = Miller indices

3.4.4 Model building, refinement and validation

The initial model building was supported by automated model building with ARP/wARP (Langer et al., 2008) that replaced the sequence of the search model by the actual protein sequence. In addition, regions with additional amino acids or gaps in contrast to the search model were mainly adapted by the program. After the initial structure of Fada5 was solved and validated, it was used as search model for MR and ARP/wARP was dispensable. All structures were modeled, adapted and built with alternating cycles of model building in the program Coot (Emsley and Cowtan, 2004) and refinements and validations with the programs Refmac 5.5.0102 (Pannu et al., 1998) or Phenix 1.9-1692 (Adams et al., 2010). During model building solvent molecules such as water or other molecules of the protein buffer or cryoprotectant solution were also included into the model. When protein modeling was advanced, co-crystallized ligands were included in the model and appropriate geometry files were calculated by the PRODRG server (Schuttelkopf and van Aalten, 2004). In both refinement programs maximum likelihood refinement was applied. The aim of every refinement, independent from the applied method, is to obtain maximum accordance between observed $|F_{obs}|$ and calculated $|F_{calc}|$ structure factor amplitudes. In case that more than one protein molecule is present in the asymmetric unit non-crystallographic symmetry (NCS) restraints can be applied during refinement. Translation-libration-screw motion refinement (TLS refinement) was used, either with the standard settings or if applicable, after determination of the TLS domains by the TLSMD server (Painter and Merritt, 2006), with self-defined groups. The model quality in general and the improvements by model building were supervised by the observation and trends of the R -values, which are a sign for the agreement between the observed and calculated structure factors.

$$R = \frac{\sum_{hkl} ||F_{obs}| - |F_{calc}||}{\sum_{hkl} |F_{obs}|}$$

Equation 10 - Calculation of the R -values: h, k, l = Miller indices; F_{obs} = observed structure factors; F_{calc} = calculated structure factors

The R -values, R_{work} and R_{free} , are defined in **Equation 10**. As the calculated data should of course be similar to the observed data, the difference between both should be as small as possible. In literature target values as 0.2 for R_{work} and 0.25 for R_{free} are defined but of course both values are highly dependent on the data resolution and quality. In general the R -values should be decreasing in the

Methods

process of model building and refinement but initial R -values above 0.5 may indicate that the model is probably wrong. The initially defined R_{work} value is prone to model bias, as it is based on the data that is used during the refinement process. This might lead and has in the past led to overfitting of the model and therefore Axel T. Brunger introduced an R -value free of model bias, which he called R_{free} (Brunger, 1992). The R_{free} follows the same definition as the R_{work} , but for its calculation 5% of the data are excluded from the refinement, and therefore also not affected by it. The excluded data are randomly chosen and the R_{free} is in general 3% to 6% higher than the R_{work} . The R_{free} indicates success or failure of model building by decreasing or increasing values respectively.

When model building and refinement had reached congruence between the initial data and the built model, the model had to be validated. Validation means that the geometry of bands, angles and the orientation of peptides, cis versus trans, is checked. This is done with the help of programs such as MolProbity (Chen et al., 2010), which outline questionable bond lengths, angles, poor rotamers or Ramachandran outlier, which concern deviations in the dihedral angles ψ and Φ . The basis of this analysis are statistical values from protein data bank entries or other libraries but small deviations from these standards are to be expected. During the validation process, outliers were systematically checked in Coot, with the help of the electron density, and, if they proved to be wrong, they were corrected. Ramachandran outliers were checked and tried to overcome, the aim here was to have less than 0.05% outlier and more than 98% of the residues in the favored geometry. In addition to the geometrical validation, MolProbity lists information about clashing atoms, between protein-protein, protein-solvent or protein-ligand, and ligands with questionable density. Atom clashes are defined by a serious steric overlap of $> 0.4 \text{ \AA}$ between adjacent atoms.

After the final evaluation, the secondary structure motifs were assigned with the program dssp (Kabsch and Sander, 1983), which was necessary to especially assign motifs as 3_{10} -helices, which are often not recognized by structure illustrating programs. Pymol (2010; DeLano, 2002) is such an illustrating program and was used for the creation of structure images but also for the examination of protein-protein, protein-solvent or protein-ligand interactions. PDBe (Protein Data Base of Europe) programs such as PDBeFold (Krissinel and Henrick, 2004) and PDBePISA (Krissinel and Henrick, 2007) were also employed to analyze and compare the structures with each other or with homologs.

3.4.5 Modeling and Hotspot analysis

One of the aims of this work was preparing a structural base for lead design to specifically target FadA5. This was addressed by a comparison of FadA5 with human thiolases and the application of a hotspot analysis, which was performed by Johannes Schiebel from Prof. Dr. Klebe's laboratory (Marburg).

Based on the structure solution of FadA5 and known human thiolase structures, the I-TASSER server (Roy et al., 2010; Zhang, 2008) was used to model so far unknown human thiolases and compare their modeled structure with FadA5. Six human thiolase have been identified, namely the mitochondrial enzymes T1-thiolase (T1), T2-thiolase (T2) and the trifunctional enzyme thiolase (TFE) as well as the peroxisomal SCP2-thiolase (SCP2), the human cytosolic thiolase (CT) and the human peroxisomal AB-thiolase (AB). For three of these enzymes, T2 (pdb entry 2ibw), CT (pdb entry 1wl4) and AB (pdb entry 2iik) structural information is available. For modeling two different modeling strategies were applied, either a distinct template for the modeling plus the Protein Data Bank (PDB) was chosen or random PDB modeling was allowed. This means, that either a distinct structure, 1wl4, 2iik or FadA5 was chosen as primary template or all structures in the PDB were available as modeling basis. For modelling of the TFE thiolase an additional pdb entry (4b3h) of a mycobacterial TFE was included. The I-TASSER server usually released five model structures per search, thus yielding 20 or 25 (TFE) models per structurally uncharacterized human thiolase.

HotspotsX (version 0.60) was used to calculate hotspots of binding with pair potentials derived from the CSD (Neudert and Klebe, 2011; Velec et al., 2005). Hotspot analyses were performed for seven different probes: a hydrogen-bond acceptor (ACC), hydrogen-bond donor (DON), hydrogen-bond acceptor/donor (AnD), hydrophobic (HYD), aromatic (ARO), fluorine (FLU) and halogen (HAL) probe. combinations of certain fconv atom types were used to define the probes ACC, DON, AnD, HYD and ARO (Neudert and Klebe, 2011) analogous to a study of Behnen et al. except that ACC also included P.3 and that S.sh/S.s,S.3 were part of AnD/ACC instead of HYD (Behnen et al., 2012). In addition, a fluorine (FLU; fconv atom type F.0) and halogen probe (HAL; atom types Cl.0, Br.0, I.0) were used.

All FadA5 residues of the OPC-complex structure located within a 10 Å radius of the ligands CoA and OPC (114 residues from subunit A, 6 from subunit B) were chosen for the hotspot analyses. Waters and ligands were not included during the calculation of the contour maps.

4 Results

4.1 Mrr1 – a transcription factor from *C. albicans*

4.1.1 Cloning

The gene sequence used for cloning was codon optimized for expression in *E. coli*. The gene synthesis, performed by Geneart (Regensburg) was necessary, as *C. albicans* follows a different codon usage than *E. coli* and most other organisms. In *C. albicans*, the codon CUG is translated into a serine instead of a leucine, therefore this codon was replaced and in general a codon optimization for *E. coli* as expression host was performed (Santos and Tuite, 1995).

4.1.1.1 Secondary structure analysis

Prior to the cloning experiments, secondary structure analyses with different online servers (Porter, Phyre² and Scratch, [Table 17](#)) were performed to identify stretches in the protein sequence that could cause problems in expression and proper folding as well as during purification. In addition, it is known that the length of the construct influences its soluble protein expression, functional purification and subsequent successful crystallization. (2008) The estimation for successful crystallization of proteins containing more than 1000 amino acids is far below 10%. The server predictions identified a long stretch of amino acids (M1 to S299) at the *N*-terminus that hardly seem to assume any defined secondary structure and might therefore be highly disordered in the absence of additional factors, such as specific ions, DNA or interacting proteins. Amino acids I300 to E1020, or E1027, are predicted to adopt mostly an α -helical fold which is hardly interrupted by disordered regions. This stretch was also identified to form a so called Armadillo repeat, which is the repetition of a not defined number of α -helices with lengths of up to 40 amino acids per helix. (Peifer et al., 1994) The remaining *C*-terminal amino acids, 1027-1108 or 1020-1108 respectively, are again predicted to be mainly disordered.

With the additional information of the location of gain of function mutations in the *MRR1* gene, which are all assigned to the above mentioned Armadillo repeat region, constructs for cloning of different lengths were designed. The following constructs were generated: Mrr1 Full length (FL); a construct containing only the Armadillo repeat region, Mrr1 Armadillo repeat (AR); and an *N*- and *C*-terminally elongated Armadillo repeat construct, Mrr1 250 ('250') ([Figure 18](#)).

Results

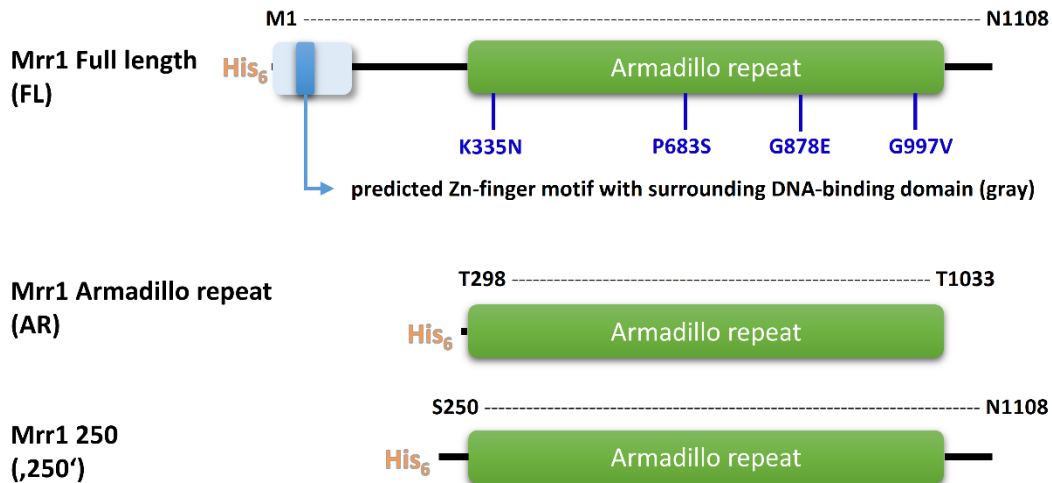


Figure 18 - Mrr1 constructs for expression in *E. coli*:

Three different constructs of Mrr1 were initially cloned for expression tests in *E. coli*. The Zn-finger (|), the surrounding DNA-binding domain (|) and the Armadillo repeat region (|) are highlighted and the amino acid range is indicated above the constructs. All three constructs contained an N-terminal His₆-tag (orange). Later a Mrr1 '250' CT construct was generated that contained a C-terminal His₆-tag but otherwise did not differ from the original Mrr1 '250' construct with the exception of the location of the tag and the absence of the TEV-cleavage site. The location of the gain of function mutations (blue) introduced into the Mrr1 '250' construct are shown in the FL protein.

4.1.1.2 Mrr1 constructs for expression in *E. coli*

Cloning of Mrr1 FL and AR constructs was performed with SLIC as described in 3.1.6.2. The DNA template for Mrr1 FL was the engineered Mrr1 DNA from Geneart, which was amplified with primers 4 and 20 (Table 24) to generate overlapping ends complementary to the amplified pBadm11 vector (primer 1 and 2, Table 24). The transformed (3.1.4.1) cloning samples produced a large number of clones out of which several were tested by colony PCR (3.1.6.4) utilizing primers 26 and 27 (Table S 1). The colony PCR revealed one potentially positive clone, which was applied to plasmid amplification and isolation (3.1.5) and subsequently was verified as being correct by sequencing (Seqlab, Göttingen).

This newly generated construct was used as the template for the generation of the Mrr1 AR truncation variant. The insert DNA was amplified with primers 22 and 23 and pBadm11 was amplified with the primer pair 1 and 2 (Table 24). The following procedure was the same as described above. Again, a lot of colonies grew and selected ones were chosen for colony PCR, which pointed to one potentially positive clone. The amplified and isolated DNA was verified by sequencing to be correct (Seqlab, Göttingen).

Attempts to generate Mrr1 '250' by SLIC as described above, did not yield positive clones and therefore the strategy was changed to restriction digestion and ligation cloning (3.1.6.2). The Mrr1 FL construct was used as a template and with primers 28 and 29 (Table 24) restriction sites for NcoI and

Results

KpnI were generated by PCR (3.1.1). The amplified insert DNA and the pBadm11 vector were separately digested with the two restriction enzymes (NcoI, KpnI) to generate overlapping sticky DNA ends. The cleaned digested vector DNA was treated with Antarctic phosphatase to prevent recircularization of the vector. Selected colonies were used for colony PCR. The amplified and isolated DNA of two potentially positive clones was verified by sequencing to be correct in both cases (Seqlab, Göttingen). Due to the cloning procedure, two additional amino acids, a methionine at position -1 and an alanine at position 0, were introduced between the TEV-cleavage site and the insert sequence.

Four out of 18 gain of function mutations (K335N, P683S, G878E, and G997V) were chosen for site directed mutagenesis (1.2.6). K335N and G997V were selected due to possible secondary structure changes calculated with the help of the relative occurrence of amino acids in secondary structure motifs (Berg et al., 2007). These values indicated either disturbances in helix formation or potential for helix elongation. P683S was chosen, due to its activation potential of Mrr1 and due to the fact that it was used in several studies as selected gain of function mutant (Schubert et al., 2011a). G878E is located in the so called fourth mutational hot spot and therefore was chosen as a representative variant (Dunkel et al., 2008a). Mutations were introduced into the Mrr1 '250' construct only, the location of the mutations, in context of the FL protein, is shown in **Figure 18**. The template was amplified (3.1.1) with a complementary primer pair (**Table 24**) that contained the base pair alteration(s) that were necessary to introduce the mutation. Several colonies were chosen for DNA amplification and isolation and were subsequently verified by sequencing (Seqlab, Göttingen). As the Mrr1 '250' construct was used as a template, two additional amino acids, M-1 and A0, are part of the constructs between the TEV-cleavage site and the insert sequence. The Mrr1 '250' CT construct with the C-terminal His₆-tag was kindly provided by Sebastian Kaiser, AG Kisker.

Table 24 - Mrr1 constructs for E. coli expression

No	construct	organism	vector	DNA sequence	amino acid sequence	Cloning primer	notes	ORF
4	pBadm11 FL	<i>E. coli</i>	pBadm11	1 - 3324	M1 - N1108	i: 3 / 4 v: 1 / 2	--	complete
5	pBadm11 AR	<i>E. coli</i>	pBadm11	891 - 3099	T298 - T1033	i: 5 / 63 v: 1 / 2	--	complete
6	pBadm11 '250'	<i>E. coli</i>	pBadm11	751 - 3324	S250 - N1108	i: 7 / 8	additional amino acids M-1 and A0	complete
16	pBadm11 '250' K335N	<i>E. coli</i>	pBadm11	751 - 3324	S250 - N1108	28 / 29	K --> N confirmed	complete
17	pBadm11 '250' P683S	<i>E. coli</i>	pBadm11	751 - 3324	S250 - N1108	20 / 21	P --> S confirmed	complete
18	pBadm11 '250' G878E	<i>E. coli</i>	pBadm11	751 - 3324	S250 - N1108	26 / 27	G --> E confirmed	complete
19	pBadm11 '250' G997V	<i>E. coli</i>	pBadm11	751 - 3324	S250 - N1108	18 / 19	G --> V confirmed	complete
20	pETm11 '250' CT	<i>E. coli</i>	pETm11	751 - 3324	S250 - N1108	i: 32 / 33 v: 30 / 31	C-terminal His ₆ -tag	complete

i: primer for insert amplification; v: primer for vector amplification

Results

Table 25 - General information about the Mrr1 *E. coli* constructs

No	construct	calculated mass (kDa)	calculated isoelectric point (pI)	calculated extinction coefficient ϵ ($M^{-1}\cdot cm^{-1}$)
4	pBadm11_FL	129.8	8.24	114,490
5	pBadm11_AR	88.5	7.29	91,110
6	pBadm11_'250'	102.2	6.17	100,050
16	pBadm11_'250'K335N	102.1	6.10	100,050
17	pBadm11_'250'P863S	102.0	6.17	100,050
18	pBadm11_'250'G878E	102.1	6.11	100,050
19	pBadm11_'250'G997V	102.2	6.17	100,050
20	pETm11_'250' CT	100.7	6.23	97,070

4.1.1.3 Mrr1 constructs for expression in *P. pastoris*

For protein expression in *Pichia pastoris* as expression host, three different vectors were available. Two vectors, pPinkHC and pPinkLC, are suitable for intracellular protein expression but differ with respect to the plasmid copy number in the cell, high copy versus low copy numbers. The pPink α HC vector can be used for secreted protein expression. To test which expression system is most suitable for the Mrr1 constructs, the three non-mutated variants (**Figure 19**) were cloned into all three vectors, with the exception that the Mrr1 Armadillo repeat construct was prolonged by the complete C-terminus of the protein to overcome expression problems observed in *E. coli*.

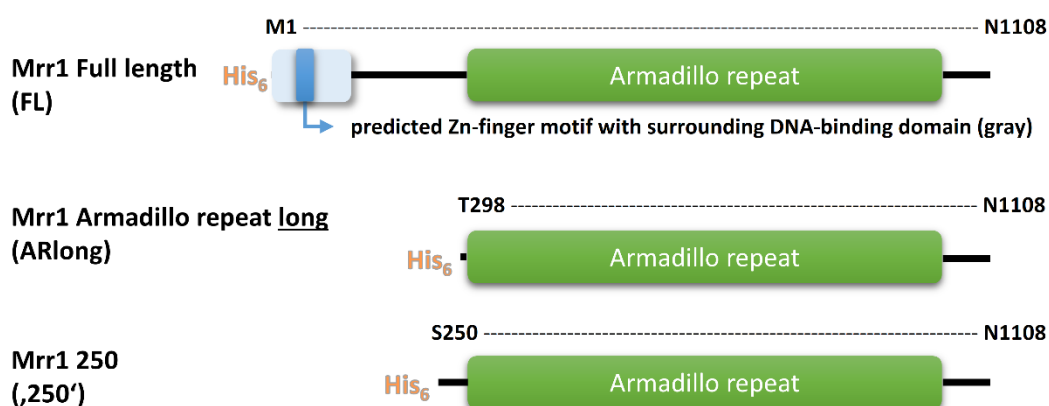


Figure 19 - Constructs of Mrr1 for expression in *P. pastoris*:

Three different constructs of Mrr1 were initially cloned for expression tests in *E. coli*. The Zn-finger (|), the surrounding DNA-binding domain (|) and the Armadillo repeat region (|) are highlighted and the amino acid range is indicated above the constructs. All three constructs contained an N-terminal His₆-tag (orange).

Cloning of Mrr1 FL into the secreting pPink α HC vector was performed with SLIC as described in 3.1.6.2 and above. The DNA template for this construct was Mrr1 FL pBadm11 for *E. coli* expression, which permitted the transfer of the His₆-tag for subsequent affinity purification. For amplification of the vector, primers 32 and 33 were used and for the insert amplification the primer pair 35 and 36 was employed (**Table 26**). The transformed (3.1.4.1) cloning samples produced a large number of clones out of which several were tested by colony PCR (3.1.6.4) with primers 40 and 67 (**Table S 1**).

Results

The colony PCR revealed two potentially positive clones which were applied to plasmid amplification and isolation (3.1.5) and were subsequently verified as correct by sequencing (Seqlab, Göttingen).

The other Mrr1 constructs for secreted protein expression were cloned, based on the pPink α HC Full Length construct, by Lisa Krebs, AG Kisker, in a variation of the in 3.1.6.2 described SLIC method with the primers listed in **Table 26**.

The constructs for intracellular protein expression were cloned, based on the pBadm11 Full Length construct, by Lisa Krebs, AG Kisker, with a protocol based on the SLIC method described in 3.1.6.2 with the primer pairs listed in **Table 26**.

All constructs for *P. pastoris* protein expression were verified by sequencing (Seqlab, Göttingen).

Table 26 - Mrr1 construct list for *P. pastoris* expression

No	construct	organism	vector	DNA sequence	amino acid sequence	Cloning primer	notes	ORF
7	pPink α HC FL	<i>P. pastoris</i>	pPink α HC	1 - 3324	M1 - N1108	i: 11 / 12 v: 9 / 10	--	complete
8	pPink α HC ARlong	<i>P. pastoris</i>	pPink α HC	891 - 3324	T298 – N1108	13 / 14	--	complete
9	pPink α HC '250'	<i>P. pastoris</i>	pPink α HC	751 - 3324	S250 - N1108	14 / 15	additional amino acids M-1 and A0	complete
10	pPinkHC FL	<i>P. pastoris</i>	pPinkHC	1 - 3324	M1 - N1108	i: 11 / 12 v: 16 / 17	--	complete
12	pPinkHC ARlong	<i>P. pastoris</i>	pPinkHC	891 - 3324	T298 – N1108	22 / 23 13 / 14	--	complete
14	pPinkHC '250'	<i>P. pastoris</i>	pPinkHC	751 - 3324	S250 - N1108	24 / 25 14 / 15	additional amino acids M-1 and A0	complete
11	pPinkLC FL	<i>P. pastoris</i>	pPinkLC	1 - 3324	M1 - N1108	i: 11 / 12 v: 16 / 17	--	complete
13	pPinkLC ARlong	<i>P. pastoris</i>	pPinkLC	891 - 3324	T298 – N1108	22 / 23 13 / 14	--	complete
15	pPinkLC '250'	<i>P. pastoris</i>	pPinkLC	751 - 3324	S250 - N1108	24 / 25 14 / 15	additional amino acids M-1 and A0	complete

i: primer for insert amplification; v: primer for vector amplification

Table 27 - General information about the Mrr1 *P. pastoris* constructs

No	construct	calculated mass (kDa)	calculated isoelectric point (pI)	calculated extinction coefficient ϵ ($M_{-1} \cdot cm_{-1}$)
7, 10, 11	pPink α HC FL	129.8	8.24	114,490
8	pPink α HC ARlong	96.8	6.64	95,580
9	pPink α HC '250'	102.0	6.17	100,050
12, 13	pPinkHC/LC ARlong	97.0	6.64	95,580
14, 15	pPinkHC/LC '250'	102.2	6.17	100,050

Results

4.1.2 Expression in *E. coli* and purification experiments

4.1.2.1 Mrr1 Full length

4.1.2.1.1 Expression test

Prior to large scale expression experiments, small scale expression tests were necessary to ensure that expression of the target protein can be detected. For this purpose, construct **4** was transformed into BL21 (DE3), BL21(DE3)-pLysS, BL21-CodonPlus(DE3)-RIL, Rosetta (DE3) or BL21 Star (DE3) cells. The test culture was supplemented with the respective antibiotics and inoculated with one colony per expression test. After the cell culture was grown to an OD₆₀₀ of 0.6, protein expression was induced by the addition of L(+)-arabinose. The temperature was reduced to 30 °C or 16 °C and samples were taken after 3, 5 and 7 h or 5 and 7 h and overnight respectively. The cells were pelleted

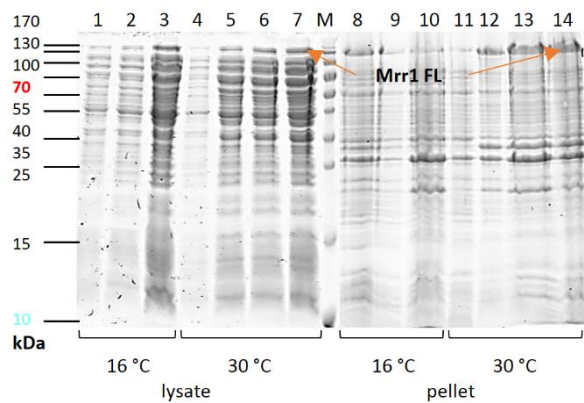


Figure 20 - 15% SDS-PAGE of the Mrr1 FL test expression with BL21 (DE3) cells:

M = marker; 1-3 (lysate) and 8-10 (pellet) fractions from the 16 °C test, 1/8 = 5 h sample, 2/9 = 7 h sample, 3/10 = overnight sample; 4-7 (lysate) and 11-14 (pellet) fractions from the 30 °C test, 4/11 = 0 h sample, 5/12 = 3 h sample, 6/13 = 5 h sample, 7/14 = 7 h sample. Arrows indicate the expected height of Mrr1 FL. Two separate gels are here combined together.

and stored at -80°C.

Cell lysis was performed by sonication in binding buffer, described in 3.2.1.1.2, and cell debris was separated by centrifugation from the lysate. As described in 3.2.1.2, the lysate was first incubated with Ni-NTA resin, then the resin bound protein was separated from the supernatant by centrifugation, washed twice and finally resuspended together with the resin. Samples from the protein bound to the resin and resuspended pellet fractions were analyzed by SDS-PAGE analysis.

The expression pattern from the transformed BL21 (DE3) (Figure 20) and Rosetta (DE3) cells was similar, with clear bands in the resuspended pellet fractions at the height of approx. 130 kDa (arrow right), which is the expected size of Mrr1 FL (Table 25). The band in the pellet fractions implicated a large amount of insoluble protein, which might eventually be extracted by a more intense cell lysis. Nevertheless, also the lysate fractions at 16 °C and 30 °C showed a band, although less prominent at the expected protein size (left arrow). The results of the expression test were thus promising with respect to soluble expression of this large 1108 amino acid long protein.

Results

4.1.2.1.2 Affinity chromatography purification tests

Large scale expression was performed with BL21 (DE3), Rosetta (DE3), or later with BL21 (DE3) arctic express cells. For this purpose, the full length construct (**4**) was transformed into the respective *E. coli* cells and a single colony was used to inoculate a preculture. The preculture grew overnight at 37 °C and was then used to inoculate the main culture, which grew at 37 °C also. Protein expression was induced with 0.02% L(+)-arabinose, when an OD₆₀₀ of 0.6 was reached. Then the cells were grown for 6 h or overnight at 30 °C or 16 °C respectively.

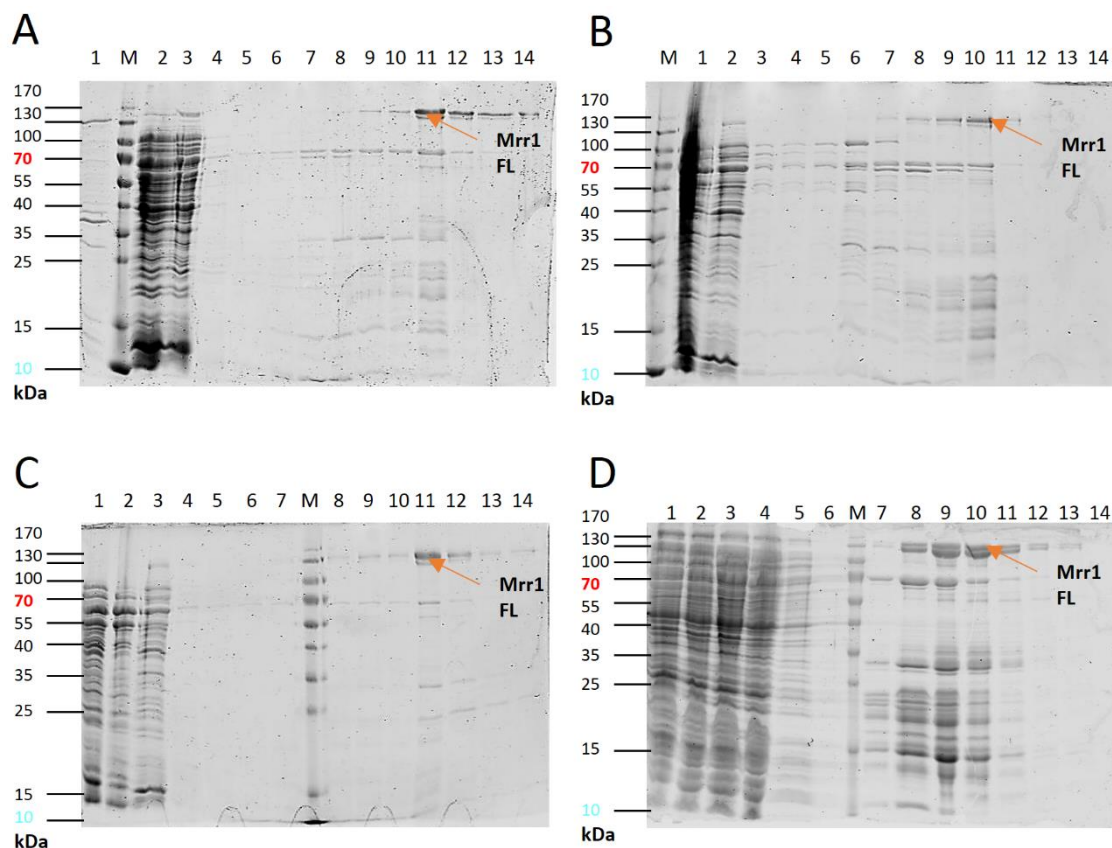


Figure 21 - 15% SDS-PAGE of the affinity chromatography purification tests of Mrr1 FL:

A) Ni-IDA resin: M = marker; 1 = resuspended lysed cells; 2 = flow through; 3-5 = wash unbound sample; 6-10 = wash fractions; 11-14 = elution fractions **B** (Ni-NTA): M = marker; 1 = flow through; 2-4 = wash unbound sample; 5-9 = wash fractions; 10-13 = elution fractions; 14 = empty **C)** Ni-TED: M = marker; 1 = lysate; 2 = flow through; 3-5 = wash unbound sample; 6-10 = wash fractions; 11-14 = elution fractions **D)** 1 ml HisTrap FF crude: M = marker; 1 = lysate; 2-5 = flow through fractions; 6-14 = elution fractions. **Arrows** indicate the expected height of

The cells were resuspended in binding buffer (**6**, **Table 12**) and lysed with the cell disruptor. Initially three different loose resins, Ni-IDA, Ni-NTA and Ni-TED, and the prepacked HisTrap FF crude column (1 ml) were tested (**Figure 21**). Therefore, the resins were loaded with the lysate and washed several times with buffer containing increasing imidazole concentrations. Elution was performed with two different buffers containing 150 mM or 250 mM imidazole. The load, wash and elution fractions were analyzed by SDS-PAGE.

Results

Purification with the Ni-TED resin led to the best results, since the sample eluted in only a few fractions and was already relatively pure (Figure 21C, lanes 11-14). The band at the height of approx. 130 kDa (orange arrows) was supposedly Mrr1 FL. Especially the purification with the HisTrap column performed at an Äkta system contained a lot of other proteins and a band shortly beneath the 130 kDa band was more prominent than the actual Mrr1 band. This second band may indicate protein degradation (Figure 21D, lanes 8-11). In addition to the different resins, several buffer combinations, with phosphate or Pipes-NaOH as buffer, varying salt concentrations and additives such as DTT were tested (Table S 2). The Tris-HCl buffer combination at pH 8.0 without any additives proved to be the best or not worse than more expensive buffer compositions. Due to these results Mrr1 FL was subsequently purified by affinity chromatography with Ni-TED resin as first step using buffers 6/7/9 (Table 12).

4.1.2.1.3 Thermofluor analysis

The initial size-exclusion chromatography results (not shown) indicated that Mrr1 FL was mainly eluting in the void volume of the size-exclusion column and therefore this construct is very likely forming higher, non-ordered oligomers or aggregates. Thermofluor analysis was therefore pursued to identify a buffer (combination) to improve protein stability and solubility, reduce the oligomerization and thereby obtain a protein in solution which is properly folded and useful for subsequent biochemical analyses or crystallization.

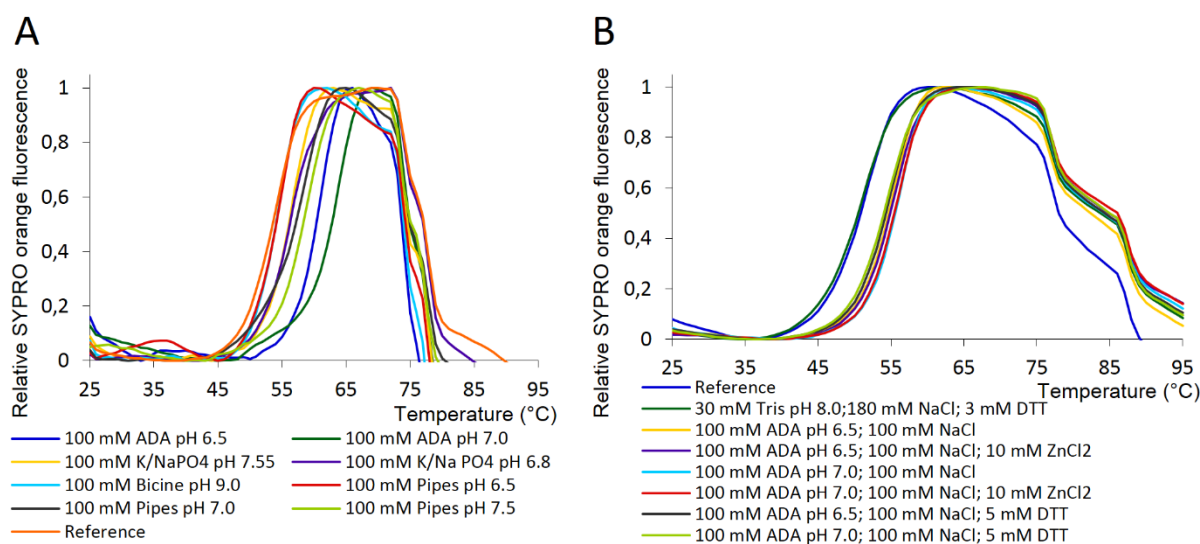


Figure 22 - Thermofluor experiments with Mrr1 FL:

A) Standard buffer thermofluor screen of the Mrr1 FL protein. The initial reference buffer was buffer 33 (Table S 3; 50 mM Tris-HCl pH 8.0, 300 mM NaCl) **B)** Thermofluor fine screen using the buffer hit conditions from A) at different salt concentrations and with other additives. The reference buffer was buffer 35 (Table S 3; 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 5 mM DTT)

Results

The thermofluor analysis was performed as described in 3.2.2.3 and initially the standard buffer screen (**Table S 3**) was used to evaluate the thermal protein stability in different buffers thereby analyzing a wide pH range (**Figure 22A**). This initial screen showed that certain buffers increased the melting temperature of the protein by up to 15 °C, which could also indicate an increase in protein stability. Stabilizing buffers were for example Ada-NaOH pH 6.5 / 7.5, Bicine-NaOH pH 9.0, Na/K phosphate pH 6.8 or Pipes-NaOH pH 6.5 / 7.0 / 7.5 (**Figure 22A**). Phosphate buffers are a difficult choice for proteins that shall be used for crystallization, as phosphate is prone to crystallize in the presence of divalent cations thus forming salt crystals which have to be distinguished from protein crystals. For this reason, the Ada-NaOH buffer at pH 6.5 was suggested to be the best buffer. Based on the initial screen a thermofluor fine screen was designed (**Figure 22B**) to examine whether further addition of salt (NaCl and/or ZnCl₂) or a reducing agent (DTT) could be beneficial. DTT or ZnCl₂ containing buffers did not significantly increase protein stability. Sodium chloride concentrations of 100 – 200 mM were sufficient for protein stabilization and of all tested buffers, Ada-NaOH pH 7.0 seemed to have the highest stabilization potential. Based on the thermofluor experiments, Ada-NaOH containing buffers were tested for further analysis by size-exclusion chromatography.

4.1.2.1.4 Size-exclusion chromatography tests and general purification procedure

Size-exclusion chromatography was used as second purification step of Mrr1 FL after the described Ni-affinity chromatography (4.1.2.1.2). This chromatography step was performed with different buffers, for example Tris-HCl or Ada-NaOH utilizing varying salt concentrations, different pH values and with DTT as additive. The first experiments (not shown) already indicated that Mrr1 FL is mainly forming protein aggregates eluting in the void volume of the size-exclusion chromatography columns. Therefore four different buffers (**33-36**, **Table S 2**) were tested in size-exclusion chromatography experiments and in addition two different size-exclusion columns and their elution behavior were analyzed. Most experiments were performed with HiLoad Superdex 200 pg columns but also a HiPrep 26/60 Sephacryl HR S-300 column was used, which has a broader separation range (10 kDa to 1.5 megaDa) compared to the Superdex columns (10 kDa to 0.6 megaDa). For the size-exclusion experiments, the elution fractions of the affinity chromatography step containing Mrr1 were concentrated to at least 5% of the size-exclusion column volume with Vivaspin concentrators (MWCO of 50,000 kDa).

The size-exclusion run was performed with an Äkta system (GE Healthcare) where protein elution was monitored with the help of the absorption at 280 nm and 260 nm. SDS-PAGE analysis of the size-exclusion chromatography run with a Sephacryl column showed that Mrr1 FL protein is mainly found

Results

in the void volume peak I (**Figure 23A**), where the $\frac{A_{260}}{A_{280}}$ ratio indicates that the protein is forming aggregates. Protein normally only elutes with the void volume of the column if it forms large clusters, which are not able to enter the resin pores. Elution peak II, which was more a tailing of peak I than a single peak, also mainly contained Mrr1 FL but still with a very high $\frac{A_{260}}{A_{280}}$ ratio. The fractions between peaks II and III are did not yield any bands in the SDS-PAGE as the protein amount was too small (**Figure 23B**). Analysis of peak III also shows no protein bands. This peak is likely the imidazole elution peak, which was a remainder of the Ni-affinity chromatography step. The protein containing peaks I and II were separately pooled, concentrated, flash frozen and stored at $-80\text{ }^{\circ}\text{C}$ for later use.

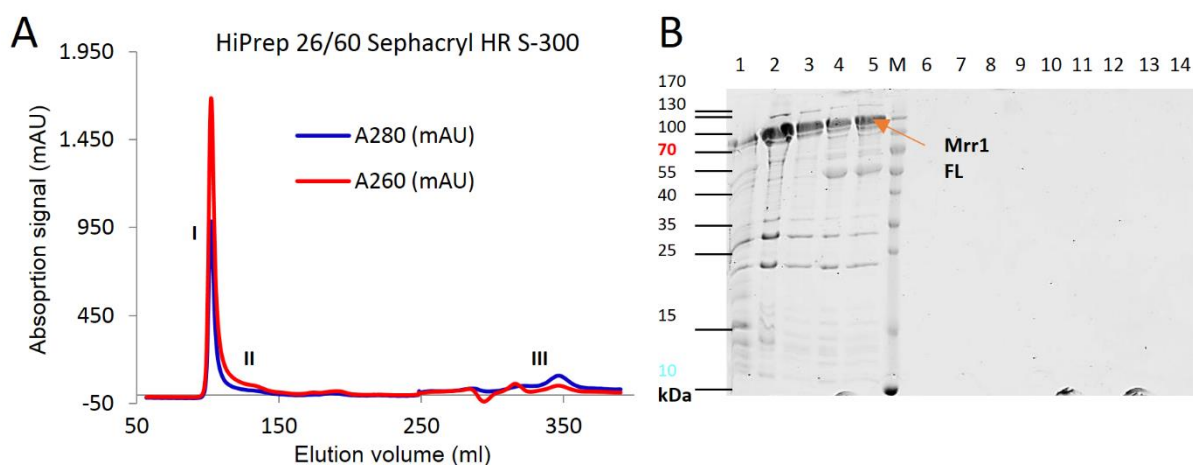


Figure 23 - Size-exclusion chromatography run with a HiPrep 26/60 Sephacryl HR S-300 column:

A) Chromatogram of the size-exclusion chromatography step: absorption at 280 nm (blue) and 260 nm (red)
B) 15% SDS-PAGE of the size-exclusion run: M = marker, 1 = load, 2/3 = concentrated peak I; 4/5 = concentrated peak II; 6-9 = elution fractions between peaks II and III; 10-14 = elution fractions peak III.

Arrows indicate the expected height of Mrr1 FL.

Based on the results from different purification trials, one standard procedure was established, which included an initial affinity chromatography step with Ni-TED resin and subsequently a size-exclusion chromatography step. The affinity chromatography was performed as described for the test runs (4.1.2.1.2) with buffers **6**, **7** and **9**. The SDS-PAGE analysis shows that already quite pure protein is eluted from the resin (fractions 11 to 14, **Figure 24A**). Those fractions were pooled, concentrated and applied to the respective Superdex column for size-exclusion chromatography with an Äkta system. Protein elution was monitored with the help of the absorption at 280 nm and 260 nm and the following SDS-PAGE analysis showed three dominant elution peaks. The first peak (**Figure 24B/C**), near the void volume of the column, mainly contains Mrr1 FL protein, with a $\frac{A_{260}}{A_{280}}$ ratio larger than 1, which again indicates that the protein is forming aggregates. Peak II (**Figure 24B/C**) has a $\frac{A_{260}}{A_{280}}$ ratio that is slightly smaller than 1 and also contains predominantly Mrr1 FL. Peak III cannot be analyzed by SDS-PAGE, as no bands are visible, but due to the size of Mrr1 it is very

Results

unlikely that the protein elutes at such a late stage (~350 ml). The protein containing peaks I and II were separately pooled, concentrated, flash frozen and stored at -80 °C for later use.

All Mrr1 FL purification experiments led to mainly aggregated protein with at least some degradation. The yields were sufficient for subsequent analyses but the aggregation problem, if verified by other techniques such as DLS, would make the protein sample useless for further experiments.

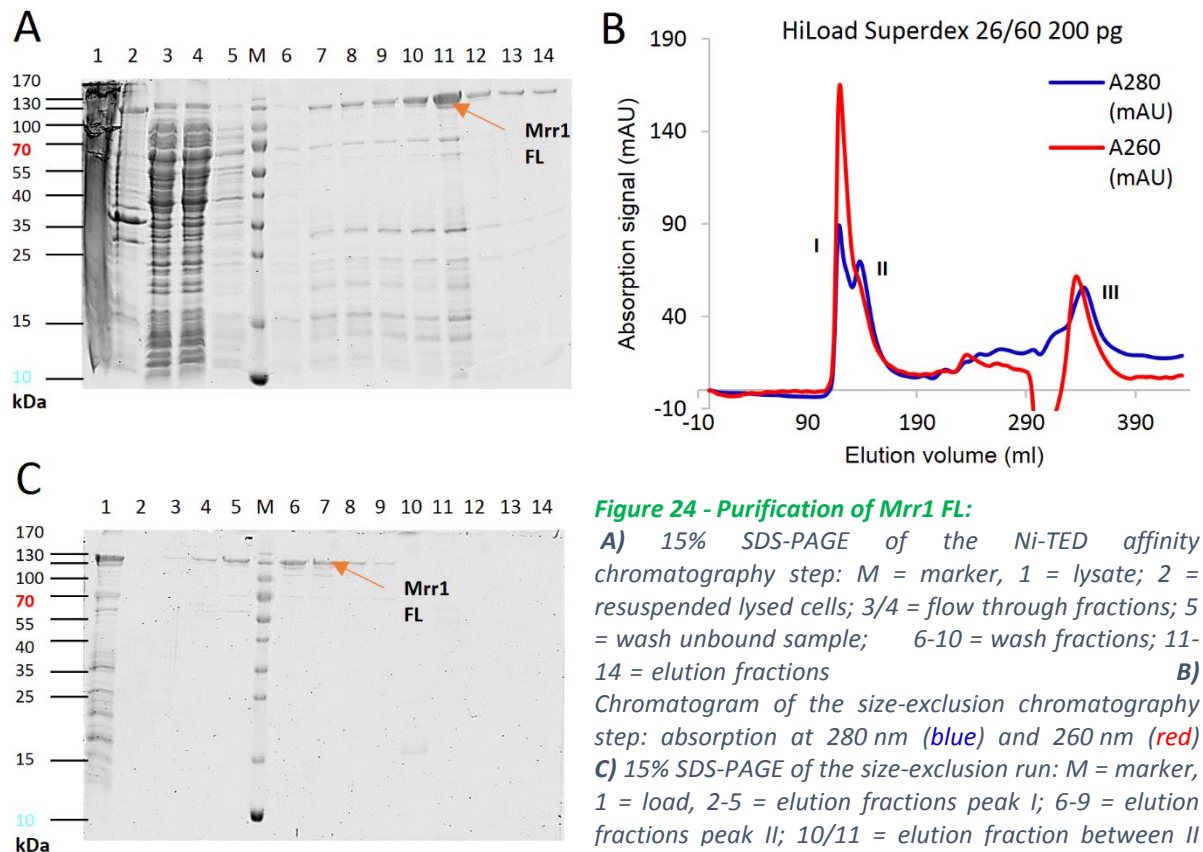


Figure 24 - Purification of Mrr1 FL:

A) 15% SDS-PAGE of the Ni-TED affinity chromatography step: M = marker, 1 = lysate; 2 = resuspended lysed cells; 3/4 = flow through fractions; 5 = wash unbound sample; 6-10 = wash fractions; 11-14 = elution fractions

B) Chromatogram of the size-exclusion chromatography step: absorption at 280 nm (blue) and 260 nm (red)

C) 15% SDS-PAGE of the size-exclusion run: M = marker, 1 = load, 2-5 = elution fractions peak I; 6-9 = elution fractions peak II; 10/11 = elution fraction between II

4.1.2.2 Mrr1 Armadillo repeat

4.1.2.2.1 Expression tests

Prior to large scale expression experiments, small scale expression tests were necessary to ensure that expression of the target protein can be detected. For this purpose, construct 5 was transformed into BL21 (DE3), BL21(DE3)-pLysS, BL21-CodonPlus(DE3)-RIL, Rosetta (DE3) or BL21 Star (DE3) cells. The test culture was supplemented with the respective antibiotics and inoculated with one colony per expression test. After the cell culture was grown to an OD₆₀₀ of 0.6, protein expression was induced by the addition of 0.02% L(+)-arabinose. The temperature was reduced to 30 °C or 16 °C and test samples were taken after 3, 5, 7 h and overnight or 5 and 7 h and overnight respectively. The cells were pelleted and stored at -80°C.

Results

The cells were lysed by sonication (3.2.1.1.2) and the lysate was then incubated with Ni-NTA resin. The purification test was performed as described in 3.2.1.2 and selected samples were in the following analyzed by SDS-PAGE.

The tested BL21(DE3)-pLysS (Figure 25), BL21-CodonPlus(DE3)-RIL and Rosetta (DE3) cells showed a similar expression pattern, with a slightly stronger band in the resuspended pellet fractions at the height of 90 to 95 kDa (left arrow), which is slightly above the expected size of Mrr1 AR with 88.2 kDa

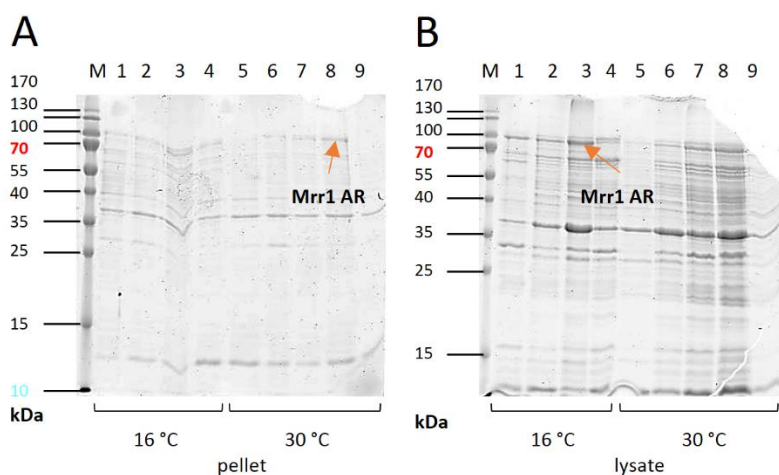


Figure 25 - 15% SDS-PAGE of the Mrr1 AR test expression with BL21(DE3)-pLysS cells:

A) pellet fractions / B) lysate fractions:

M = marker; 1/5 = sample at induction; 2-4 = 5 6-9= 3, 5, 7 h and overnight fractions (30 °C); Arrows indicate the expected height of Mrr1 FL.

(Table 25). The lysate fractions at 16 °C and 30 °C also showed a band slightly higher than the expected protein size (right arrow). The results of the expression tests were promising but due to the small difference of the observed band to the expected band height, larger expressions and purifications were performed to evaluate these results.

4.1.2.2.2 Expression and purification

Large scale expression was performed with BL21(DE3)-pLysS cells. For this purpose the Armadillo repeat construct (5) was transformed into the *E. coli* strain and was grown on LB-agar overnight at 37 °C. A preculture supplemented with the respective antibiotics was inoculated with a single colony from the LB agar plates. The cells were grown at 37 °C until an OD₆₀₀ of 0.6 was reached, then protein expression was induced through the addition of 0.02% L(+)-arabinose. The temperature was reduced to 16 °C and the culture was grown overnight. The cells were harvested by centrifugation and stored at -80 °C.

The cells were resuspended in binding buffer (6, Table 12), homogenized by sonication and finally lysed with the cell disruptor. Initially four different loose resins, Ni-IDA, Ni-NTA, Ni-TED and Co-Talon were tested (Figure 26). The resins were loaded twice with the lysate, washed with binding buffer and subsequently with buffer containing increasing imidazole concentrations (7). Elution was performed with two different buffers containing 150 mM and 250 mM imidazole (9). The load, wash and elution fractions were analyzed by SDS-PAGE. The analysis of all performed tests did not indicate

Results

a clear prominent band at the expected height. For Ni-TED or Co-Talon (**Figure 26A/B**) none of the elution bands is assumed to be Mrr1 AR. The Ni-IDA purification (**Figure 26C**) revealed two bands at approx. 95 kDa (orange arrow) and 100 kDa that might indicate the presence of the protein of interest. The Ni-NTA purification (**Figure 26D**) in general was not very successful since a lot of different proteins were eluted but one band in the wash fractions (5 to 7, orange arrow) might be Mrr1 AR. Elution and wash fractions of the Ni-IDA purification were pooled, concentrated with Vivaspin concentrators (MWCO 50,000 kDa) and applied to size-exclusion chromatography.

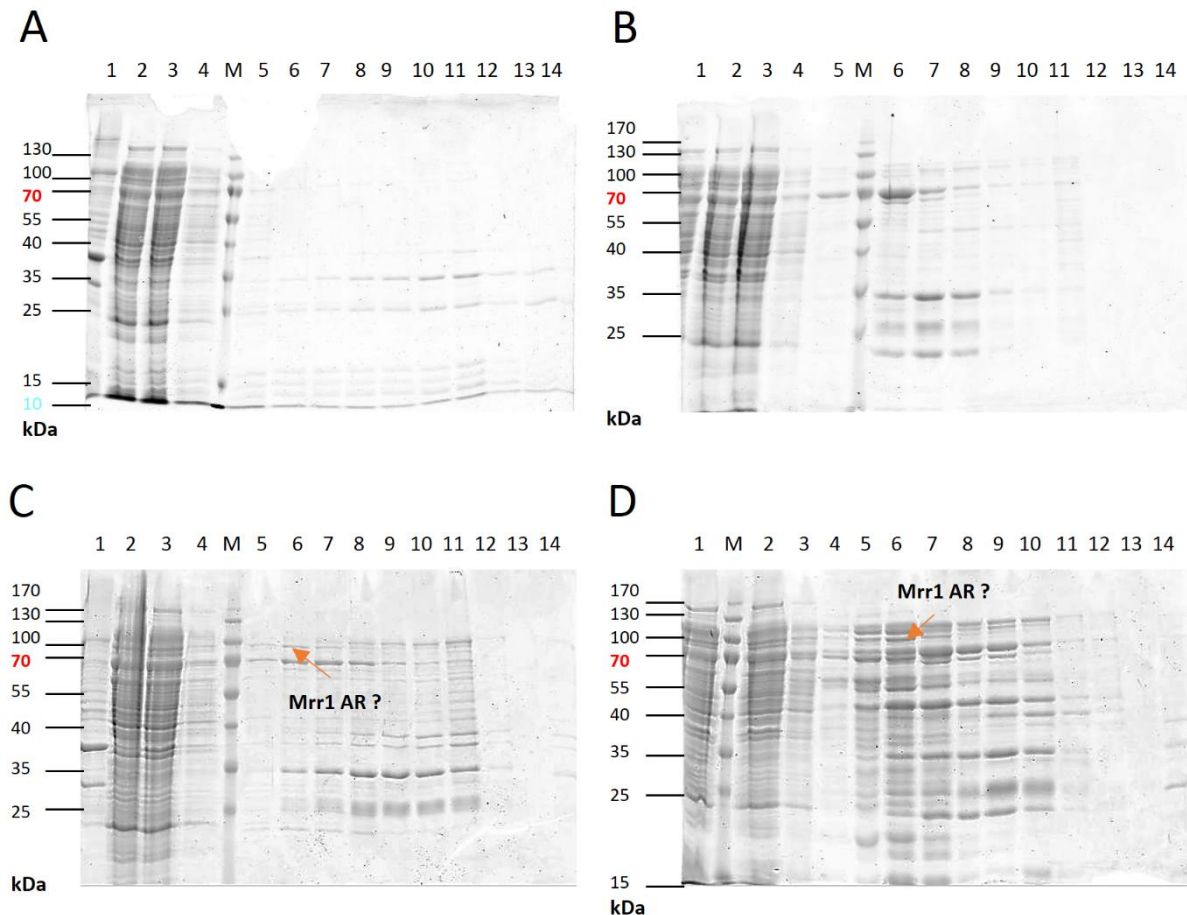


Figure 26 - 15% SDS-PAGE of the affinity chromatography purification tests of Mrr1 AR:

A) Ni-TED resin: M = marker; 1 = resuspended lysed cells; 2/3 = flow through fractions; 4 = wash unbound sample; 5-9 = wash fractions; 10-14 = elution fractions **B)** (Co-Talon): M = marker; 1 = lysate; 2/3 = flow through fractions; 4 = wash unbound sample; 5-9 = wash fractions; 10-14 = elution fractions **C)** Ni-IDA: M = marker; 1 = resuspended lysed cells; 2/3 = flow through; 4 = wash unbound sample; 5-9 = wash fractions; 10-14 = elution fractions **D)** Ni-NTA: M = marker; 1/2 = flow through; 3 = wash unbound sample; 4-8 = wash fractions; 9-12/14 = elution fractions; 13 = empty; **Arrows** indicate the expected height of Mrr1 AR.

Size-exclusion chromatography was tested as the second purification step with a Superdex 200 10/300 GL column (GE Healthcare) at an Äkta system with buffer **33**. Protein elution was monitored at 280 nm and 260 nm followed by SDS-PAGE analysis of the eluted fractions. The elution profile (**Figure 27A**) shows three peaks out of which peaks I and II showed a $\frac{A_{260}}{A_{280}}$ ratio above 1 which indicates that the proteins in these peaks form aggregates. The SDS-PAGE of the peaks (**Figure 27B**) does not show a (prominent) band at the expected protein size and therefore Mrr1 AR is likely not

Results

present in these fractions. Peak III shows a $\frac{A_{260}}{A_{280}}$ ratio smaller than 1 but the SDS-PAGE shows again no band at the expected protein size of 88 kDa.

The expression and purification results indicate that Mrr1 AR was not overexpressed or only a degradation product was present and could not be purified with the tested methods. Mass spectrometry of excised SDS-PAGE bands was not available at the time when the experiments were performed. Therefore the nature of the bands remained elusive. In addition, yield and purity were not promising and it was questionable if reasonable amounts of protein for the intended experiments would have been possible to obtain.

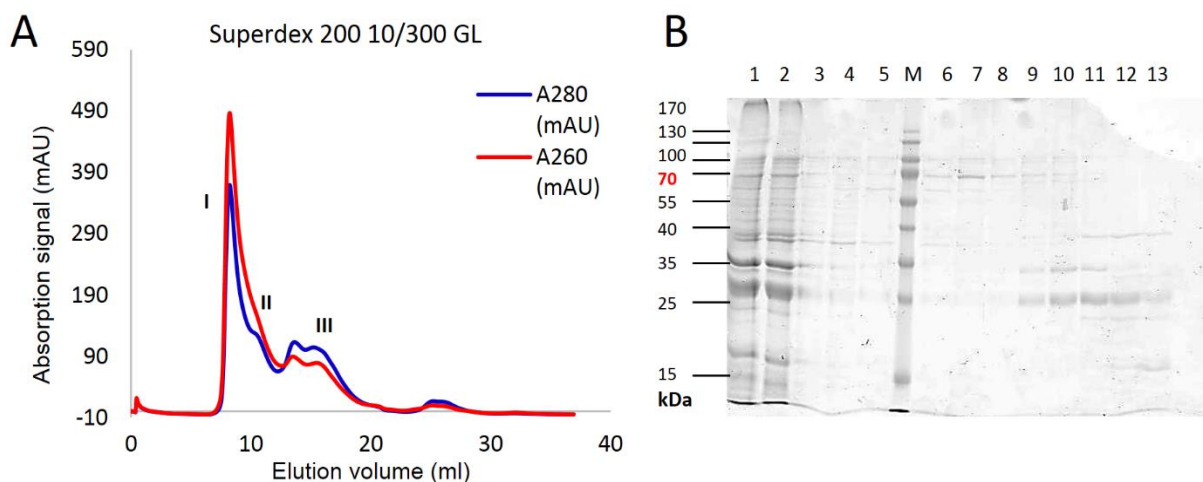


Figure 27 - Size-exclusion chromatography with a Superdex 200 10/300 GL column:

A) Chromatogram of the size-exclusion chromatography step: absorption at 280 nm (blue) and 260 nm (red)
B) 15% SDS-PAGE after size-exclusion chromatography: M = marker, 1 = load, 2-5 = elution fractions peak I; 6-8 = elution fractions peak II; 9-13 = elution fractions peaks III.

4.1.2.3 Mrr1 '250' and variants

4.1.2.3.1 Expression tests

Prior to large scale expression experiments, small scale expression tests were necessary to ensure that expression of the target protein can be detected. For this purpose, construct **6** was transformed into BL21 (DE3), BL21(DE3)-pLysS and BL21-CodonPlus(DE3)-RIL cells. The test culture was supplemented with the respective antibiotics and inoculated with one colony per expression test. After the cell culture was grown to an OD_{600} of 0.6, protein expression was induced by the addition of 0.02% L(+)-arabinose. The temperature was reduced to 30 °C or 16 °C and test samples were taken after 3, 5, 7 h and overnight or 5 and 7 h and overnight respectively. The cells were pelleted and stored at -80°C.

The test purification was performed as described above. Samples from the protein bound to the resin and resuspended pellet fractions were analyzed by SDS-PAGE analysis.

Results

BL21 (DE3) cells transformed with the vector containing the Mrr1 250 construct expressed the protein of interest at relatively high amounts (Table 25) but most of it was present in the cell pellet and hardly any soluble protein expression was observed (not shown). In BL21-CodonPlus(DE3)-RIL cells the results are similar but overall less protein was expressed (not shown). In contrast, in BL21(DE3)-pLysS cells a slight overexpression band (Figure 28A, arrow) can be observed for the insoluble pellet fractions but also a distinct overexpression band for the lysate fractions (Figure 28B, arrows). A remarkable observation was also obtained through the time course. Most likely the protein was degraded when it was expressed overnight (16 °C) or even after 7 h at 30 °C, which can be seen in lanes 9 and 5 respectively. Nonetheless, the results of the test expression were promising with respect to the soluble expression of this 859 amino acids long Mrr1 variant.

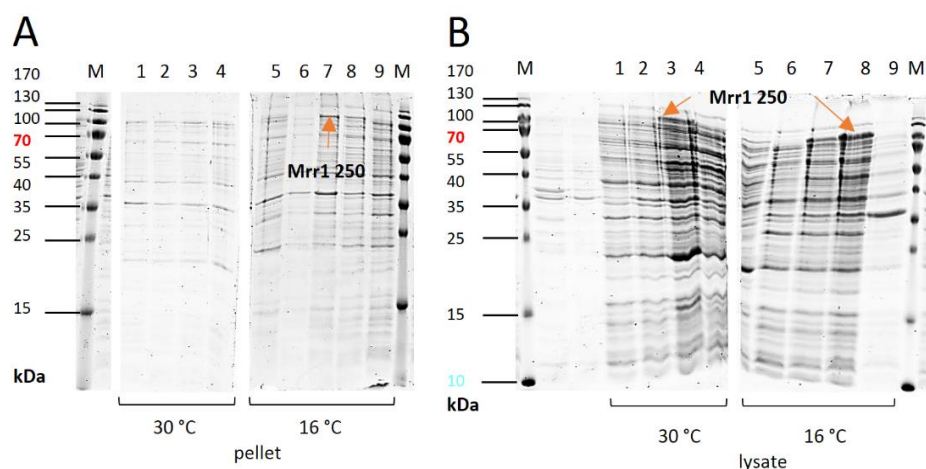


Figure 28 - 15% SDS-PAGE of the Mrr1 250 test expression with BL21(DE3)-pLysS cells:

A) pellet fractions / **B)** lysate fractions: M = marker; 1/5 = sample at induction; 2-5 = 3, 5 6-9= 5, 7 h and overnight fractions (16 °C); Arrows indicate the expected height Mrr1 '250'

4.1.2.3.2 Thermofluor analysis

Initial purification results (not shown) already indicated that Mrr1 '250' in contrast to the full length protein is not aggregating. Nevertheless, thermofluor experiments were performed to elucidate the stability of the protein in the initial size exclusion buffer (33, Table 12) and to identify agents or buffers that might enhance its stability in solution. The initial screens indicated a melting temperature of approx. 50 °C, which is slightly lower than measured for the Full length protein, but especially buffer substances as Ada-NaOH, Pipes-NaOH or Bicine-NaOH at different pH values indicate a clear stabilization potential (Figure 29A).

Results

The Ada-NaOH buffers were subsequently used, but as shown for other proteins this buffer is prone to display false positives, so that also other buffers as Bicine were investigated in further screening experiments. Follow up screens with the help of compounds of the Additive Screen and the Silver Bullets Screen (both Hampton Research, [Table 15](#)) and around derived stabilizing agents were performed. The experiments indicate high stabilization potential for KCl, NaF, K/Na tartrate, Na citrate or CsCl. The physiological salt KCl was tested in different concentrations as replacement for NaCl with different buffer compounds and for the optimal crystallization buffer composition, which should contain the buffer substance in low molarities, as 20 or 30 mM. Ada-NaOH and Bicine-NaOH with 300 mM KCl at pH values of 6.5, 7.0 and 8.0 show good stabilization properties ([Figure 29B](#)). Based on the thermofluor experiments and the known ability of Ada-NaOH to display a false positive, 20 mM Bicine-NaOH pH 8.0 with 300 mM KCl is chosen as buffer for size-exclusion runs. As the necessity of an additional, intermittent purification step came into focus, buffer combinations with Bicine-NaOH at different pH values were tested and anion-exchange buffers containing 50 mM Bicine-NaOH pH 8.5 and 50 mM or 1 M KCl ([Table 12](#)) were established.

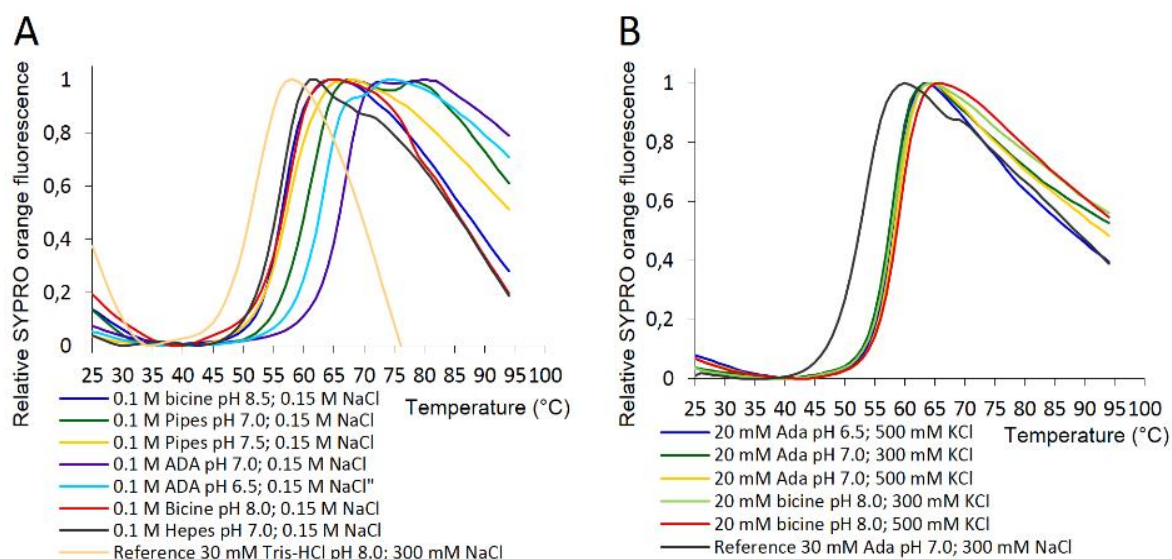


Figure 29 - Thermofluor experiments with Mrr1 '250':

A) Standard buffer thermofluor screen of the Mrr1 '250' protein. The initial reference buffer was buffer **33** ([Table 12](#); 50 mM Tris-HCl pH 8.0, 300 mM NaCl) **B)** Thermofluor fine screen around the buffer hit conditions from A) and a pursued additive screen, to elucidate the optimum buffer and KCl concentrations Reference buffer was buffer **34** ([Table 12](#); 30 mM Ada-NaOH pH 7.0, 300 mM NaCl)

4.1.2.3.3 Expression and purification of the wild type protein

Large scale expressions were performed with BL21(DE3)-pLysS cells. The '250' construct (**6**) within the pBadm11 vector was transformed into the cells and was grown on LB agar plates overnight at 37 °C. A preculture, supplemented with the respective antibiotics, was inoculated with a single colony from the LB agar plates. The cells were grown at 37 °C until an OD₆₀₀ of 0.6 was reached; then protein

Results

expression was induced by the addition of 0.06% L(+)-arabinose. The temperature was reduced to 30 °C and the culture was grown for another 6 h, then the cells were harvested by centrifugation and stored at -80 °C.

The cells were resuspended in binding buffer (**22**, **Table 12**), homogenized by sonication and subsequently with the cell disruptor. Initially three different loose resins, Ni-IDA, Ni-NTA and Ni-TED, and a prepacked column (Ni-MAC) were tested (not shown). The tests were performed as described for the Full length protein but with buffers **22-24** or **18/19** respectively (**Table 12**). The load, wash and elution fractions were analyzed by SDS-PAGE. With all resins used a strong protein band in the elution fractions at the expected protein height was observed. The best results, with the highest purity, were obtained with Ni-NTA resin and therefore this resin was used for all future affinity chromatography experiments when this construct was purified.

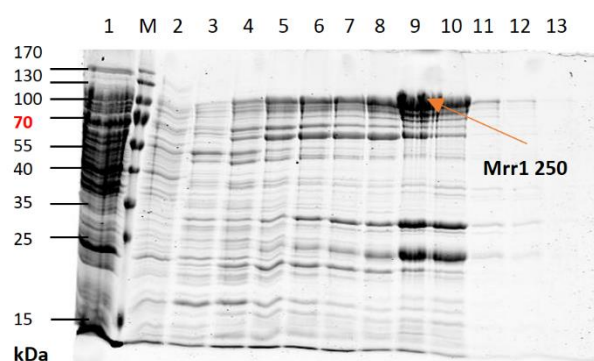


Figure 30 - 15% SDS-PAGE of Mrr1 '250' after affinity chromatography with Ni-NTA:

M = marker; 1 = flow through; 2 = wash unbound sample; 3-8 = wash fractions; 9-13 = elution fractions. The arrow indicates the expected height of the target protein.

The Ni-NTA purification was established with the following protocol utilizing buffers **22-24**. The protein was loaded overnight onto the column under permanent circulation of the lysate. The resin was washed several times with buffer containing increasing imidazole concentrations. The protein was eluted with elution buffers containing 150 or 250 mM imidazole respectively. The collected fractions were analyzed by SDS-PAGE, which showed a clear elution peak in elution fraction I, but also the other elution fractions contained Mrr1 '250' (**Figure 30**). The elution fractions were pooled and concentrated with Vivaspin concentrators (MWCO 50,000) to approx. 1ml.

Although Mrr1 '250' can be overexpressed (**Figure 30**), dominant impurities with sizes of ~65, 35 and 25 kDa, were still present after the first purification step. Initial size-exclusion experiments did not lead to a significant reduction of these impurities. Therefore, an anion-exchange chromatography step was included into the purification procedure. Initial scouting with three different buffers, a MonoQ 5/50 GL column (GE Healthcare) and with the help of an Äkta system, to monitor the absorption and to optimize the high salt buffer gradient, the best buffer composition and gradient could be identified. The final anion-exchange chromatography protocol therefore consists of an elution that is subdivided into three gradients and one step and is performed with buffers **29** and **30** (3.2.1.3, **Table 12**). The elution was monitored at 280 and 260 nm and the chromatogram (**Figure 31A/B**) showed that the elution profile contained two main peaks I and II, with an additional small

Results

peak in between. SDS-PAGE analysis revealed that peak I mainly contains Mrr1 '250' with some minor impurities (Figure 31B). Large amounts of the above mentioned impurities were thus separated from Mrr1 and mainly eluted in peak II (Figure 31B). The Mrr1 '250' containing protein fractions with the least impurities were pooled, concentrated and applied to the His₆-tag cleavage reaction.

The cleavage was performed with TEV protease in buffer 38 (Table 12) at 4 °C overnight. After cleavage, the protein sample was diluted with binding buffer in a ratio of 1:9 and was applied to a reverse Ni-chromatography step. The sample was loaded once onto the equilibrated Ni-NTA resin, the flow through was collected and eventually bound protein eluted from the column in several wash and elution steps. The fractions were analyzed by SDS-PAGE (not shown) and the Mrr1 '250' containing flow through fraction was concentrated and applied to the size exclusion column.

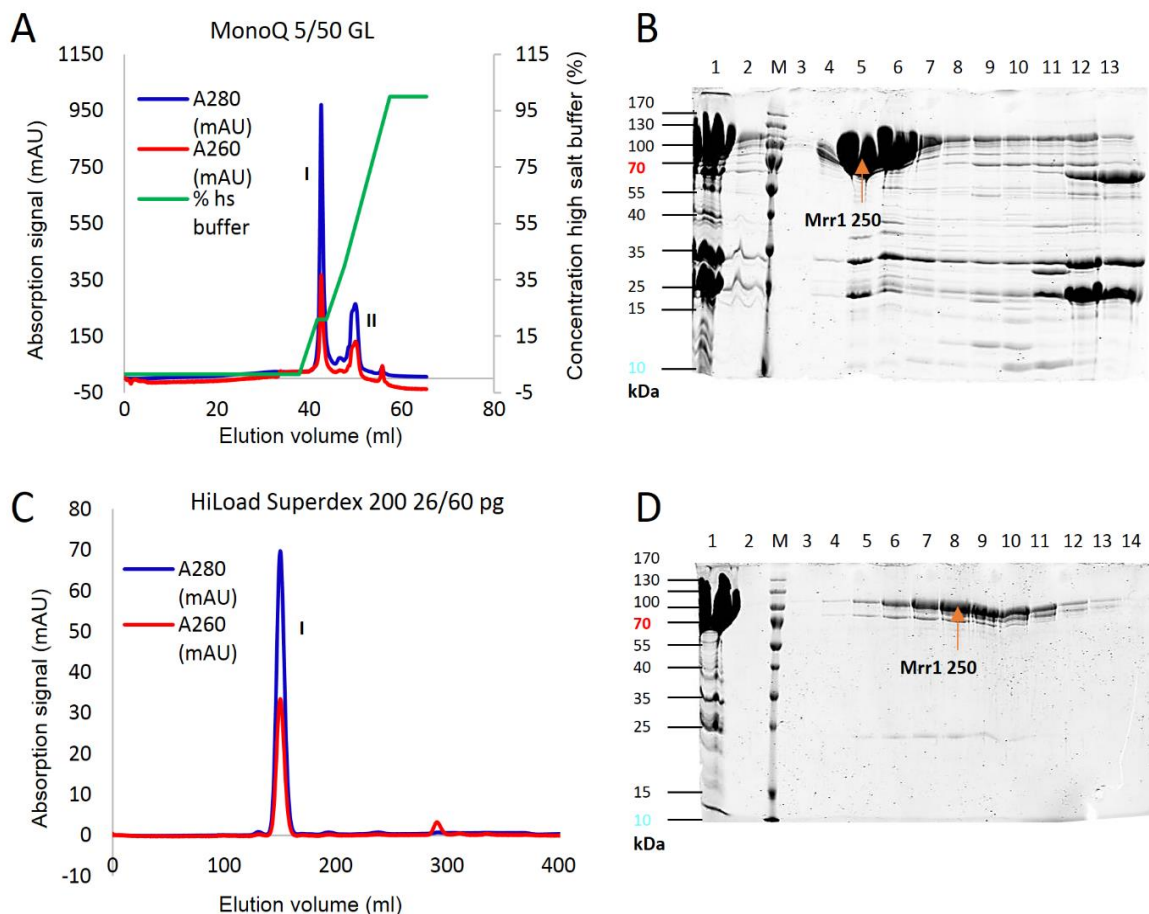


Figure 31 - Purification of Mrr1 '250':

A) Chromatogram after anion-exchange chromatography: absorption at 280 nm (blue) and 260 nm (red); % high salt buffer (green) **B)** 15% SDS-PAGE after anion-exchange chromatography: M = marker, 1 = load; 2 = flow through, 3 = wash unbound sample, 4-7 = elution fractions peak I; 8-10 = fractions between peaks I and II; 11-13 = elution fractions peak II **C)** Chromatogram after size-exclusion chromatography: absorption at 280 nm (blue) and 260 nm (red) **D)** 15% SDS-PAGE after size-exclusion chromatography: M = marker, 1 = load, 2-14 = elution fractions peak I.

Arrows indicate the expected height of the target protein.

Results

Size-exclusion was performed with the HiLoad Superdex 200 26/60 µg or with Superdex 200 (Increase) 10/300 GL columns (GE Healthcare) and an Äkta system. Protein elution was monitored at 280 nm and 260 nm and the collected fractions were analyzed by SDS-PAGE (**Figure 31C/D**). Analysis of peak I shows that Mrr1 '250' is very pure but unfortunately double or triple bands can be observed indicating protein degradation. Mass spectrometry analysis and western blot experiments with an α His-antibody confirmed this hypothesis. The bands contained amino acids from the Mrr1 '250' construct in different lengths. Degradation most likely took place from the C-terminus since the degradation bands were also visible in the western blot and thus the N-terminally located His-tag must have been present.

The protein containing peaks were pooled, concentrated, flash frozen and stored at -80 °C for later use.

4.1.2.3.4 Expression and purification of the K335N, P683S, G878E and G997V Mrr1 '250' variants

Expression and purification was performed as described for the WT protein, therefore only the final size-exclusion elution profiles and the SDS-PAGE analysis of the different variants will be shown.

Purity and protein stability of all variants differ in comparison to each other and for some variants also throughout the different purifications. Protein stability was especially affected by degradation processes that were mainly observed for the G to E/V variants.

The highest homogeneity was observed for the purification of the K335N variant. Elution patterns and SDS-PAGE analysis show similar profiles, degradation behavior and purity as observed for the WT protein. In size-exclusion chromatography one single peak, I, was obtained (**Figure 32A**), which showed some tailing, but the protein degradation, that is causing this, is comparably low and comparable to the WT protein (**Figure 32B**). This one peak elution pattern was observed for all three performed purifications of the K335N variant.

It was possible to purify the P683S variant (**Figure 32C/D**), but it was more difficult to handle than the WT protein or the K335N variant. Although the purification protocol did not differ, degradation was observed (**Figure 32D**), starting already after the Ni-affinity chromatography step. The size-exclusion profile was sometimes as homogeneous as for the K335N variant but more often one prominent peak (I) plus additional peaks or tailing (peak II) were observed (**Figure 32C**). The SDS-PAGE after size-exclusion chromatography, clearly showed a prominent double band, with higher intensities for the lower (~80 kDa) band.

Results

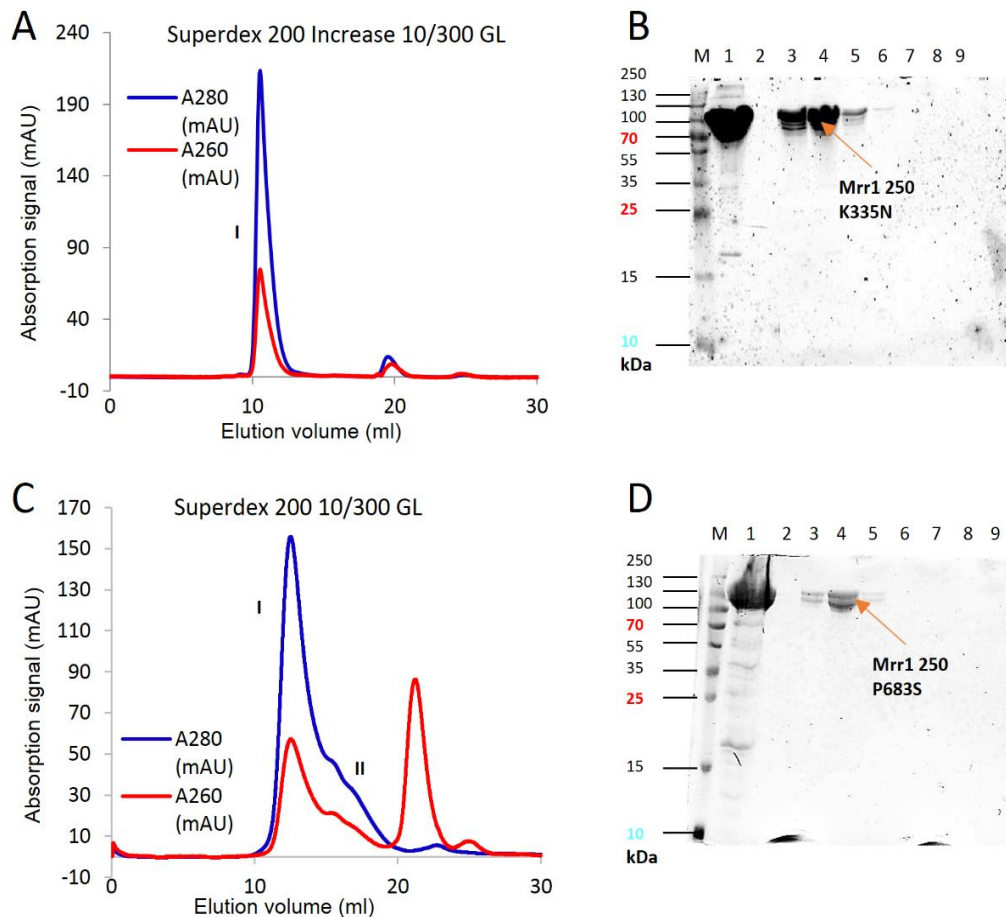


Figure 32 - Size-exclusion elution profiles and SDS-PAGE analyses of Mrr1 '250' K335N and P683S:

A) Chromatogram of the size-exclusion chromatography step of the K335N variant: absorption at 280 nm (blue) and 260 nm (red) **D)** 15% SDS-PAGE of the size-exclusion step: M = marker, 1 = load, 2-7 = elution fractions peak I; 8/9 = elution fractions after peak I; **C)** Chromatogram of the size-exclusion chromatography step of the P683S variant: absorption at 280 nm (blue) and 260 nm (red) **D)** 15% SDS-PAGE of the size-exclusion step: M = marker, 1 = load, 2-5 = elution fraction of peak I; 6-9 = elution fractions of peak II. Arrows indicate the expected height of the target protein.

Both G variants showed differing purification patterns, either protein with hardly any degradation was obtained, or intense degradation bands were already visible after the initial purification step. The elution profiles and SDS-PAGE analyses shown in **Figure 33** represent the observed purification results. For the G878E variant, it was once possible to obtain a pure and homogenous sample (**Figure 33A/B**). The same results were obtained for G997V, which can be purified to relatively high homogeneity. In contrast to a successful G997V protein purification, with hardly any degradation, heavy degradation and an inhomogeneous elution and SDS-PAGE profile was observed for the G997V protein, too (**Figure 33C/D**). An extreme example is shown in **Figure 33D** where the protein elution profile revealed that a prominent peak (I) is present in the void volume and the analysis of the protein containing peak II showed heavy degradation with low percentages of the full length '250' protein variant. The G878E variant once also showed a similar degradation behavior. Purification of the two G to E/V variants also led to lower yields compared to the two other variants, which was also

Results

caused by the necessity to more strictly remove fractions containing heavier degradation bands after anion-exchange chromatography (not shown).

Dependent on the impurities contained in the elution fractions, differing amounts of the protein variants were pooled, concentrated, flash frozen and stored at -80 °C for later use.

Despite the quite prominent degradation problem, several gain of function variants were utilized for subsequent biochemical experiments and partially for crystallization approaches.

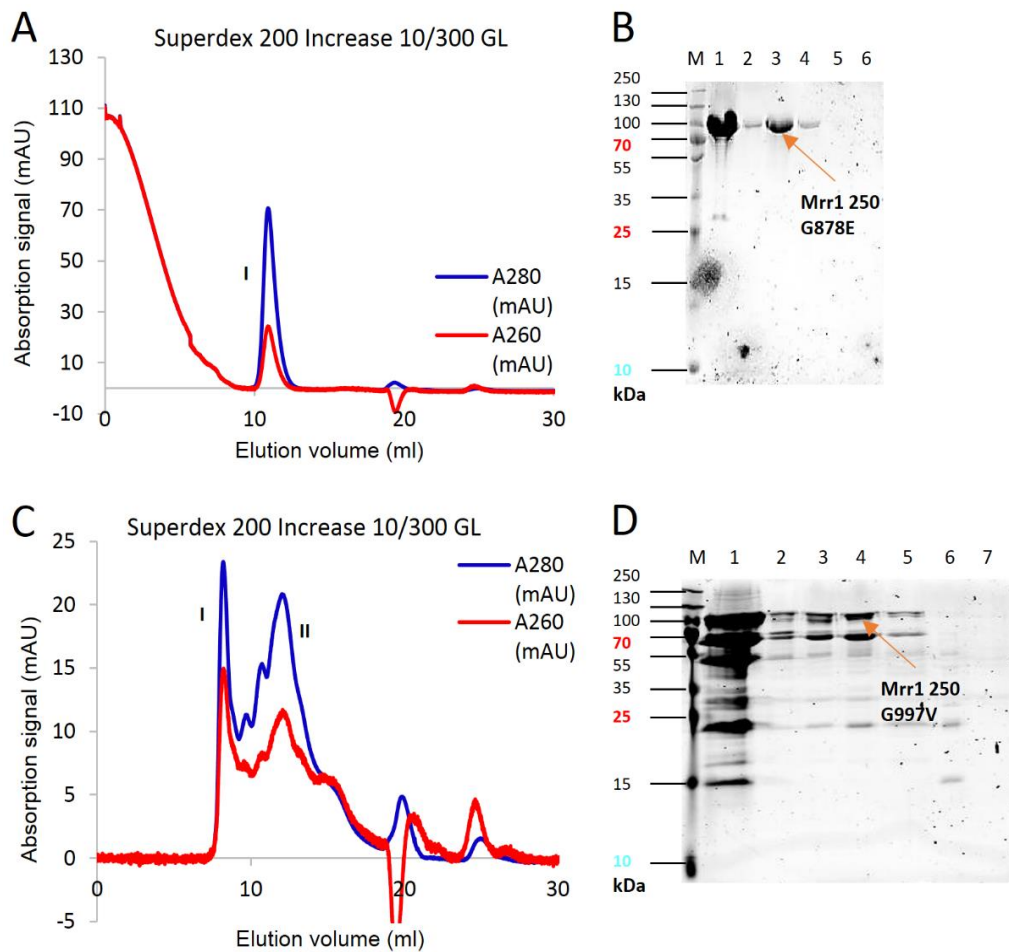


Figure 33 - Size-exclusion elution profiles and SDS-PAGE analyses of the Mrr1 '250' G878E and G997V variants:

A) Chromatogram of the size-exclusion chromatography step of the G878E variant: absorption at 280 nm (blue) and 260 nm (red) **D)** 15% SDS-PAGE after size-exclusion chromatography: M = marker, 1 = load, 2-4 = elution fractions peak I; 5/6 = elution fractions after peak I; **C)** Chromatogram of the size-exclusion chromatography step of the G997V variant: absorption at 280 nm (blue) and 260 nm (red) **D)** 15% SDS-PAGE after size-exclusion chromatography: M = marker, 1 = load, 2 = elution fraction of the tailing of peak I; 3-5 = elution fractions of pre-peak II and peak II; 6/7 = elution fractions of the tailing of peak II. Arrows indicate the expected height of the target proteins running.

Results

4.1.2.3.5 Expression and purification of the Mrr1 '250' wild type construct with a C-terminal His₆-tag

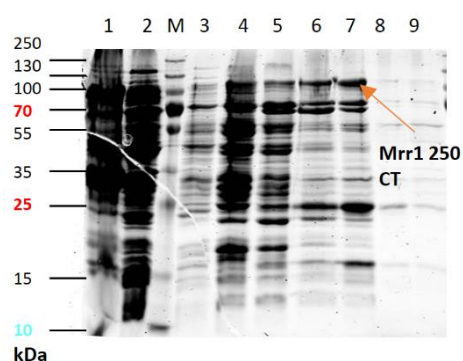


Figure 34 - 15% SDS-PAGE after Ni-NTA affinity chromatography of the Mrr1 '250'CT construct:

Ni-NTA resin: M = marker; 1 = resuspended cell pellet; 2 = flow through; 3 = wash unbound sample; 4-6 = wash fractions; 7-9 = elution fractions; **Arrow** indicates the expected height of the target protein.

Large scale expression was performed in Rosetta (DE3) cells. The '250'CT construct (**20**) in the pPETm11 vector was transformed into the cells and was grown on LB agar overnight at 37 °C. A preculture, supplemented with the respective antibiotics, was inoculated with a single colony from the LB agar plates. The cells were grown at 37 °C until an OD₆₀₀ of 0.6 was reached; then protein expression was induced by the addition of 1 mM IPTG. The temperature was reduced to 16 °C, the culture was grown overnight and the cells were harvested by centrifugation and stored at -80 °C.

The cells were resuspended in binding buffer (**42, Table 12**), homogenized by sonication and subsequently lysed with the cell disruptor. The lysate was loaded overnight onto Ni-NTA resin and subsequently washed several times with buffer containing increasing imidazole conditions. The elution was performed with buffers containing 100 or 250 mM imidazole respectively. Fractions were analyzed by SDS-PAGE. The analysis showed that the protein was already eluting at the end of the wash procedure (lanes 5/6, **Figure 34**) and in the first elution fraction whereas all the following fractions hardly contained our target protein (lanes 8/9, **Figure 34**). The last wash fraction and the elution fractions were pooled and concentrated with Vivaspin concentrators (MWCO 50,000) to approx. 1ml. Overall, Mrr1 '250' CT was overexpressed but not with a high yield. In addition, impurities around 70 kDa and at 25 and 15 kDa protein size were observed which are most likely due to the early elution of the protein and its weak interaction with the affinity resin (**Figure 34**).

To reduce the impurities, anion-exchange chromatography was performed as described for the Mrr1 '250' constructs with the N-terminal His₆-tag. The elution was monitored at 280 nm and 260 nm, using an Äkta system. SDS-PAGE analysis (**Figure 35A/B**) of the two main peaks, I and II, revealed that Peak I mainly contained Mrr1 '250' CT with an increasing number of impurities corresponding to the tailing in the Mrr1 elution profile (**Figure 35B**). The impurities at a height of ~70 kDa can be observed from peak I to peak II. To avoid the impurities only fractions from lanes 2 and 3 of the SDS-PAGE were pooled, but this led to a yield of the protein which was excessively diminished. The C-terminal His₆-tag is not cleavable, therefore the TEV cleavage step cannot be performed. The pooled fractions were directly applied onto a Superdex 200 (Increase) 10/300 GL column (GE Healthcare). Size-exclusion chromatography was performed with an Äkta system, the protein elution was monitored at

Results

280 nm and 260 nm and the collected fractions were analyzed by SDS-PAGE (Figure 35C/D). Analysis of the main peak, peak I, showed that Mrr1 '250' CT is very pure but also shows some degradation was observed; the yield however is low. Peak II only contained impurities. The protein containing fractions were pooled, concentrated, flash frozen and stored at -80 °C for later use.

The resulting low yields, obtained for this construct, hamper broad range crystallization experiments.

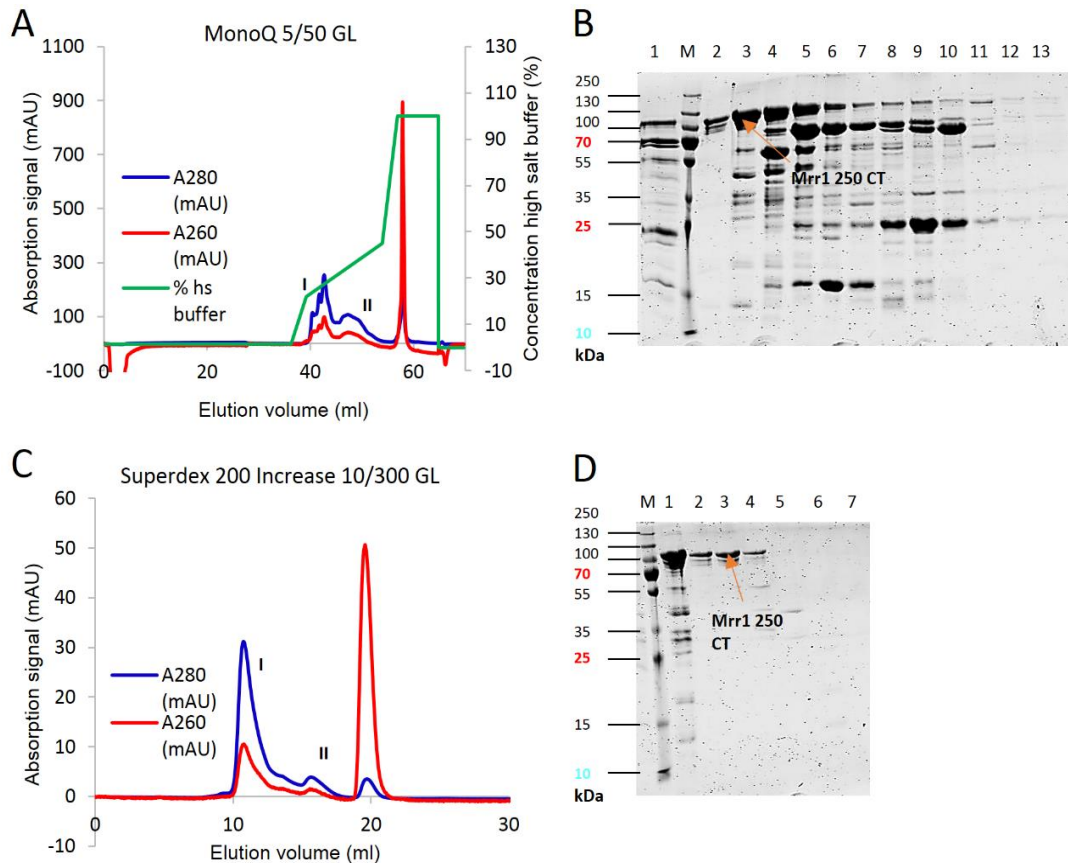


Figure 35 - Purification of the Mrr1 '250' CT construct:

A) Chromatogram of the anion-exchange chromatography step: absorption at 280 nm (blue) and 260 nm (red); % high salt buffer (green) **B)** 15% SDS-PAGE after anion-exchange chromatography: M = marker, 1 = load; 2-7 = elution fractions peak I; 8-13 = elution fractions peak II **C)** Chromatogram of the size-exclusion chromatography step: absorption at 280 nm (blue) and 260 nm (red) **D)** 15% SDS-PAGE after size-exclusion chromatography: M = marker, 1 = load, 2-4 = elution fractions peak I; 5 = elution fraction between the two peaks; 6-7 = elution fractions peak II.

Arrows indicate the expected height of the target protein.

4.1.3 Expression tests in *P. pastoris*

The expression and purification approaches of Mrr1 in *E. coli* showed that only Mrr1 '250' can be expressed and purified without aggregation in reasonable amounts, which are suitable for biochemical analyses and crystallization experiments. The aim, however, was the structural characterization of the complete Mrr1 construct and the structural comparison of the wild type protein with the gain of function variants.

Results

To overcome the expression, purification and aggregation problem of the *E. coli* system, initially three constructs (**Figure 19**) were cloned into the three pichia vectors, FL, ARlong and '250'. The verified constructs were genetically integrated into the four available *P. pastoris* strains (1 to 4) as described in 3.1.4.3. Successful construct integration was first identified by the creamy white colony color and later also partially verified by colony PCR (3.1.6.5). Colonies that did not contain the integrated gene were significantly smaller, required longer times to grow and showed a prominent pink coloring.

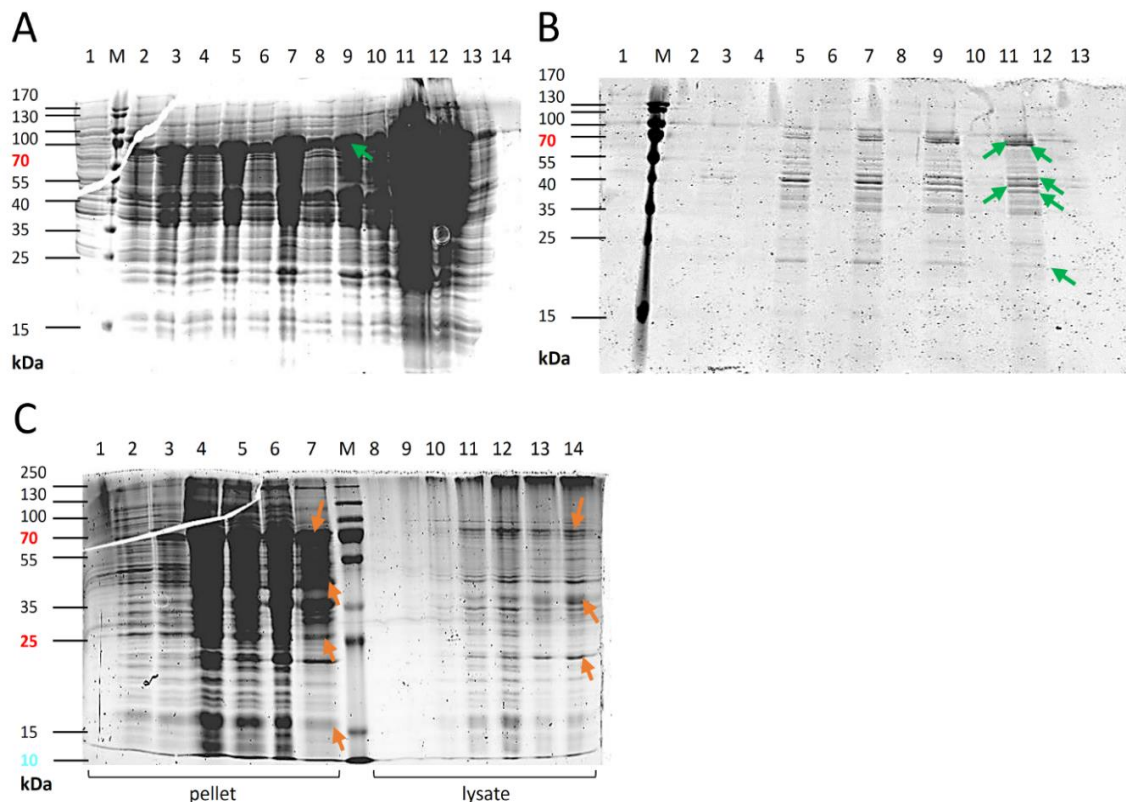


Figure 36 - 15% SDS-PAGE analyses of the test expressions in *P. pastoris*:

A/B) Test of secreting protein expression of Mrr1 '250' in pichia strain 1: A = pellet fractions; B = culture supernatant; M = marker; 1 = time of induction; 2/3 = 18 h sample; 4/5 = 24 h sample; 6/7 = 40 h sample; 8/9 = 66 h sample; 10/11 = 90 h sample; 12/13 = 110 h sample; even numbers: with additional induction; odd numbers: initial induction with methanol only **C)** Test of non-secreting protein expression of Mrr1 FL (pPinkHC) in pichia strain 1: samples 1-7 resuspended pellet fractions and 8-14 lysate fractions; M = marker; 1/8 = time of induction; 2/9 = 3 h sample; 3/10 = 7 h sample; 4/11 = 20 h sample; 5/12 = 26 h sample; 6/13 = 52 h sample; 7/14 = 76 h sample. The orange arrows indicate protein bands, which are running at a similar height as the ones marked in **Figure 37B**.

Prior to large scale expression experiments, small scale expression tests were performed. All constructs, in the secreting vector pPink α HC, were successfully integrated into all available pichia strains. The ARlong construct was not used for expression experiments with the non-secreting pPinkHC/LC vector system. As no protein expression was observed with the secreting vector system, the expression with the non-secreting vectors was mainly focused on Mrr1 FL and partially on Mrr1 '250'. Mrr1 FL in pPinkHC or LC was attempted to be integrated into all four strains, which was

Results

successful with the exception of strain 3, where the integration of the high copy plasmid failed. For Mrr1 '250' in pPinkHC or LC integration failed into strain 1 (**Table 28**).

For all positive integrants, expression tests (3.1.7.3) were performed. Two expression cultures per construct and strain were cultivated for the four *pichia* strains containing the secreting vectors. One of the cultures was twice induced with methanol, one time by the transfer of the grown yeast cells from BGM medium into BMMY inducing medium and a second time 3 days post the first induction. The second culture was regularly, every 24 h, supplemented with methanol after the initial induction. These two induction procedures were used, to check if and how protein expression is influenced by low methanol levels in comparison to regular methanol provision. During the expression test, samples were taken on a regular basis, to monitor the expression of the proteins of interest and the appearance of impurities. In addition to the experiments with the Mrr1 constructs, tests with empty vectors or without induction were performed to analyze and compare the intrinsic protein expression.

Analysis of the expression pattern utilizing the secreting vectors showed significant bands at approx. 65, 40 and 37 kDa, especially in the pelleted cells. These bands can be observed independently from the integrated Mrr1 construct and are present in all samples. Two exemplary SDS-PAGEs for Mrr1 '250' integrants are shown in **Figure 36A/B**. Gel A displays the expression pattern over time for the harvested cells and gel B contains the fractions taken from the culture supernatant in the same time frame. The observed expression pattern is hardly discernable for all constructs in any of the strains. It is however interesting to note, that the expression patterns differ between the twice induced cultures in comparison to the regularly with methanol supplied cultures. The expression of some proteins is triggered or induced by the regular addition of methanol, (see lanes 3, 5, 7, 9, 11, compared to the same cultures that were supplemented with methanol only twice). The prominent bands cannot be annotated to one of the constructs, as the running height is smaller than expected for all Mrr1 variants. Mass spectrometry analyses from excised bands (**Figure 36B**, green arrows) confirmed that the expressed proteins were not corresponding to Mrr1 or degraded Mrr1 protein fragments.

Expression tests with the non-secreting vectors were performed with the Mrr1 FL constructs in all strains and Mrr1 '250' in strain 1. Mrr1 FL was pursued more intensely because the full-length protein is also of special interest for DNA interaction studies as well as co-crystallization approaches and it was not possible to obtain functional full-length protein via *E. coli* expression. Similar to the results obtained for the strains containing the secreting vectors, comparable expression patterns were observed as shown exemplarily in **Figure 36C** of Mrr1 FL in the high copy vector. The band pattern showed high similarities to the results obtained in the secreted protein expression

Results

experiment. Prominent bands at approx. 65, 40 or 37 kDa are visible in the resuspended cells and in the lysate. The lysate was incubated with Ni-NTA beads and washed several times to reduce the background and thereby enrich the protein of interest. These fractions show similar bands as observed in the SDS PAGE of the secreted system (Figure 36) that were analyzed by mass spectrometry and did not contain Mrr1 (Figure 36B). Therefore it is very unlikely that the expressed proteins are Mrr1.

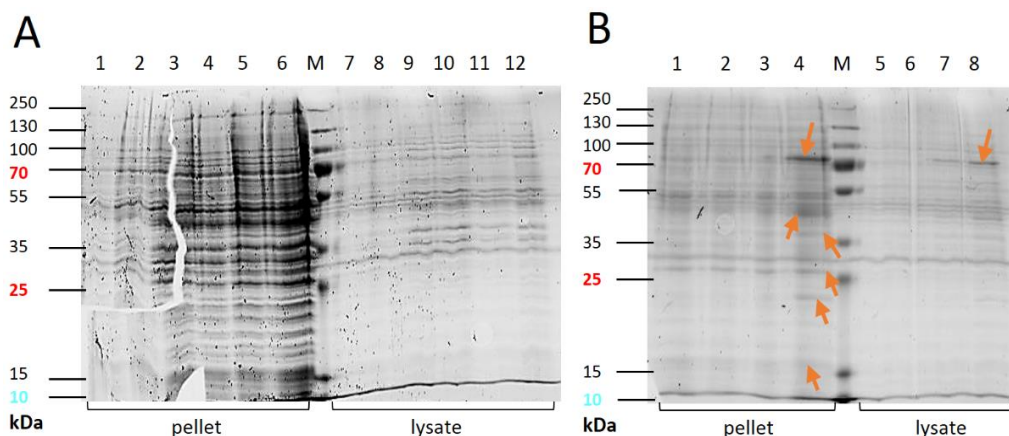


Figure 37 - 15% SDS-PAGE analyses of the test expression control experiments in *P. pastoris*:

A) Protein expression pattern of the non-induced Mrr1 FL protein in pPinkHC with pichia strain 1: The indicated time points are referring to the hours post induction of the co-cultured induced expression samples. 1-7 resuspended pellet fractions and 8-14 lysate fractions; M = marker; 1/7 = 0 h sample; 2/8 = 3h sample; 3/10 = 7 h sample; 4/11 = 20 h sample; 5/12 = 26 h sample; 6/13 = 52 h sample; 7/14 = 76 h sample **B)** Protein expression pattern of the empty, integrated pPinkHC with induction in strain 1: samples 1-4 resuspended pellet fractions and 5-8 lysate fractions; M = marker; 1/5 = 1 day before induction; 2/6 = time of induction. The orange arrows indicate protein bands, which are at a similar height as the ones marked in Figure 36C.

Control expression experiments with the non-induced Mrr1 FL or the empty but induced pPinkHC vector confirmed that the prominent bands that appear in the expression pattern originate from the organism itself after methanol induction (Figure 37). Lysate and pellet analysis of the non-induced cells shows hardly any prominent bands (Figure 37A, lanes 1-6 and 7-12 respectively) whereas the empty vector expression pattern shows weak but clear expression of the above mentioned proteins, indicated by the orange arrows in Figure 36C and Figure 37B.

None of the performed pichia expression experiments led to any Mrr1 expression in *P. pastoris*, which is also summarized in Table 28. To ensure that this was not due to a single integrant, which may have been an outlier, all of the Mrr1 FL pPinkHC/LC integrants were expressed and analyzed by SDS-PAGE, but none of them showed any differences. In addition to the initial integration control by colony color, colony PCR experiments were performed to check if the positive integrants definitively contained the gene of interest. In most of the analyzed colonies the gene of interest could be amplified and the length was confirmed in a DNA agarose gel. For some colonies, even repeated amplification trials, did not result in any DNA amplification or a signal in the DNA agarose gel. In addition, a western blot experiment (not shown) with an α His-antibody was performed with the

Results

same fractions depicted in **Figure 36**. Only for the positive control, FadA5, a band was observed in the Western blot whereas none was present for the Mrr1 samples. This again confirmed that the yeast expression experiments did not yield any Mrr1 protein.

Table 28 - Summary of the performed protein expression tests for all three expression vectors and the four strains

	strain	<i>Full length</i>		<i>Armadillo repeat long</i>		<i>'250'</i>	
		integration / verification	expression	integration / verification	expression	integration / verification	expression
pPink α HC (secreting)	1	color	no	color	no	color	no
	2	color	no	color	no	color	no
	3	color	no	color	no	color	no
	4	color	no	color	no	color	no
pPinkHC	1	color, PCR	no	--	<i>n.d.</i>	color, PCR	<i>no</i>
	2	color	no	--	<i>n.d.</i>	--	<i>n.d.</i>
	3	failed	<i>n.d.</i>	--	<i>n.d.</i>	--	<i>n.d.</i>
	4	color	no	--	<i>n.d.</i>	--	<i>n.d.</i>
pPinkLC	1	color, PCR	no	--	<i>n.d.</i>	failed	<i>n.d.</i>
	2	color	no	--	<i>n.d.</i>	--	<i>n.d.</i>
	3	failed	no	--	<i>n.d.</i>	--	<i>n.d.</i>
	4	color	no	--	<i>n.d.</i>	--	<i>n.d.</i>

n.d. = not determined

4.1.4 Biophysical characterization

4.1.4.1 Dynamic light scattering (DLS) analysis of Mrr1 Full length protein

To analyze the oligomeric state, the polydispersity and the suitability of the Mrr1 FL samples for crystallization, dynamic light scattering experiments were performed. Several experiments with protein samples from different purifications and in different buffers were performed at 4 °C and 20 °C. As they all led to very similar results only some representative measurements with size-exclusion chromatography elution samples of the pooled peak I and peak II fractions are shown. The two protein samples were analyzed at concentrations of 1.2 mg/ml (peak I) and 1.3 mg/ml (peak II). Initially DLS was performed as described in 3.2.2.5 at 4 °C, subsequently the sample was heated to 20 °C in the measuring cell to also obtain room temperature results.

DLS analysis of peak I at 4 °C (**Figure 38A**) showed three peaks. The left (49.8%) and the middle (49.0%) peak represent nearly the total of the measured protein mass. The left peak shows a hydrodynamic radius $R = 5.6$ nm, which results in a calculated molecular weight of 189 kDa. This might display the protein monomer or dimer. The middle peak ($R = 17.2$ nm) displays a higher oligomer with a molecular weight of 2,610 kDa and the right peak ($R = 1988$ nm) might represent protein aggregate or most likely dust particles, as the mass was calculated to be 175,614 megaDa. Although the protein sample was centrifuged before the measurement and the cuvette was extensively cleaned and dried in an air flow, remaining dust particles might still be present. The same

Results

protein sample measured at 20 °C (**Figure 38B**) shows one broad peak. The hydrodynamic radius, $R = 26.3$ nm, indicates a higher oligomer or protein aggregate with a calculated mass of 7,070 kDa. Overall, the protein sample of peak I led to low to medium polydispersity values, but as the majority of the protein sample is likely to be aggregated it was not suitable for crystallization experiments or protein-DNA-interaction studies.

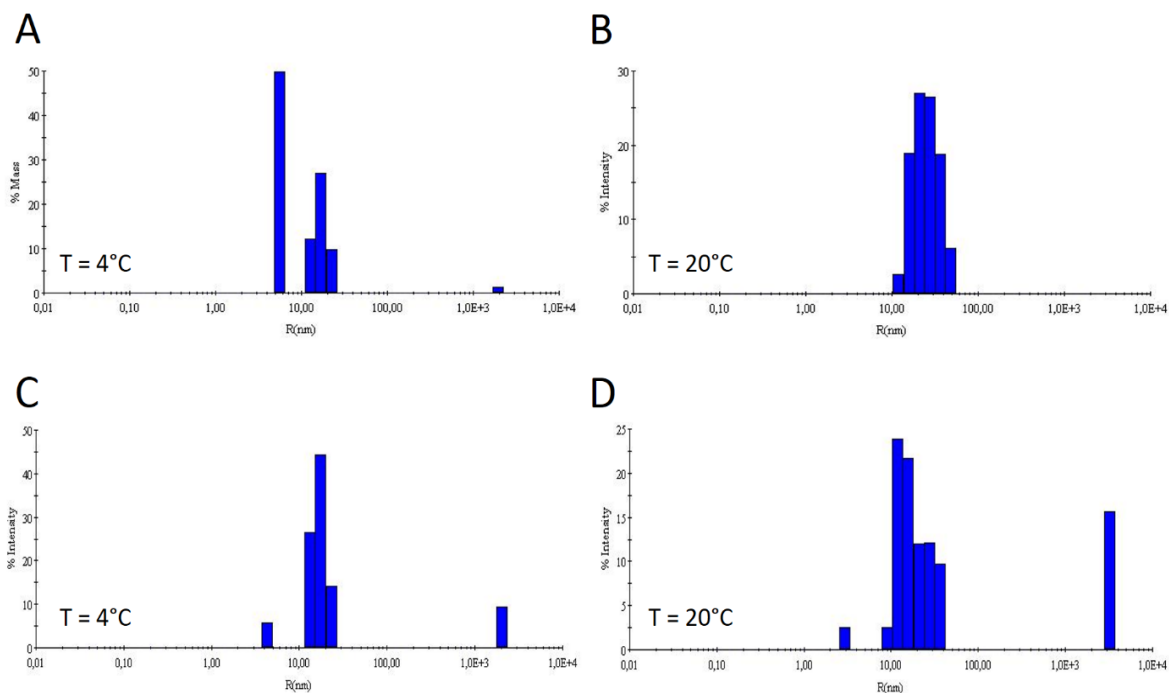


Figure 38 - Selected DLS measurements:

R = hydrodynamic radius (nm) **A/B)** Measurement of a sample from peak I (1.2 mg/ml) at 4 °C and 20 °C **C/D)** Measurement of a sample from peak II (1.3 mg/ml) at 4 °C and 20 °C

Protein samples from a purification of peak II after size exclusion chromatography measured at 4 °C again shows three peaks in the DLS experiments. 77% of the mass are accumulated in the left peak at a hydrodynamic radius of $R = 4.4$ nm. The calculated molecular weight of 106 kDa might represent a Mrr1 monomer. The middle peak ($R = 17.3$ nm) represents a higher oligomer with a calculated molecular weight of 2,642 kDa and a mass share of 22.8%. The right peak, with a calculated molecular mass of 189,391 megaDa ($R = 2,053$ nm), again might represent aggregated protein or more likely remaining dust particles. The same protein sample examined at 20 °C again led to three peaks. The left peak accumulates 79% of the mass but represents a calculated molecular weight of 43 kDa ($R = 3.0$). This low molecular weight cannot be assigned to the Mrr1 FL protein. The middle peak, with a hydrodynamic radius of $R = 15.1$ nm, is a higher oligomer with a calculated molecular weight of 1,928 kDa, representing 20% of the measured protein mass. The right peak, with $R = 31.8$ nm, depicts either protein aggregate or more likely dust particles. The protein sample of peak II led to polydispersity values, which might be sufficient for crystallization experiments. At 4 °C the result was promising for the left peak in the DLS experiment, but due to the presence of large

Results

aggregates in the same sample as observed in the other peaks the suitability for crystallization experiments or protein-DNA-interaction studies remained questionable. At higher temperature the results indicate an increase in protein aggregation, thus suggesting that crystallization is most likely, if at all, only possible at lower temperature.

4.1.4.2 Circular-dichroism experiments

Circular-dichroism measurements were performed as described in 3.2.2.4. The experiments were performed and analyzed by a bachelor student, Julius Heuchert, indicated by *, or by Christin Schäfer, indicated by **.

Julius Heuchert examined samples of Mrr1 '250' WT with and without cleaved His₆-tag and a sample of the not cleaved P683S variant. In addition samples of the cleaved variants Mrr1 '250' K335N and G878E as well as the Mrr1 '250' construct with the C-terminal His₆-tag (CT) were analyzed and an additional uncleaved Mrr1 '250' WT sample was included for comparison. K335N and the CT construct were analyzed twice with two different samples, but only the samples with the smoothest shape were used for comparison.

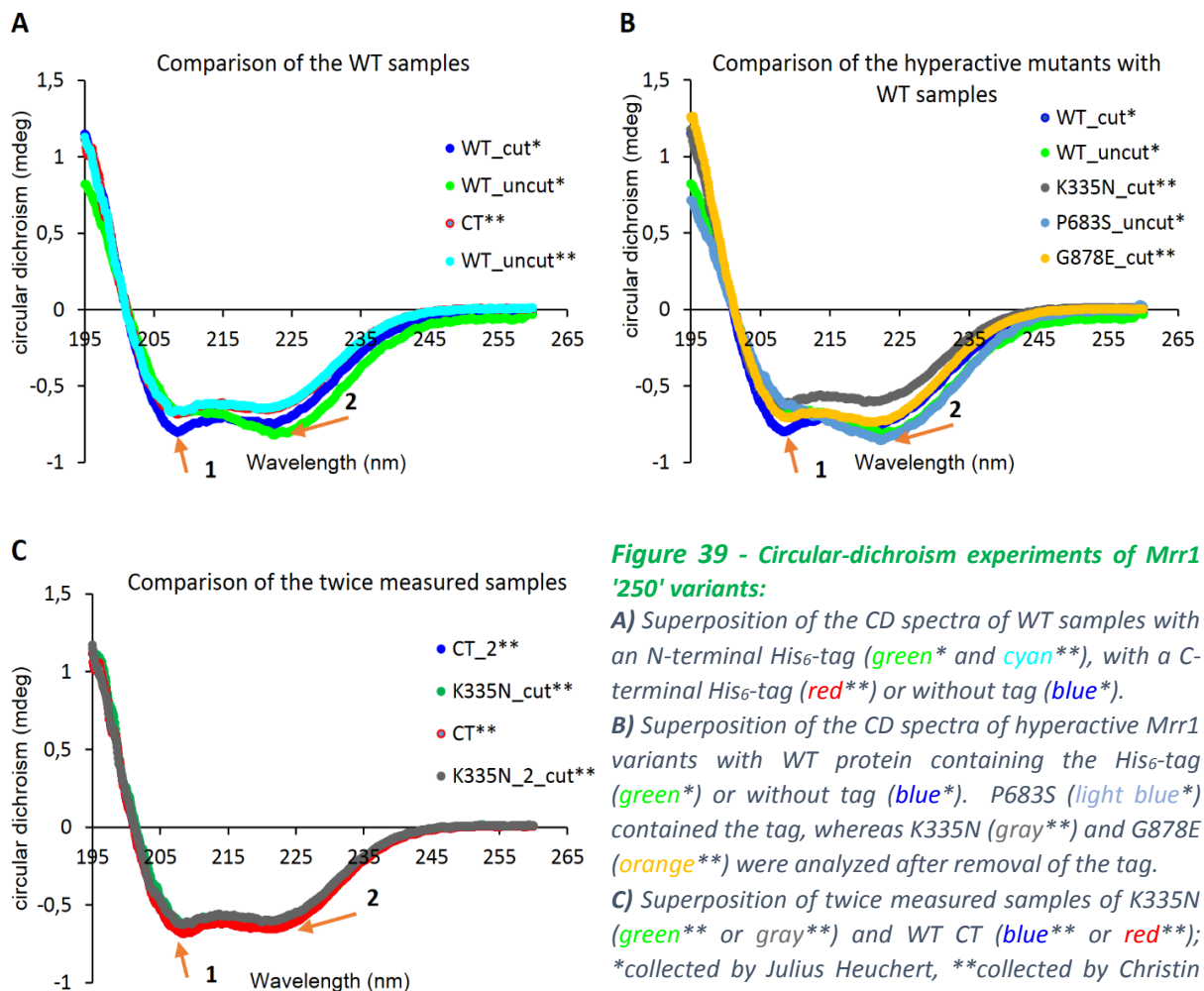


Figure 39 - Circular-dichroism experiments of Mrr1 '250' variants:

A) Superposition of the CD spectra of WT samples with an N-terminal His₆-tag (green* and cyan**), with a C-terminal His₆-tag (red**) or without tag (blue*).

B) Superposition of the CD spectra of hyperactive Mrr1 variants with WT protein containing the His₆-tag (green*) or without tag (blue*). P683S (light blue*) contained the tag, whereas K335N (gray**) and G878E (orange**) were analyzed after removal of the tag.

C) Superposition of twice measured samples of K335N (green** or gray**) and WT CT (blue** or red**); *collected by Julius Heuchert, **collected by Christin Schäfer.

Results

The results are summarized in **Figure 39** and **Table 29**. In general, three differently folded protein species have been obtained, out of which two were already described by Julius Heuchert in his bachelor thesis. There Mrr1 '250' shows a predominant α -helical folding profile for the protein that was cleaved from the tag, whereas the folding profile of the uncleaved protein somewhat looks irregular or indicates misfolding. Arrow 1 (**Figure 39**) points towards a typical minimum around 210 nm that is normally more pronounced than the second minimum at approx. 220 nm (arrow 2). These two minima define an α -helical fold, where the first minimum is normally deeper in comparison to the second minimum as can be observed for the WT_cut* sample (blue). Protein samples, analyzed by Julius Heuchert, that still contained the *N*-terminal tag (WT_uncut*, green) clearly displayed a less pronounced first minimum but a deeper second minimum (arrow 2, **Figure 39A/B**). Interestingly, this was not confirmed by a second measurement with protein still containing the *N*-terminal tag (WT_uncut**, cyan) as it shows a CD-curve with nearly equally pronounced minima. This profile represents the third obtained protein species. All measured WT or variant proteins could be categorized to belong to one of the three species (**Table 29**). This table provides an overview of the measurements and highlights similarities or differences between the measured protein samples.

Table 29 - Summary of all performed CD experiments with Mrr1 '250' variants

	<i>number of accumulations /measurement</i>	<i>observation</i>	<i>comparison of measured samples</i>	<i>respective Figure</i>
WT_uncut sample 1*	10	minimum 2 >> minimum 1	The two measurements differ from each other. Sample 1 behaves similar as the	Figure 39 A/B
WT_uncut sample 2**	10	minimum 1 = minimum 2	P683S_uncut sample but sample 2 shows two nearly equally pronounced minima.	Figure 39A
WT_cut*	10	minimum 1 >> minimum 2	This sample looks similar as the CT and K335N_cut measurements but with a stronger pronounced minimum 1	Figure 39 A/B
CT ** sample 1	10	minimum 1 > minimum 2	The measurements for the two samples are nearly identical and look similar as the two K335N_cut measurements	Figure 39 A/C
CT ** sample 2	10	minimum 1 > minimum 2		Figure 39C
K335N_cut** sample 1	10	minimum 1 > minimum 2	The measurements for the two samples are nearly identical and look similar as both CT measurements.	Figure 39C
K335N_cut** sample 2	10	minimum 1 > minimum 2		Figure 39 B/C
P683S_uncut*	10	minimum 2 >> minimum 1	This result is comparable to the measurement of WT_uncut sample 1.	Figure 39B
G878E_cut**	10	minimum 2 > minimum 1	This measurement looks similar as the result of the WT_uncut sample 1 but is less pronounced.	Figure 39B

The obtained results are difficult to evaluate, as it is not clear if the spectra of the Mrr1 samples containing the *N*-terminal tag were artifacts, but the CD-curves were at least obtained twice

Results

(WT_uncut* and P683S_uncut). Assuming that these spectra are correct, then the uncut proteins with the *N*-terminal tag assume a different fold compared to the proteins where the tag was removed, which might explain the observations obtained during crystallization as described in **Figure 39**. The CD results may also indicate that the construct with the *C*-terminal tag is more properly folded compared to the construct with the *N*-terminal tag, which might make this construct preferable for crystallization, if other issues such as degradation would not contradict this hypothesis. In the literature, however, it is stated, that for strictly α -helical proteins CD bands with negative ellipticity around 208 nm and 222 nm can be observed. Normally the excited state of the amide bond induced around 210 nm is expressed by a stronger negative minimum than the excited state represented by the negative bands around 220 nm. But predominantly α -helical proteins as myoglobin produce also CD spectra with a more prominent negative peak around 220 nm. (Dodero et al., 2011) In analogy to these results, the CD spectra of the uncut proteins likely do not represent an improperly folded protein but more likely a protein with a slightly different fold.

4.1.4.3 Size-exclusion coupled multiangle laser light scattering (SEC-MALLS)

SEC-MALLS experiments were performed with Mrr1 '250' WT and all purified variants (3.2.2.6). The wild type protein was analyzed with and without the *N*-terminal His₆-tag and with the *C*-terminal His₆-tag. The variants P683S and G997V were analyzed with the His₆-tag whereas the K335N and G878E proteins were analyzed after cleavage of the His-tag. The first experiments with the Mrr1 '250' WT cut/uncut as well as the uncut P683S and G997V variants, were performed by Julius Heuchert. The follow up experiments were performed with the knowledge that the His₆-tag seemed to be detrimental to the crystallization process and therefore it was cleaved off if possible. The WT

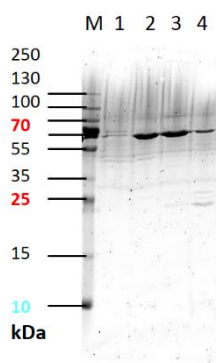


Figure 40 - 15% SDS-PAGE analysis of Mrr1 '250' G878E after size-exclusion chromatography:

The gel represents the sample used for the SEC-MALLS experiments of the G878E variant. M = marker; 1-4 = elution fractions main peak

protein with the *C*-terminal tag was of special interest, as the proteins with the *N*-terminal tag were prone to degradation and their secondary structure (CD-experiments 4.1.4.2) seemed to be influenced by the His₆-tag. It was therefore speculated that the *C*-terminal tag could change one of the above mentioned characteristics. In addition, MALLS was performed to obtain the oligomeric state and the polydispersity for all variants.

After 100 μ l of the protein sample (0.9 to 5 mg/ml) were injected onto the column, SEC-MALLS experiments were performed by monitoring light scattering and the refractive index for 30 minutes with an applied flow of 0.5 mg/ml. The results were analyzed with the ASTRA 6 software, where initially the baseline was selected and

Results

then peaks were picked. After these defining steps, the software calculated the molar mass moments for the chosen peaks and released polydispersity values.

The results for all performed experiments are listed in **Table 30**.

In general, the polydispersity of all samples was between $1.03\% \pm 1.61\%$ and $1.45\% \pm 22.7\%$, which indicates a reasonable homogeneity for crystallization experiments. The polydispersity for WT_cut seems to be an exception. All proteins samples show that at least two species are present in solution, a higher oligomer with more than $1,000 \cdot 10^3$ g/mol was recorded for most of them, except for WT CT. All tested variants also show a peak, which can be described as a protein dimer with 175 to $190 \cdot 10^3$ g/mol, and this species was predominant for all samples with the exception of the G878E variant. Some of the samples seem to contain a trimeric Mrr1 (around $280 \cdot 10^3$ g/mol) or tetrameric (around $400 \cdot 10^3$ g/mol) state or a more transient state in between. The G878E variant seems to be the outlier among the different samples as a predominant protein peak with only $1221 \cdot 10^3$ g/mol was detected, confirming the presence of a monomer. The SDS-PAGE of the applied G878E protein sample (**Figure 40**) showed that the protein was already heavily degraded, as the prominent bands are located at heights that correspond to approx. 65 kDa. The degraded G878E variant is therefore also most likely present in a predominant dimeric state, together with smaller fractions of the protein in a trimeric or tetrameric state.

Table 30 - Summary of all performed SEC-MALLS experiments with Mrr1 '250' variants

		Calculated values monomer	Peak I	Peak II	Peak III	Peak IV
Mrr1 250 WT*uncut	Molar mass ($\cdot 10^3$ g/mol)	99.2 without His ₆ -tag 102.2 with His ₆ -tag	7,317	340	190	--
	Oligomeric state		oligomer	trimer/tetramer	dimer	--
	Polydispersity (%)		1.03% \pm 1.61%			
Mrr1 250 WT cut*	Molar mass ($\cdot 10^3$ g/mol)	99.2 without His ₆ -tag 102.2 with His ₆ -tag	2,999	188	--	--
	Oligomeric state		oligomer	dimer	--	--
	Polydispersity (%)		10.4% \pm 73.2%			
Mrr1 250 CT**	Molar mass ($\cdot 10^3$ g/mol)	100.7 with His ₆ -tag	571	272	182	--
	Oligomeric state		hexamer	trimer	dimer	--
	Polydispersity (%)		1.45% \pm 22.7%			
Mrr1 250 K335N cut**	Molar mass ($\cdot 10^3$ g/mol)	99.1 without His ₆ -tag 102.1 with His ₆ -tag	3,061	333	175	--
	Oligomeric state		oligomer	trimer/tetramer	dimer	--
	Polydispersity (%)		1.05% \pm 2.06%			
Mrr1 250 P683S*	Molar mass ($\cdot 10^3$ g/mol)	99.0 without His ₆ -tag 102.0 with His ₆ -tag	7,055	405	188	--
	Oligomeric state		oligomer	tetramer	dimer	--
	Polydispersity (%)		1.15% \pm 3.03%			
Mrr1 250 G878E cut**	Molar mass ($\cdot 10^3$ g/mol)	99.1 without His ₆ -tag 102.1 with His ₆ -tag	4,208	252	181	122
	Oligomeric state		oligomer	tetramer	trimer	dimer

Results

		Polydispersity (%)	1.28% ± 4.26%			
Mrr1 250 G997V*	Molar mass (· 10 ³ g/mol)	99.2 without His ₆ -tag 102.2 with His ₆ -tag	7,535	368	177	--
	Oligomeric state		oligomer	tetramer	dimer	--
	Polydispersity (%)		1.29% ± 15.9%			

* collected by Julius Heuchert; ** collected by Christin Schäfer

The G878E variant points to a problem that was already alluded to in the purification chapter, which is degradation. This result underlines the fact that dependent on the expression or purification, the stability of Mrr1 '250' variants varies significantly although they were expressed and purified utilizing the same protocol. A comparison of **Figure 33B** and **Figure 40** shows that the same variant was purified nearly without degradation (**Figure 33B**) or being almost completely degraded to a significantly shorter polypeptide (**Figure 40**). Independent from this extreme example nearly all variants of Mrr1 '250' analyzed by SEC-MALLS indicate degradation, as the expected molar mass for the dimeric protein with the His₆-tag is approx. 204 · 10³ g/mol and without the His₆-tag approx. 198 · 10³ g/mol. The expected accuracy of the determined molar masses is ≤ ±5% by MALLS, and thus only the cut or uncut WT proteins with the *N*-terminal tag indicate no or minor degradation. The *N*-terminal tag containing variants, show minor or strong degradation, which is especially pronounced for the G997V, K335N and G878E variants. The *C*-terminal tagged Mrr1 construct also shows comparably strong degradation tendencies and therefore the approach of using a *C*-terminal instead of an *N*-terminal tag does not seem to be favorable.

4.1.5 Crystallization

Due to the aggregation problem of the Mrr1 Full length protein and the expression problems with the Armadillo repeat truncation variant crystallization experiments were only performed with the Mrr1 '250' protein and its variants.

4.1.5.1 Mrr1 '250' wild type protein

Crystallization experiments for the wild type protein were performed with the protein in different size-exclusion buffers. These buffers differed in buffer type, Tris-HCl, Ada-NaOH or Bicine-NaOH, buffer concentration (20, 30 or 50 mM) and in the pH (7.0 or 8.0). In addition, the buffers varied in their salt content, 100, 200, 300 or 500 mM, and the nature of the salt, NaCl or KCl. The experiments were pursued at a protein concentration of approx. 11 mg/ml at 20 °C and 4 °C. The described buffer variations were applied for crystallization trials with the uncleaved Mrr1 WT protein and all available

Results

commercial screens with the exception of the MIDAS screen (Table 15 3, 5 to 7, 9 to 14) in the 96-well format.

None of these crystallization trials led to the formation of crystals; many conditions led to precipitation or phase separation processes. In a few cases small micro nuclei or crystalline precipitate was observed. Extensive fine screening trials of selected conditions that showed micro crystals or crystalline precipitate in former screens were performed in 96-well and 24-well format. These secondary screens did also not lead to protein crystallization, only some salt crystals were obtained.

As the above mentioned experiments did not yield any Mrr1 '250' WT crystals, crystallization trials were pursued with the cleaved protein at concentrations of 10 to 11 mg/ml and in size-exclusion buffer 37 (20 mM Bicine-NaOH pH 8.0, 300 mM KCl). Initial crystallization trials were performed with the twelve available commercial screens in 96-well format.

These crystallization experiments led to small needle like or small cubic protein crystals in four different conditions (Figure 41). Unfortunately, subsequent attempts to reproduce the crystals in fine screen experiments in 96-well and 24-well format with increased drop size, were unsuccessful. Reproduction of the crystals was only successful when the initial screens with the Honeybee in the 96-well format and with drops of 0.3 + 0.3 μ l were repeated.

As crystal improvement failed, some of the small needles were harvested, transferred into cryoprotectant solution and flash frozen in liquid nitrogen. These crystals were tested at the microfocus beamline at the ESRF, Grenoble, and showed some diffraction spots around the beam center, indicating that they are protein crystals. However, analysis of dissolved crystals by SDS PAGE and silver staining failed as probably too few crystals were fished. If sufficient amounts of protein would have been present for silver staining it would have been possible to evaluate if the crystallized protein was further degraded in the drop prior to crystallization. This information could have been used to tailor constructs of a distinct length that might crystallize more readily.

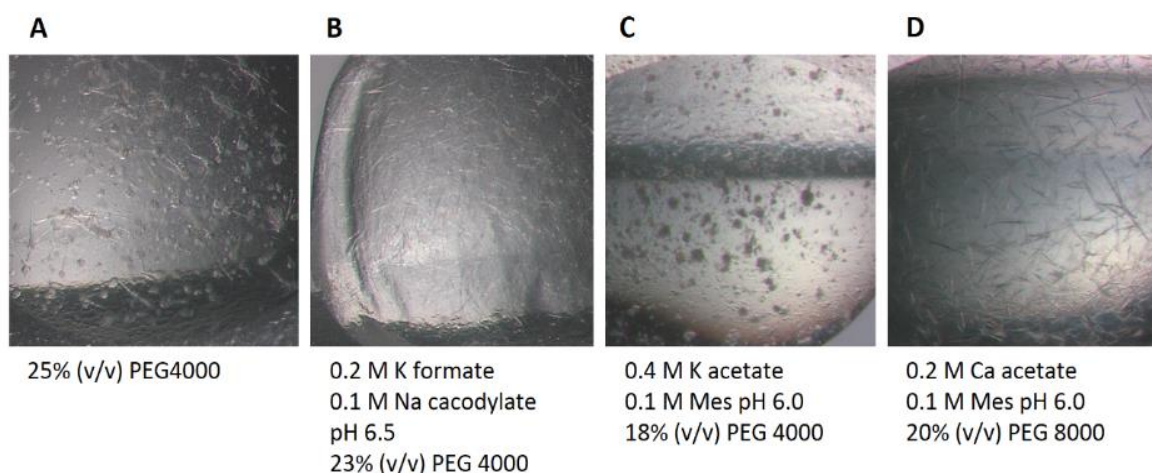


Figure 41 - Crystallization conditions of Mrr1 '250' WT with cleaved His₆-tag

Results

4.1.5.2 Mrr1 '250' mutant proteins

In addition, to the crystallization experiments with the WT protein some of the expressed gain of function variants, which were available in reasonable amounts, were applied to crystallization.

The crystallization set ups of the P638S and G997V variants were initially performed by Julius Heuchert. The G997V variant was only available in small amounts, and therefore not all crystallization screens could be used. Both protein variants were crystallized in the 96-well format, containing the His₆-tag, at concentrations of 10 to 11 mg/ml in size-exclusion buffer **37** (20 mM Bicine-NaOH pH 8.0, 300 mM KCl). These crystallization trials did not yield crystals, again only very small nuclei or crystalline precipitate were observed.

Subsequently, P683S_cut was used for crystallization set ups at similar concentrations and in the common size-exclusion buffer. These set ups yielded thin plate like crystals (**Figure 42A**), oval short sticks (**Figure 42B**) or small globular (pseudo)crystals (**Figure 42C**). The obtained crystallization conditions were used for fine screens in the 24- and 96-well format but only yielded crystals for condition A. A subsequent analysis of these crystals revealed that they were salt crystals.

The G997V and the G878E variants were not used for further crystallization experiments, because both proteins were not available in reasonable amounts or were heavily degrading during the purification process.

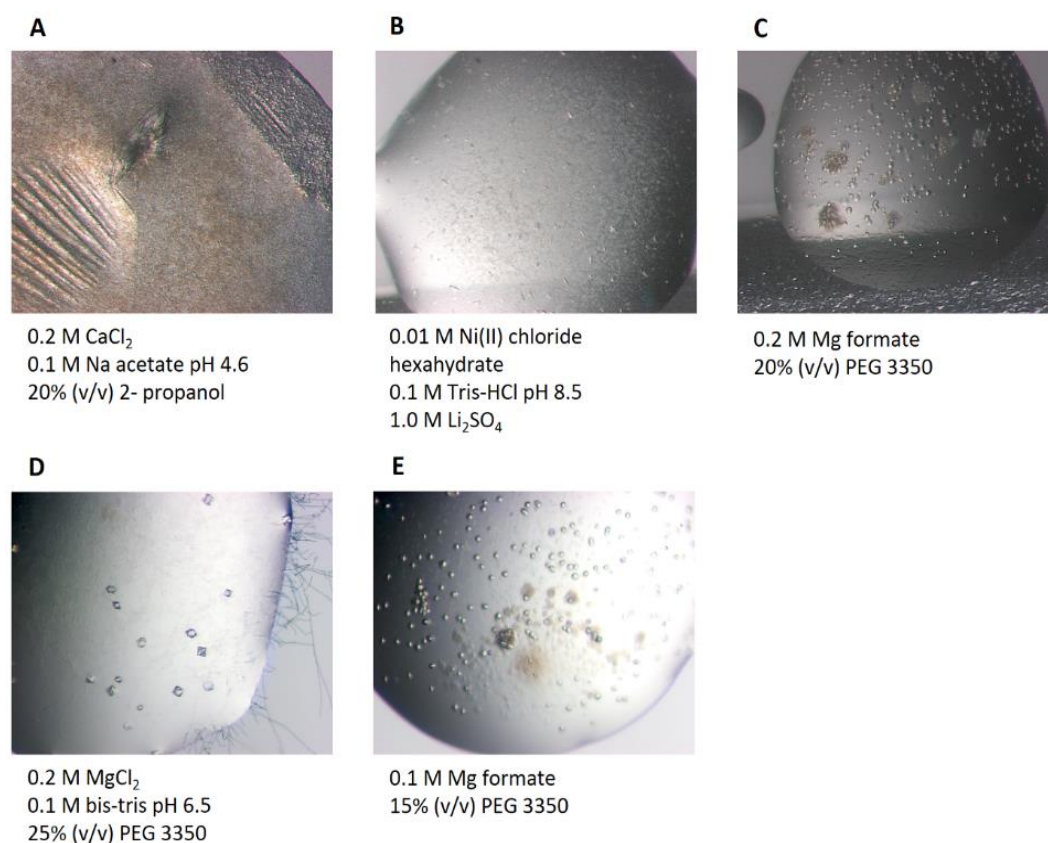


Figure 42 - Crystallization conditions of the Mrr1 '250' variants P683S and K335N with cleaved His₆-tag:
A-C) Crystals of P683S cut; **D/E)** Crystals of K335N cut

Results

The TEV-cleaved K335N Mrr1 '250' variant was applied to crystallization set ups at concentrations of approx. 11 mg/ml with all available commercial screens in the 96-well format. These initial screens led to small globular (pseudo)crystals (**Figure 42E**) and crooked octahedral but clear shaped small crystals (**Figure 42D**), the latter type was unfortunately grown in a condition that also led to fungal growth.

The crystals from condition D were harvested, transferred into cryoprotectant solution and flash frozen in liquid nitrogen. These crystals were tested at the microfocus beamline at the ESRF, Grenoble, and were identified to be salt crystals.

So far neither the WT nor the gain of function mutant constructs yielded crystals that were sufficient in size and quality to obtain a dataset. Without protein crystallization, structure determination or comparison is not possible, and therefore only modeling studies with the I-TASSER server were performed to gain some structural insights into Mrr1 (Modeling).

4.2 FadA5 – a thiolase from *M. tuberculosis*

The FadA5 wild type (WT) and variants were expressed, purified and crystallized. As no structural information was published for this enzyme the structure was initially solved by molecular replacement (MR) with a thiolase structure from *T. thermophilus* as search model. Of all thiolase structures deposited in the protein data bank the thiolase from *T. thermophilus* possessed the highest sequence identity to FadA5 (43%) and was therefore deemed to be the most suitable search model. All subsequent structures were solved with the initial FadA5 structure as search model, and again by MR. One aim of this work was the characterization of FadA5 as potential drug target, but as no inhibitors are known, the characterization of the active site and potential binding pockets was pursued by co-crystallization experiments with the enzyme products acetyl-CoA and 3-oxo-pregn-4-ene-20-carboxyl-CoA (3-OPC-CoA) as well as the substrate analog 3-oxo-chol-4,22-diene-24-oyl-CoA (3-OCD-CoA). After these structures were solved, a thorough comparison with the known human thiolases was performed to obtain an initial glimpse into the possibility to specifically inhibit FadA5 without affecting similar human pathways.

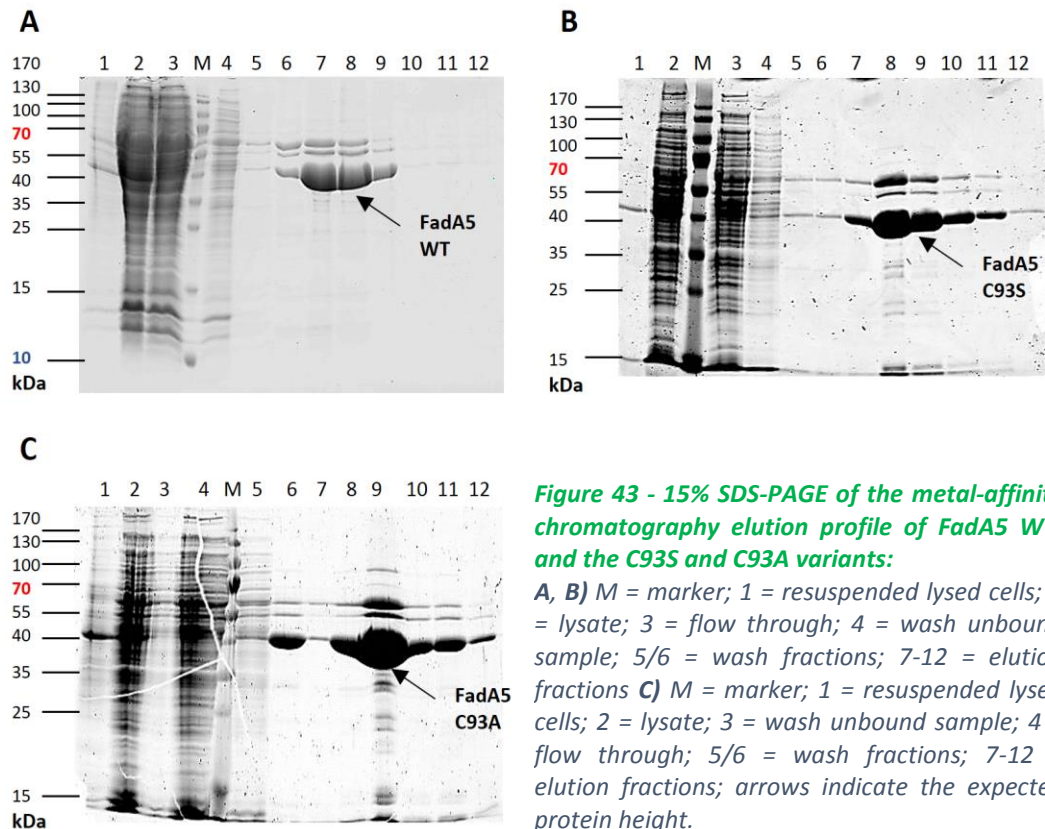
4.2.1 Expression and purification of the FadA5 variants

Expression and purification of FadA5 has been described (Nesbitt et al., 2010) and was established in a precedent work. (Schaefer, 2010). For the expression of wild type FadA5 or the, C93S and C93A variants, the FadA5 gene containing vectors were transformed in electro-competent *M. smegmatis* cells (mc²155) and a single colony was used to inoculate a preculture. The preculture grew for 3 to 4 days and was then used to inoculate the main culture. Protein expression was induced after 30 h culture growth by the addition of 0.2% (w/v) acetamide and the bacteria were harvested 40 h post induction.

FadA5 was purified in a three-step purification process, the first was an affinity-chromatography step, followed by further cleaning via anion-exchange chromatography and a final polishing step by size-exclusion chromatography. The metal-affinity chromatography was performed as described in 3.2.1.2 with the standard buffer combination for FadA5 (Table 12). The Co-Talon resin was loaded with the cleared lysate by gravity flow and unbound proteins were subsequently washed out with binding buffer. After two additional wash-steps, the protein was eluted with elution buffer yielding six fractions of eluted protein. The success of the purification step was analyzed by SDS-PAGE, which showed for WT and variant proteins a similar elution pattern (Figure 43). The band slightly above the 40 kDa marker band is in good agreement with the expected band at 43 kDa. The elution fractions

Results

were pooled, concentrated and applied to anion exchange chromatography (3.2.1.3) with the standard buffers, **3** and **4** (Table 12), for FadA5. The elution process was monitored by protein absorption at 280 nm and 260 nm (Figure 44A). The elution profile and the following SDS-PAGE analysis (Figure 44B) clearly indicates two peaks A and B, with peak B containing nearly pure FadA5.



The bands were also analyzed by mass spectrometry which identified the band at approx. 60 kDa (peak A) as the GroEl chaperone (*M. smegmatis*), the band at approx. 50 kDa as methyltransferase (*M. smegmatis*) and the band at approx. 40 kDa as FadA5. The elution profiles and the SDS-PAGE analysis of the same steps during the purification of the FadA5 C93S and C93A variants led to similar peaks and bands (Figure S 1 / Figure S 2) but mass spectrometry was only performed with the bands from the WT SDS-PAGE. The pooled and concentrated fractions from the anion-exchange chromatography, were applied to size exclusion chromatography (3.2.1.5), with buffer **5** (Table 12). The elution was monitored at 280 nm and 260 nm and was analyzed by SDS-PAGE. The elution profile reveals one slightly asymmetric peak with a smaller pre-peak A and a large main peak B (Figure 44C). The SDS-PAGE analysis (Figure 44D) identified FadA5 at a height of 43 kDa in both peaks and a slight impurity at approx. 50 kDa. All FadA5 containing fractions were pooled; the impurity at 50 kDa was thought to be minor relative to the amounts of FadA5. The size-exclusion chromatography profiles of the variant proteins show a similar picture for the C93S variant as observed for the WT protein (Figure 44C/D) and a nearly symmetrical peak A without a pre-peak for the C93A variant (Figure 44C/D). The pure protein was concentrated to approx. 12 mg/ml, aliquoted

Results

in 35 μ l, flash frozen with liquid nitrogen and stored at -80 $^{\circ}$ C for later use.

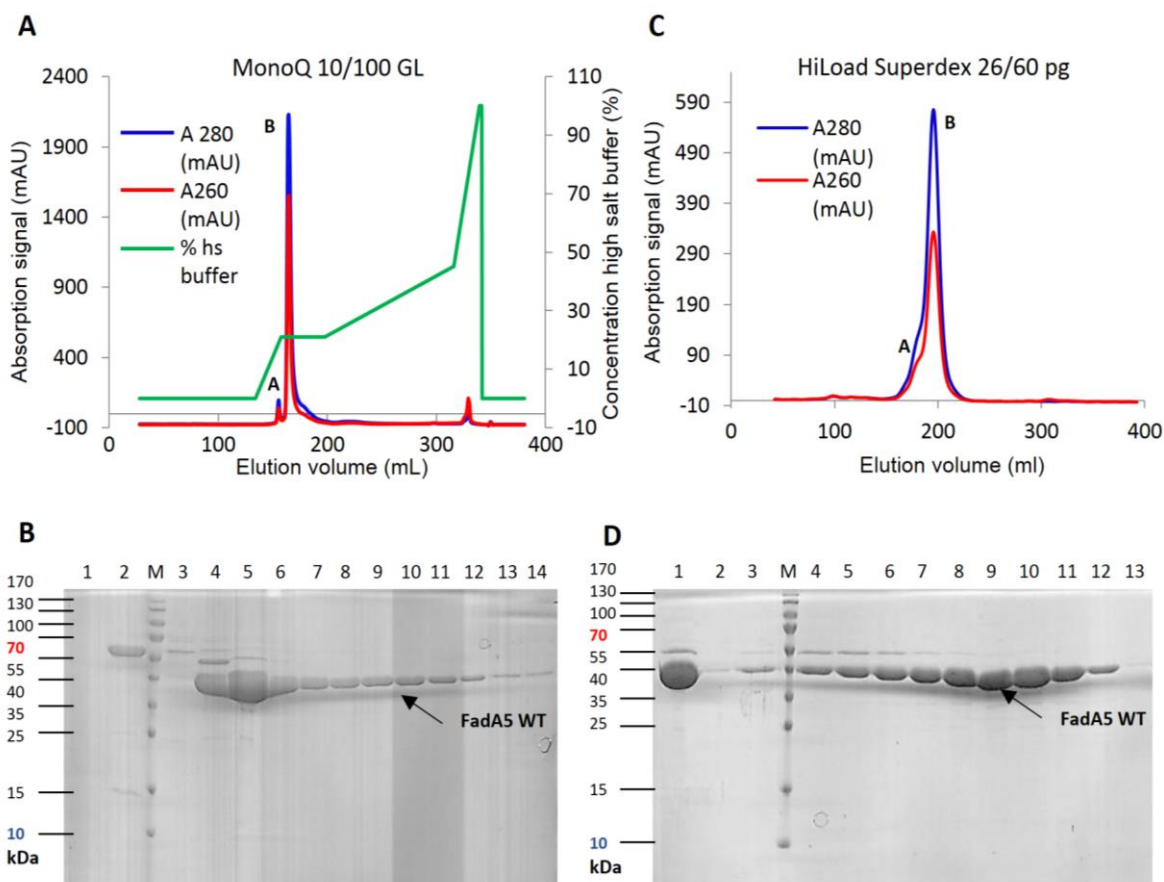


Figure 44 - Purification of FadA5 WT:

A) Chromatogram after anion-exchange chromatography: absorption at 280 nm (blue) and 260 nm (red); % high salt buffer (green) **B)** 15% SDS-PAGE after anion-exchange chromatography: M = marker, 1 = flow through, 2/3 = elution fractions peak A, 4-14 = elution fractions peak B **C)** Chromatogram after size-exclusion chromatography: absorption at 280 nm (blue) and 260 nm (red) **D)** 15% SDS-PAGE after size-exclusion chromatography: M = marker, 1 = load, 2-4 = elution fractions peak A, 5-13 = elution fractions peak B

4.2.2 Biochemical and biophysical characterization

4.2.2.1 Size-exclusion coupled multiangle laser light scattering (SEC-MALLS)

The pure FadA5 variants were analyzed by SEC-MALLS (3.2.2.6), to determine the homogeneity and the oligomeric state of the proteins.

The MALLS measurements show a very similar peak pattern for the three samples. The predominant state of FadA5 is a dimer, peak III **Figure 45A**, peak II **Figure 45B** and peak II **Figure 45C**. In all chromatograms minor fractions of the protein either formed higher oligomers or monomers but the homogeneity or polydispersity was very low (**Table 31**). These experiments confirmed an applicability of the FadA5 samples for crystallization and biochemical characterization. Peak III in C is likely to be

Results

an artifact, as it is at the height of the salt peaks and the calculated mass of 8 g/mol cannot be assigned.

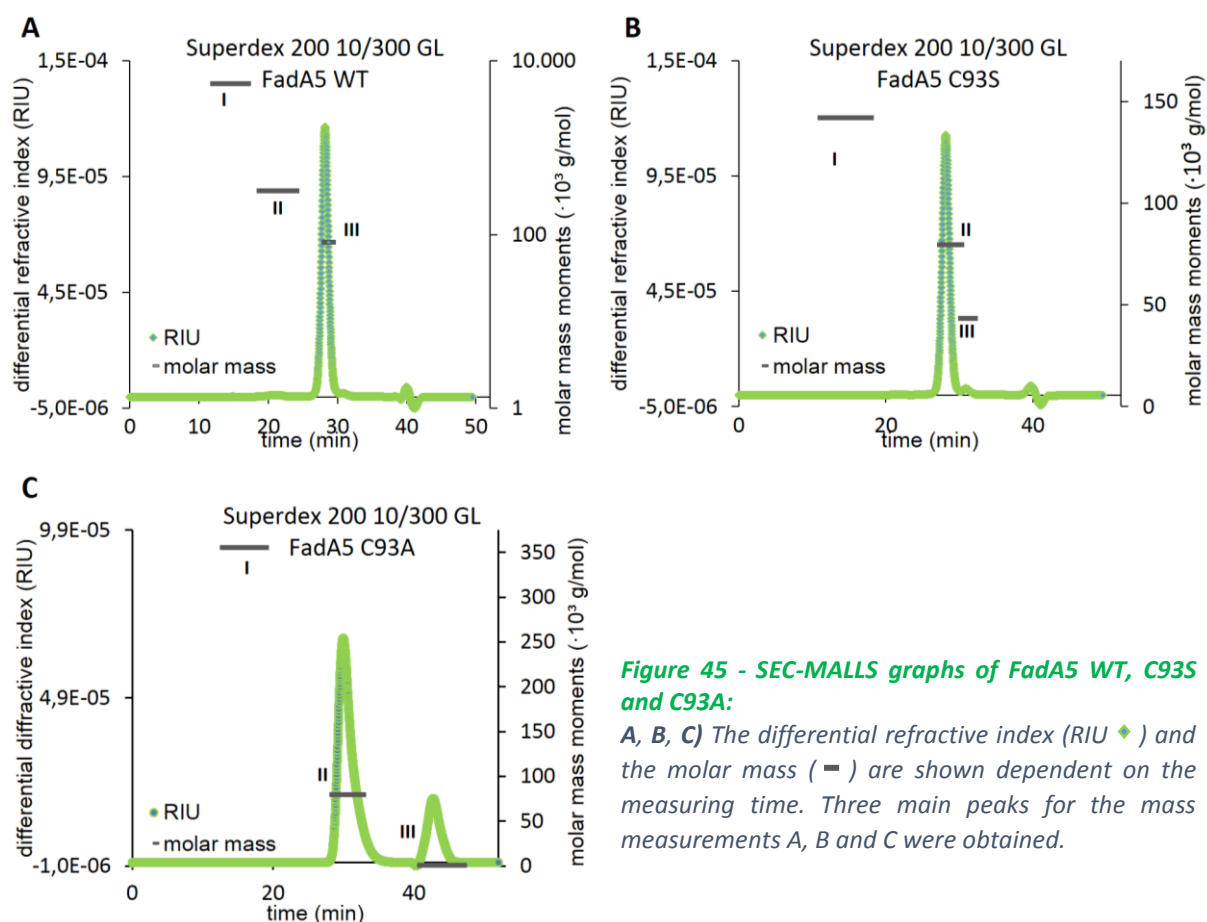


Figure 45 - SEC-MALLS graphs of FadA5 WT, C93S and C93A:

A, B, C) The differential refractive index (RIU \blacklozenge) and the molar mass (—) are shown dependent on the measuring time. Three main peaks for the mass measurements A, B and C were obtained.

Table 31 - Summary of the SEC-MALLS experiments: The predominant oligomeric state is highlighted in bold letters

		Calculated values monomer	Peak I	Peak II	Peak III
FadA5 WT	Molar mass (g/mol)	41.3 without His ₆ -Tag 43.3 with His ₆ -Tag	$6,991 \cdot 10^3$	$346 \cdot 10^3$	$81.4 \cdot 10^3$
	Oligomeric state		oligomer	octamer	dimer
	Polydispersity (%)			$1.06\% \pm 2.80\%$	
FadA5 C93S	Molar mass (g/mol)	41.3 without His ₆ -Tag 43.3 with His ₆ -Tag	$638 \cdot 10^3$	$79.7 \cdot 10^3$	$45.2 \cdot 10^3$
	Oligomeric state		oligomer	dimer	monomer
	Polydispersity (%)			$1.04\% \pm 2.67\%$	
FadA5 C93A	Molar mass (g/mol)	41.3 without His ₆ -Tag 43.3 with His ₆ -Tag	$355 \cdot 10^3$	$79.4 \cdot 10^3$	(320)
	Oligomeric state		octamer	dimer	--
	Polydispersity (%)			$1.00\% \pm 0.821\%$	

Results

4.2.2.2 Isothermal titration calorimetry

The ITC experiments were performed to obtain information about the binding affinities between the FadA5 proteins (WT and C93A) and an enzyme product, 3-oxo-pregn-4-ene-20-carboxyl-CoA (3-OPC-CoA).

During the experiments 3-OPC-CoA, in buffer **41**, at a 20-fold molar excess was titrated to the protein-buffer (**41**) solution into the sample cell under constant stirring. Dependent on the respective experiment, measurements were performed under presence or absence of a 20-fold molar excess of CoA in the ligand and the protein-buffer solutions. 21 injections were performed and the binding curve was obtained by monitoring the heat release ($\mu\text{cal/s}$) at every point of injection in comparison to a reference cell. In addition to the protein experiment, a reference measurement injecting the ligand into buffer was performed and subtracted from the protein measurement prior to its evaluation. For the analysis with the ORIGIN 7 software the data from the initial injection were deleted and the fitting model 'One Site of binding' was applied.

The results of the ITC measurements are listed in [Table 32](#).

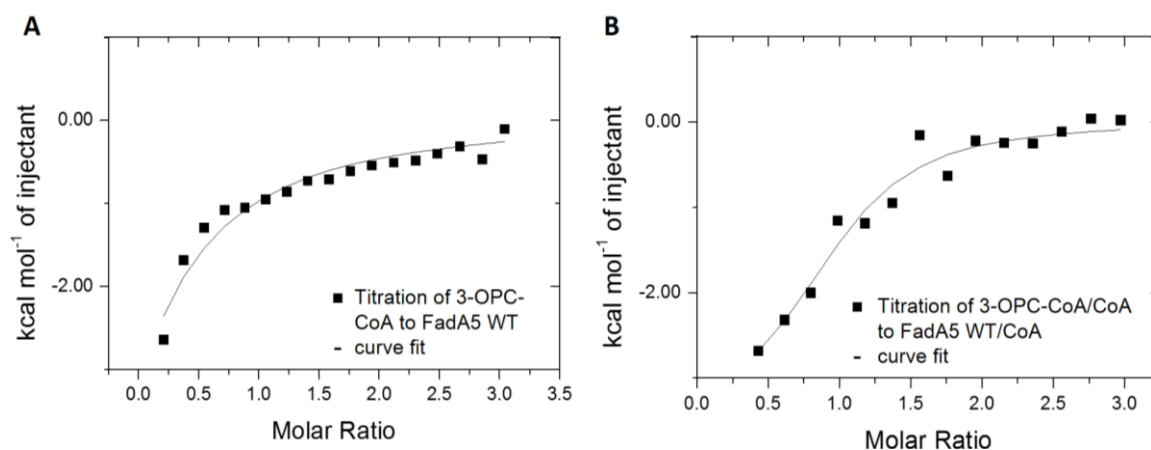


Figure 46 - Isothermal plots of the ITC experiments with FadA5 WT in the presence of different ligands:

A) The heat release by addition of 3-OPC-CoA to FadA5 WT is low, which results in low binding energies and a very low number of binding sites. **B)** The heat release by addition of 3-OPC-CoA and equimolar amounts of CoA to FadA5 WT and CoA shows a binding in the low μM range and a reasonable number of binding sites could be calculated ([Table 32](#)).

The measurement with FadA5 WT and 3-OPC-CoA ([Figure 46A](#)) indicated a K_d^{app} of 34 μM with a large error of more than 60 μM but the stoichiometry could not be determined properly as the error is 600-fold higher than the actual value for N, which is 0.001. The absolute obtained values for the enthalpy and entropy were in the range of 3,000 kcal/mol with very high error ranges, which renders the measured values doubtful.

One ITC experiment with FadA5 WT, 3-OPC-CoA and CoA was performed in triplicate ([Figure 46B](#)). These experiments indicate a K_d^{app} in the low μM range with a 1:1 ratio between enzyme and ligand

Results

although the error range is very high. For these measurements, a competitive binding of CoA and the product has to be assumed, as the protein was initially saturated with a 20-fold molar excess of CoA prior to the titration with 3-OPC-CoA. The obtained values are thus likely to show an average heat release due to CoA dissociation and 3-OPC-CoA binding to the enzyme.

In addition, the FadA5 C93A variant, a catalytically inactive variant, was titrated against 3-OPC-CoA / CoA and 3-OPC-CoA alone. Similar to the WT experiment with CoA only, these titrations did not yield data, which could be fitted in a reasonable manner and hence neither stoichiometry values nor binding affinities could be calculated.

Table 32 - Results of the ITC measurements

	addition of CoA	addition of 3-OPC-CoA	number of measurements	K_d^{app} (μ M)	stoichiometry N	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)
FadA5 WT	no	yes	1	34 ± 62	0.001 ± 0.653	$-2.4 \cdot 10^3 \pm 1,5 \cdot 10^6$	$2.88 \cdot 10^3$
	yes	yes	3	8.1 ± 22	1.32 ± 1.12	-4.16 ± 1.45	-2.97

These ITC experiments can only be regarded as very preliminary results and should be interpreted cautiously, if at all. The measurements were performed under wrong assumptions, as 3-OPC-CoA was thought to be the substrate of the thiolytic cleavage reaction, whereas it is the product. Due to this assumption CoA was added, to enable substrate turnover, because the initial WT 3-OPC-CoA experiment had showed very low stoichiometry. It was suggested, that the low stoichiometry values are caused by the lack of the second reaction partner and that affinity and thereby binding would be triggered, if this second reaction partner of the ping-pong mechanism (Figure 11, state III), CoA, is present. Since 3-OPC-CoA is the product, eventually a competitive reaction was taking place in the presence of both ligands. Another problem that might have occurred was enzyme inhibition by CoA. It is known from the literature, that CoA inhibits thiolase reactions if it is present in > 1000-fold molar excess to the enzyme. (Middleton, 1974; Schaefer et al.) This excess however, was not reached at all, as CoA was applied at a maximum of 40-fold molar excess.

The ITC experiments require intensive optimization to evaluate the optimum injection numbers and amounts, to enable saturation after each injection, which was not always the case. In addition, optimization of the protein to ligand ratio could be performed and measurements, with FadA5 and 3-OPC-CoA only, could be optimized to monitor dissociation constants between the two compounds and to compare binding behavior of the WT and the C93A protein. A limiting factor for all those measurements, however, is and was the high amount of ligands required during the experiments.

4.2.3 Crystals structures of FadA5 apo

To obtain structural information at the atomic level crystallization and a valid structure solution are inevitable processes. The protein was initially tried to be crystallized in its apo form. This, however, proved to be difficult and the first structure solved was therefore FadA5 in complex with (acetyl-)CoA. Nevertheless, the apo structures will be described first in the following chapters.

4.2.3.1 Crystallization

The FadA5 WT protein was used for crystallization in concentrations of approx. 12 mg/ml. For the majority of crystallization experiments protein in SE buffer **5** (20 mM Bicine-NaOH pH 8.5, 250 mM NaCl) was used. However, one crystallization condition (0.1 M citrate pH 5.5, 1.0 M $(\text{NH}_4)_2\text{HPO}_4$, 0.2 M NaCl; final pH 7.6) was reproduced with protein in SE buffer **5o** (60 mM Bicine-NaOH pH 8.5, 250 mM NaCl). Crystal growth in the latter condition was very slowly and took four to six months.

In general crystallization of FadA5 was achieved readily but as described in Schaefer 2010 never led to a dataset which permitted refinement. (Schaefer, 2010) Data sets of reasonable quality at a maximum resolution of 2.3 Å, most likely in space group $I4_122$, were collected and could be solved by molecular replacement (MR) with a thiolase structure from *T. thermophilus* (PDB entry 1ulq) as search model. However, none of the data sets could be refined, as some symmetry problems, possibly a difficult case of twinning, prevented successful data refinement and model building.

Due to the structure refinement and model building problems other crystallization conditions were pursued. Initial crystallization experiments were performed with all available commercial screens, i.e. Crystal Screen I/II, Index Screen, Topaz OptiMix I-V, Topaz OptiMix-PEG, Nextal PEG and Wizard Screen I/II in a 96-well format. In later stages of this work screening was also performed with the self-designed FadA5 Screen.

In addition to the already known condition (0.1 M MES pH 6.5, 1.6 M MgSO_4), FadA5 crystallized in various other conditions (**Figure 47**). The predominating components were MES buffer in a pH range between 6.0 and 6.7 in combination with high salt (MgSO_4 or $(\text{NH}_4)_2\text{SO}_4$), which mainly resulted in octahedral or lens like crystals (**Figure 47B/D/E/F**). Some polyethylene glycol (PEG) conditions could also be identified, but crystals (**Figure 47A/C**) from the PEG conditions did not diffract very well and so these conditions were not pursued. The majority of the well diffracting crystals, especially of octahedral shape, showed the already previously identified space group, $I4_122$, and the data sets again did not permit refinement. Finally crystallization conditions were identified in a modified $(\text{NH}_4)_2\text{SO}_4$ -MES condition, containing 5% dimethyl sulfoxide (DMSO), in space group $P4_12_12$ and a

Results

data set to 2.7 Å resolution was obtained. This optimized condition, however, did not consistently produce crystals in space group $P4_12_12$ but $I4_122$ was predominant again. This condition was therefore not appropriate for reliable reproduction experiments or ligand soaking approaches.

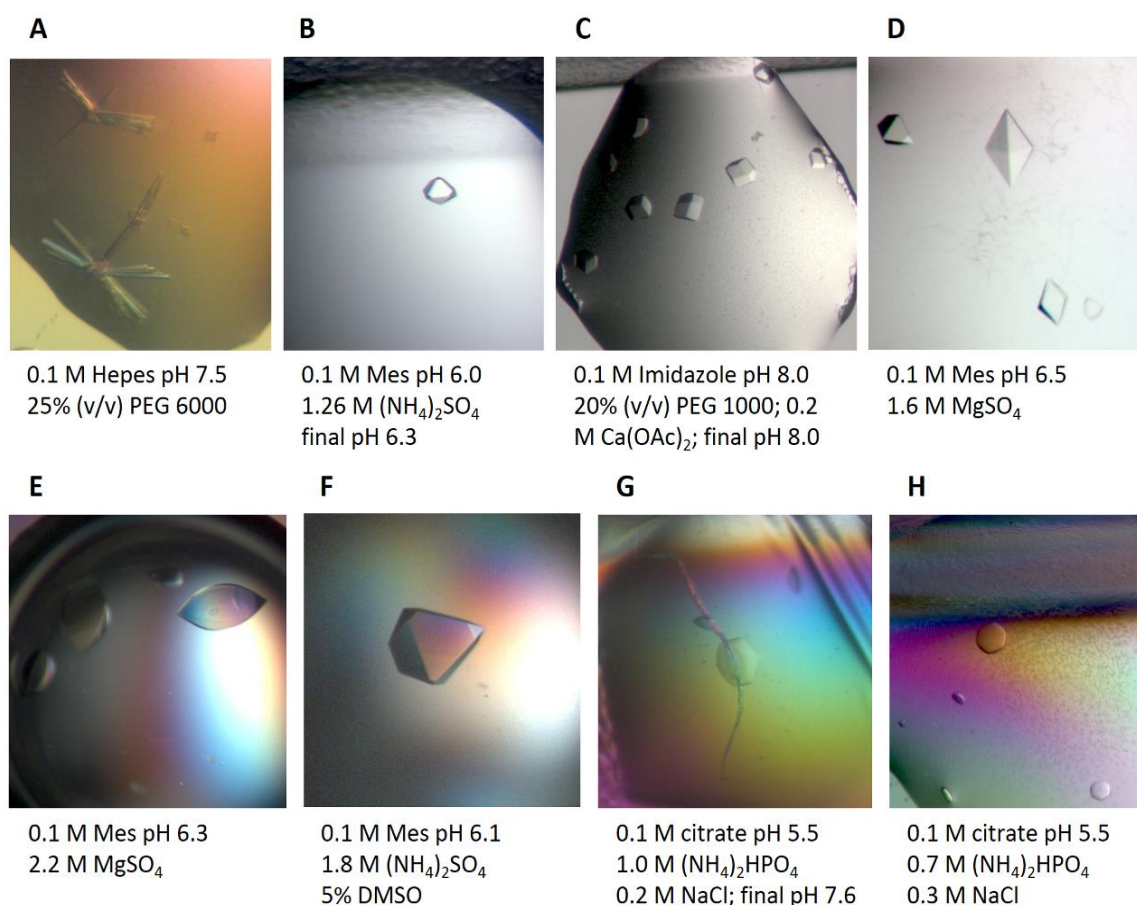


Figure 47 - Crystallization conditions of FadA5 WT apo

As mentioned in the beginning of this chapter, another crystallization condition (0.1 M citrate pH 5.5, 1.0 M $(\text{NH}_4)_2\text{HPO}_4$, 0.2 M NaCl; final pH 7.6), where crystal growth took four to six months, was attempted to be reproduced in the 96-well format. The initial crystal was of hexagonal shape (**Figure 47G**), crystallized in space group $P6_322$ and diffracted to 3.5 Å resolution. Fine screening of this crystallization condition, with protein in the old SE buffer, yielded crystals of the same space group after six months that diffracted to 2.2 Å (**Figure 47H**).

4.2.3.2 Data collection, structure solution and validation

For data collection, the apo crystals (I) were harvested, transferred with a nylon loop into the according cryoprotectant solution (0.1 M MES pH 6.1, 2.0 M $(\text{NH}_4)_2\text{SO}_4$, 5% (v/v) DMSO, 27% (v/v) glycerol) and flash-frozen in liquid nitrogen. The data set with space group $P4_12_12$ was collected at the BESSY II beamline MX 14.1 at a wavelength of 0.9184 Å. The obtained data diffracted to 2.7 Å and the unit cell dimensions were 128.31 Å x 128.31 Å x 114.07 Å. The data were processed with iMosflm,

Results

scaled with Scala and molecular replacement (MR) was performed with phaser. Since the acetyl-CoA bound structure was solved prior to the apo structure, the search model for phasing was a monomer of the (acetyl-)CoA bound FadA5 WT structure (4.2.4.2). Subsequent model building was performed by alternating manual model building in Coot (Emsley and Cowtan, 2004) and refinement steps in Refmac5 (Murshudov et al., 1997).

Apo structure I contains two molecules per asymmetric unit with 385 residues in chain A and all 391 residues in chain B. 161 water molecules were added and other solvent molecules such as glycerol, dimethyl sulfoxide, sodium and chloride were identified. The data were fully refined to R_{work} and R_{free} values of 17.1% and 24.8% respectively. Validation of the data set was performed with program tools in Coot and the independent MolProbity server (Chen et al., 2010), where clashing atoms, deviations from bond geometry or bond length or rotamer and Ramachandran outliers are reported. The final analysis shows bond and angle deviations of 1.6953° and 0.0122 \AA and an overall clash score of 11.83, which means that apo structure I is better than 94% of all PDB-released entries of the same resolution with respect to clashes. The clash score is defined as the number of serious steric overlaps ($>0.4 \text{ \AA}$) per 1000 atoms. The final analysis of the Ramachandran plot shows that 93.2% of all residues follow the most favored backbone geometry, whereas 0.8% of the residues are in the disallowed geometry region. These outliers are mainly (3 out of 4) located in regions with interrupted electron density and could therefore not be modeled correctly. The complete data collection and refinement statistics are shown in **Table 33**.

Another group of apo crystals (II) which were obtained at lower pH (see cryoprotectant solution) were harvested, transferred with a nylon loop into the according cryoprotectant solution (0.1 M citrate pH 5.5, 1.0 M $(\text{NH}_4)_2\text{HPO}_4$, 0.2 M NaCl, 25% (v/v) glycerol) and flash-frozen in liquid nitrogen. The crystals belong to space group $P6_322$ and a dataset was collected at the ESRF at a wavelength of 0.873 \AA . The diffraction data were collected to 2.19 \AA resolution and the unit cell dimension were $120.29 \text{ \AA} \times 120.29 \text{ \AA} \times 206.03 \text{ \AA}$ and $90^\circ, 90^\circ, 120^\circ$. The data were processed with iMosflm, scaled with Scala and molecular replacement (MR) was performed with phaser. The initial model for phasing was a monomer of the (acetyl-)CoA bound FadA5 WT structure (4.2.4.2). Phasing was repeated with a monomer of the OPC-complex (4.2.4.4), which represents the FadA5 structure at the highest resolution. Subsequent model building was performed by alternating manual model building in Coot (Emsley and Cowtan, 2004) and refinement steps in Refmac5 (Murshudov et al., 1997). Anomalous maps for sulfur localization were calculated with the help of Cad and FFT (CCP4 package (Winn et al., 2011)).

Results

Table 33 - Data collection and refinement statistics *FadA5 WT apo*

	<i>FadA5 WT apo</i>	<i>FadA5 WT apo</i>
Description	apo I	apo II
Relevant buffers		
Protein buffer	20 mM Bicine-NaOH pH 8.5, 250 mM NaCl	60 mM Bicine-NaOH pH 8.5, 250 mM NaCl
Crystallization condition	0.1 M Mes pH 6.1, 5% (v/v) DMSO, 1.8 M (NH ₄) ₂ SO ₄	0.1 M citrate pH 5.5, 0.7 M (NH ₄) ₂ HPO ₄ , 0.3 M NaCl
Cryoprotectant solution	0.1 M MES pH 6.1, 2.0 M (NH ₄) ₂ SO ₄ , 5% (v/v) DMSO, 27% (v/v) glycerol	0.1 M citrate pH 5.5, 1.0 M (NH ₄) ₂ HPO ₄ , 0.2 M NaCl, 25% (v/v) glycerol
Data collection		
Wavelength (Å)	0.9184	0.873
Temperature (K)	100	100
Space group	P4 ₁ 2 ₁ 2	P6 ₃ 22
Unit cell parameters		
a/b/c (Å); α/β/γ	128.31 / 128.31 / 114.07; 90° / 90° / 90°	120.29 / 120.29 / 206.03; 90° / 90° / 120°
Resolution (Å)	52.0-2.7	46.17-2.19
Number of subunits	2	2
Total reflections	190,844 (20,941)	451,806
Unique reflections	25,129 (3,358)	46,017
Completeness (%)	94.4 (99.5)	100 (100)
∅ Redundancy	7.6 (7.7)	9.8 (9.8)
<I / σ (I)>	11.8 (2.2)	15.4 (2.7)
R _{merge} ^a (%)	14.3 (85.3)	10.1 (89.3)
Wilson B-factor(Å ²) ^b	37.4	37.8
Refinement		
Total number of atoms	6,172	5,514
R _{work} / R _{free} (%)	17.1 / 24.8	20.4 / 25.8
RMSD		
bond angle (°)	1.6953	1.0603
bond length (Å)	0.0122	0.0078
Estimated coordinate error (Å) (Refmac / Phenix)	0.24	0.15
Average B factor (Å ²) ^c	42.6	35.5
Most favored (%)	93.2	88.5
Allowed (%)	6.0	7.5
Disallowed (%)	0.8	4.0
PDB code	4ubw	-
Values in parentheses refer to the highest-resolution shell		
^a R _{merge} = $\sum_{hkl} \sum_i I_i - \langle I \rangle / \sum_{hkl} \sum_i I_i$		
^b truncate (French and Wilson, 1978); ^c baverage (Winn et al., 2011); ^d (Chen et al., 2010)		

Apo structure II contains two molecules per asymmetric unit with 367 residues in chain A and 361 residues in chain B. 88 water molecules and some chloride ions were identified. The data were refined to *R_{work}* and *R_{free}* values of 20.4% and 25.8% respectively. The structure was validated and clearly indicated significant problems in geometry and clashing. In addition the electron density of monomer B is undefined in large areas, especially in comparison to monomer A, and is worse than expected for data at 2.2 Å resolution. This undefined electron density probably reflects enzyme

Results

regions with higher atomic disorder. Interestingly the apo II structure contains two disulfide bridges per monomer which were not present in the apo I structure. It is very likely that the formation of these disulfide bridges led to a reorientation of nearly 30 amino acids (region S67 to C91) and in addition to a disorder of 20 amino acids (region G127 to I146) in both monomers. Combined, this resulted in clashes and Ramachandran outliers in the disulfide bridge regions and problems with the geometry of residues which were in close proximity to the disordered region. Ramachandran outliers and clashes were tried to be resolved but due to the strained geometry near the S-S-bridges the applied changes were reversed in the refinement processes. Therefore the current bond and angle deviations are 1.0603° and 0.0078 Å but the overall MolProbity score is with 3.44 very low, indicating that this structure is worse than 94% of all PDB-released entries at the same resolution concerning overall geometry and clashes. The Ramachandran plot shows that 88.5% of all residues follow the most favored backbone geometry, but 4% remain in the disallowed geometry region. The complete data collection and refinement statistics are shown in [Table 33](#).

4.2.3.3 Structural overview

The initial apo structure of FadA5 shows a typical three sub-domain architecture, as described by Sundaramoorthy *et al.* (Sundaramoorthy et al., 2006) This three sub-domain model consists of the N-terminal sub-domain I, with residues 1 to 125 and 246 to 266, sub-domain II, with residues 126 to 245, and the C-terminal sub-domain III, with residues 267 to 391. Sub-domains I and III form the stable and rather inflexible platform of the protein and are structurally and functionally related. They are both characterized by a central β -sheet and surrounding helices. The central β -sheet, consisting of β 1, β 3, β 4, β 5 and β 8 in sub-domain I and β 9, β 10, β 11 and β 12 in sub-domain III, is embraced by the larger α helices α 1, α 2 and α 3 in the first sub-domain and by α 8, α 9 and α 11 in the third sub-domain ([Figure 48A](#)). The two subdomains harbor the catalytic triad of the thiolase, which are cysteine 93 at the very beginning of η 3 ($\eta = 3_{10}$ helix) in sub-domain I and histidine 347, encased by α 10 and α 11, plus cysteine 377 at the end of β 11 that are both located in sub-domain III ([Figure 48B](#)). The active site residues are located on top of the two described domains but they are occluded from the solvent by sub-domain II, which is also called the lid domain. (Sundaramoorthy et al., 2006) The lid sub-domain, characterized by a more extended structure is comprised of two shorter β -strands, β 6 and β 7, and three helices, namely α 4, α 5 and α 6. These secondary structure elements are interrupted by larger non-structured regions.

A remarkable fact is the position of cysteine 93 within helix η 3 and right before α 3. C93 is the nucleophile of the thiolase reaction and deprotonation by H347 leads to a reactive thiolate that attacks the β -ketoester of the substrate ([Figure 11](#) state I). The deprotonation is likely to be

Results

supported by the dipole moment of both helices which might lower the pK_a of C93. (Hol, 1985; Lodi and Knowles, 1993; Price et al., 2003)

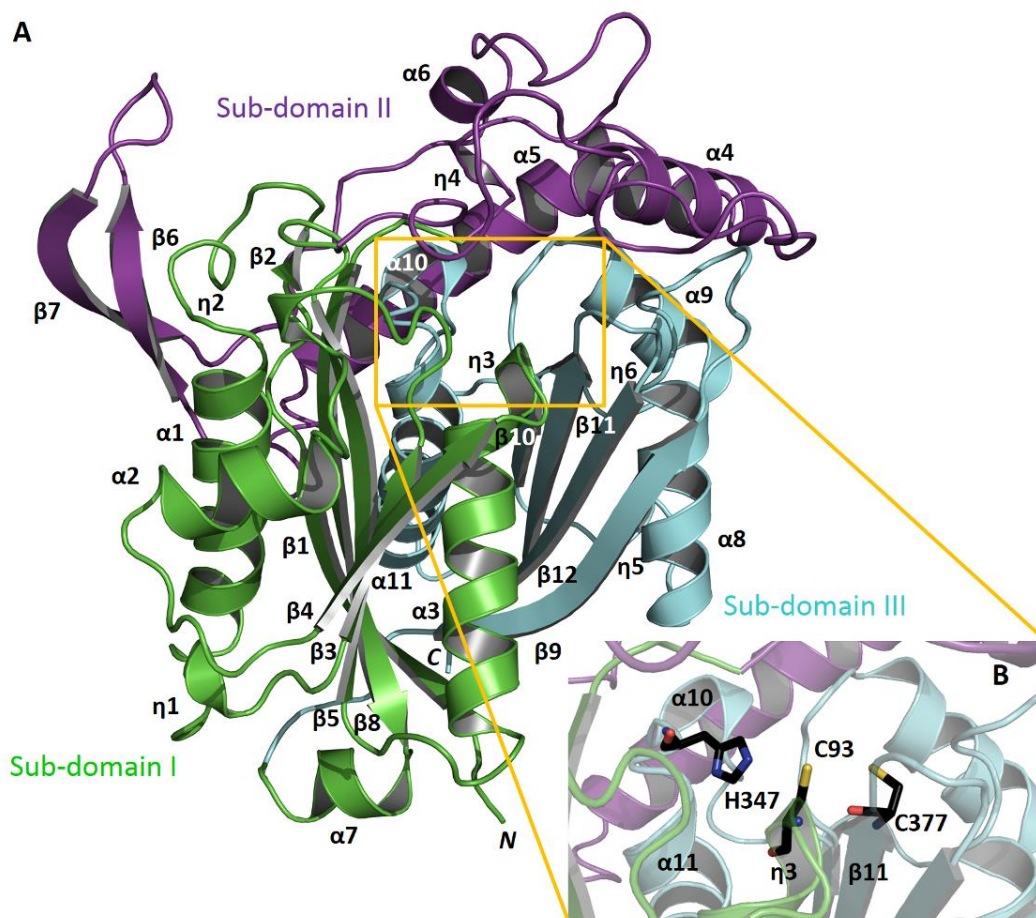


Figure 48 - Overall structure of FadA5 WT apo I:

A) The sub-domain structure of the FadA5 apo monomer is shown as a cartoon, consisting of the N-terminal sub-domain (I, green), the lid sub-domain (II, purple) and the C-terminal sub-domain (III, cyan).

B) Zoom into the thiolase active site, built by residues C93, H347 and C377, which are buried between sub-domains I and III and are 'occluded' from the solvent by sub-domain II. Residues are shown in stick presentation. ($\eta = 3_{10}$ helix)

As already observed in SEC-MALLS experiments (4.1.4.3), FadA5's functional composition is a dimer (**Figure 49A**). The dimeric structure was also observed for other thiolases but some thiolases, do not only form dimers but weaker interacting dimers of dimers with the help of an extended (tetramerization) loop region. (Haapalainen et al., 2006) Tetramerization is mainly a feature of biosynthetic thiolases but some degradative enzymes are also known to form tetramers. The tetramerization loop (arrows **Figure 49B**) is for example present in the degradative human peroxisomal AB-thiolase (AB, PDB code 2iik) or in the bacterial biosynthetic thiolase from *Zoogloea ramigera* (PDB code 1dlu (Modis and Wierenga, 2000)). Interestingly, the thiolase fold seems to be highly conserved, despite the fact that the sequence identity among the thiolases is rather low (**Figure 49B, Table 34**).

Results

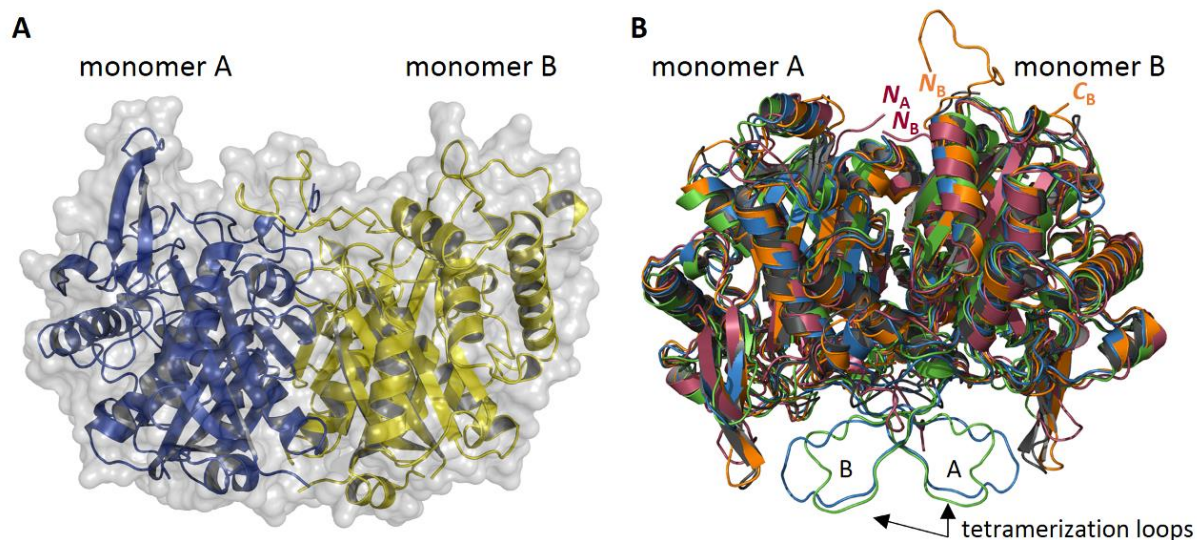


Figure 49 - Dimeric structure of thiolases

A) FadA5 forms the typical thiolase dimer (blue = chain A, yellow = chain B) **B)** Superposition of thiolases that share a sequence identity between 37 and 40% to FadA5. Despite the lack of sequence identity, the root mean square deviations (RMSD) among them are low, ranging from 1.15 to 1.37 Å, and therefore the structural similarity amongst each other is very high according to PDBeFold (Krissinel and Henrick, 2004) (Table 34). Visible N- and C-termini are indicated by arrows and labelled in the according color code (chain identifier assigned by subscripts). The tetramerization loops of the two monomers are indicated by arrows. (FadA5 (dark red); 1ulq (*Thermus thermophilus*, blue); 1dlu (*Zoogloea ramigera*, green); 2iik (*Homo sapiens*, orange); 2wua (*Helianthus annuus*, grey))

Table 34 - RMSD value comparison: FadA5 apo I compared to similar thiolases (left) and to the FadA5 complex structures (right).

	RMSD	structure similarity	sequence identity	sequence homology		N_{res}	N_{SSE}	RMSD
	[Å]		(%)					[Å]
FadA5 WT apo	chain B apo				FadA5 WT apo	776	44	0.37
<i>T. thermophilum</i> 1ulq	1.17	84	43	55	acetylated FadA5 WT-(acetyl-)CoA complex	783	46	0.30
<i>H. annuus</i> 2wua	1.33	86	37	54	acetylated FadA5 C93S-CoA complex	782	52	0.52
<i>H. sapiens</i> 2iik	1.37	92	38	53	FadA5 C93S-OPC complex	788	46	0.33
<i>Z. ramigera</i> 1dlu	1.15	80	38	53	1ulq	796	54	1.01
N_{res}: Number of analyzed residues								
N_{SSE}: Number of secondary structure elements								

Results

4.2.3.4 The other 'apo' structure (apo II)

The second apo structure immediately reveals two striking differences to the above analyzed structure. The protein dimer interface is differently oriented (**Figure 50A/B**), although the size of the surface area is similar with 1,969 Å² for apo I and 2,070 Å² for apo II, respectively (Krissinel and Henrick, 2007), and the original orientation of the η2-α2-β4 motif is changed or reversed with respect to the sub-domain I orientation. This three secondary structure elements containing turn is running upwards, towards sub-domain II, and is thereby forming an additional interface between the

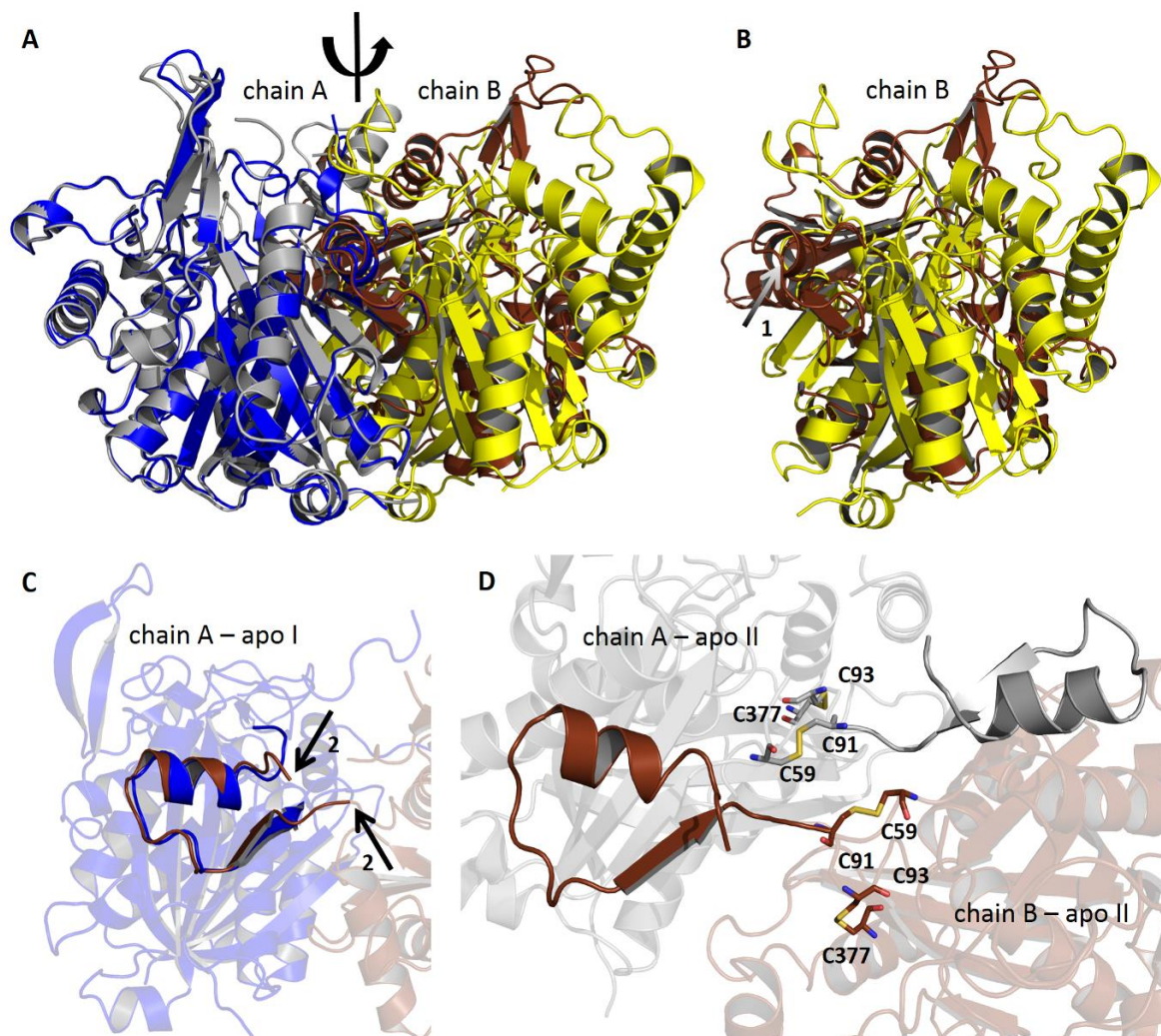


Figure 50 - Comparison of the two apo FadA5 structures:

A) Superposition of the apo I (chain A, blue; chain B, yellow) and apo II (chain A, grey; chain B, brown) dimers. The monomers of apo I are related to each other by a 2-fold axis. Conversion between the monomers of apo II by the same axis is not possible. **B)** Orientation of the two B chains, resulting from the superposition of the A chains. The yellow (apo I) and the brown (apo II) monomer are clearly not superimposing. Arrow 1 indicates the orientation of the swapped domain from residues S67 to C91. **C)** Arrows 2 indicate the starting and the end points of the reoriented domain. The blue domain is folding back to chain A of apo I whereas the brown domain is coming from chain B of apo II. These highlighted cartoons show that in the apo II structure a domain swap to the adjacent monomer was observed. Transparent cartoon representation of chain A apo I (blue) and chain B apo II (brown). **D)** The apo II dimer is shown in cartoon presentation with highlighted swapped domains. The cysteines forming disulfide bonds in each monomer are shown in stick presentation. Color code as described in A.

Results

two monomers (Figure 50C/D). This reorientation and domain swap leading to a reorientations of the catalytic cysteines is most likely due to the formation of two disulfide bonds which are formed between C59 and C91 as well as C93 and C377 (Figure 50D).

Superposition of chain A of each dimer clearly shows that the two monomers in apo I, seem to be convertible by a 2-fold axis (indicated in Figure 50A) whereas the same 2-fold axis is not present any longer for the conversion of the apo II monomers. Furthermore, superposition of monomers A of each dimer shows that monomers B assume different positions relative to monomer A. Arrow 1 points to the orientation of the swapped domain in apo II.

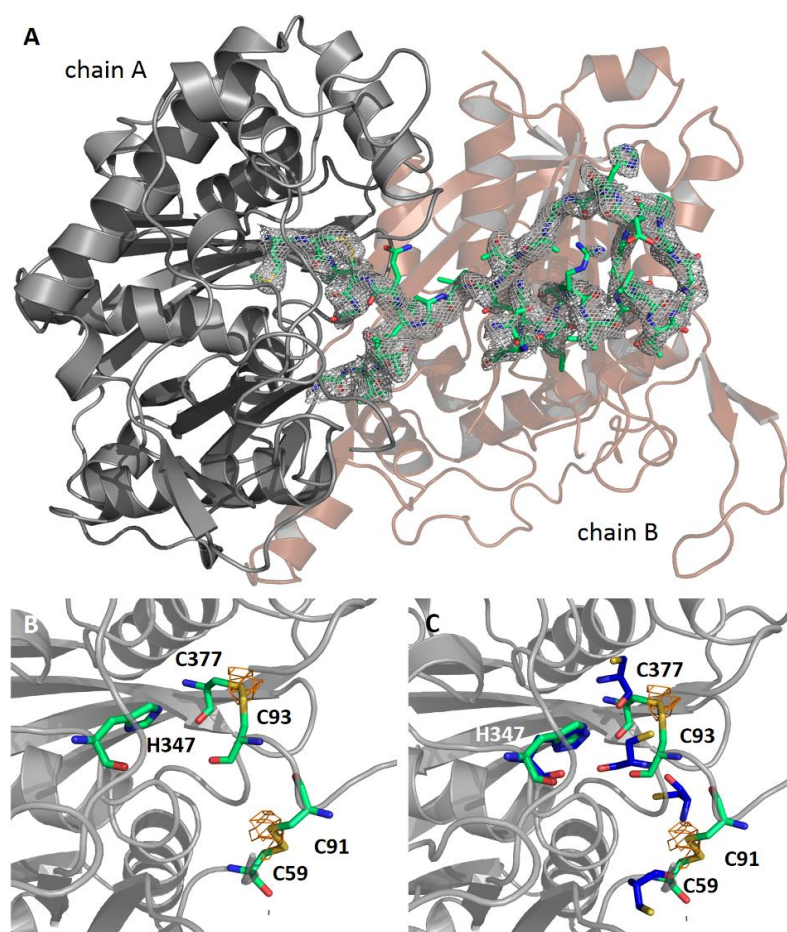


Figure 51 - Details of apo structure II and its active site:

A) SIGMAA-weighted $2F_o-F_c$ electron density map ($\sigma = 1.0$) of the swapped residues S67 to C91 (green) and the residues surrounding the disulfide bridges. The protein is shown in cartoon presentation, chain A in grey and chain B in brown (transparent). B/C) Detail of the catalytic triad (green) and the disulfide bridges with anomalous electron density ($\sigma = 2.9$; orange) in apo structure II and superimposed with apo structure I (blue).

In contrast to that, continuous electron density between S67 and C91 is observed in the model where the domain swap was built. Residues Q62 to Q66 (chain A) or F63 to Q66 (chain B) are not resolved for respective monomer. However, the transition between T88 and

The general three sub-domain architecture is conserved in the apo II structure with the exception of the described $\eta 2$ - $\alpha 2$ - $\beta 4$ motif, which does not fold back into its own monomer but extends to the other monomer and replaces the $\eta 2$ - $\alpha 2$ - $\beta 4$ motif of the other monomer. It is difficult to evaluate, if the domain swap is a result of the crystallization process or may also be present in solution, since the region of the crossing domains is also the region where the disulfide bonds cause a significant reorientation in comparison to the other apo structure and the electron density for some residues is not well defined. Model building and refinement cycles with and without the domain swap, show a gap in the electron density for

Results

C91 is clearly resolved in the model with the domain swap (**Figure 51A**) whereas it is not in the model without the swap. It is also likely, that the domain swap was necessary to stabilize the shifted dimer, as only the swap leads to a similar buried surface area compared to the apo I structure and thereby makes dimerization probable. In comparison, the calculated buried surface area for apo II without a domain swap is with 783 Å² comparably low and complex formation would not be very likely.

In addition to the intensive examination of the domain swap anomalous sulfur data were investigated to verify the location of the disulfide bridges and to exclude alternative orientations of the involved cysteines. **Figure 51B** clearly shows that anomalous density (orange) is present at the position of the disulfide bridges and thereby confirms their presence. As indicated above, the formation of these disulfide bonds is likely to cause the domain swap and the strained geometry in the regions around the disulfide bonds. Superposition of the catalytic triad from the two apo structures (**Figure 51C**) shows that H347 adopts the same conformation in both structures but the catalytic cysteines and the involved C59 residues occupy altered positions.

4.2.4 Crystal structures of FadA5 in complex with ligands

A general aim with respect to the characterization of FadA5 was to obtain insight towards the feasibility to specifically target this thiolase without targeting the human enzymes. After the initial structural characterization, complex structures with ligands such as acetyl-CoA or even steroid-CoAs were therefore pursued to permit the identification and analysis of the binding pockets.

4.2.4.1 Crystallization

FadA5 WT, FadA5 C93S and FadA5 C93A, in SE buffer 5 (20 mM Bicine-NaOH pH 8.5, 250 mM NaCl), were used for crystallization at concentrations of approx. 12 mg/ml. The ligands, acetyl-CoA, 3-oxo-pregn-4-ene-20-carboxyl-CoA (3-OPC-CoA) or 3-oxo-cholesterol-4,22-diene-24-oyl-CoA (3-OCD-CoA), were added prior to crystallization in 5 to 10-fold molar excess.

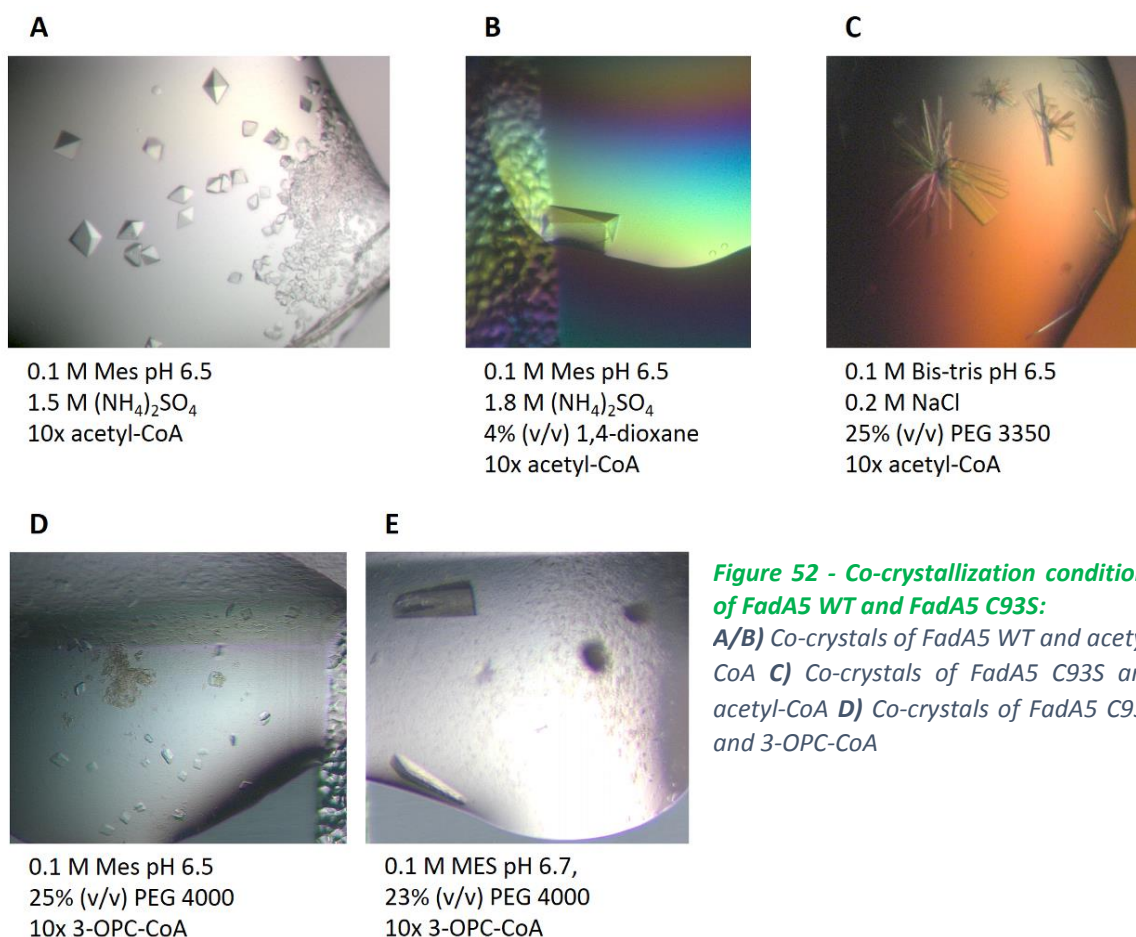
Initial crystallization screens were performed with all available commercial screens, i.e. Crystal Screen I/II, Index Screen, Topaz OptiMix I-V, Topaz OptiMix-PEG, Nextal PEG and Wizard Screen I/II in the 96-well format. The screening experiments with 3-OPC-CoA and 3-OCD-CoA were in addition performed with the Nextal pHClear Suite, Nucleix Suite and the self-designed FadA5 and Thiolase Screen.

Similar to the crystallization of apo FadA5, the co-crystallization approaches led to several different crystallization conditions. Some of them were not reproducible, the crystals did not diffract well or

Results

produced well diffracting crystals in the aforementioned problematic space group. The predominant crystallization results of the acetyl-CoA co-crystallization trials led to crystals that could not be used. However, a new crystallization condition, with $(\text{NH}_4)_2\text{SO}_4$ at higher concentrations eventually produced crystals in space group P4_32_12 (**Figure 52A**). Follow up fine screens with different additives and the WT enzyme led to an asymmetric looking crystal, of the same space group, that diffracted to 1.95 \AA (**Figure 52B**). Co-crystallization trials with FadA5 C93S led to thick plate-like crystals (**Figure 52C**) in space group P12_11 that diffracted to 3.0 \AA .

Co-crystallization with 3-OPC-CoA was only successful for the FadA5 C93S variant. Initial thin plate-like crystals (**Figure 52D**), grown in a new PEG condition, could be optimized in a secondary screen and led to thick trapezoid crystals. These crystals (**Figure 52E**) diffracted to 1.7 \AA and the space group was again P12_11 .



Results

4.2.4.2 Acetylated FadA5 WT(-acetyl)-CoA complex

4.2.4.2.1 Data collection, structure solution and validation

For data collection, the crystal from **Figure 52B** was harvested, transferred with a nylon loop into the according cryoprotectant solution (0.1 M Mes pH 6.5, 1.6 M $(\text{NH}_4)_2\text{SO}_4$, 10% (v/v) 1,4-dioxan, 25% (v/v) glycerol + 10x acetyl-CoA) and flash-frozen in liquid nitrogen. The data set with space group $P4_32_12$ was collected at the BESSY II beamline MX 14.1 at a wavelength of 0.9184 Å. The obtained resolution is 1.8 Å and the unit cell dimension are 124.68 Å x 124.68 Å x 124.71 Å and 90°, 90°, 90°. The data were processed with iMosflm, scaled with Scala, where the data were truncated to 2.0 Å, and molecular replacement (MR) was performed with phaser. Since this was the first FadA5 structure solved the model used for phasing was the structure of a thiolase from *T. thermophilus* (PDB code 1ulq), which was the model with the highest sequence identity among the different thiolases at the time of the MR approach. Initial automated model building was performed with ARP/wARP in the course of which the model sequence was automatically replaced by the FadA5 sequence. Later, the dataset was rescaled with Aimless and thereby truncated at 1.95 Å, as the density could be improved significantly and the R_{merge} as well as the correlation coefficient $\text{CC}_{1/2}$ still indicated a good data quality. Subsequent model building was performed by alternating manual model building in Coot (Emsley and Cowtan, 2004) and refinement steps in Refmac5 (Murshudov et al., 1997). Fitting of the ligands into the electron density was facilitated with ligand PDB and geometry files from the Coot library. The geometry file contains the ligand's geometry restraints necessary for real space refinement by Coot and during refinement cycles in Refmac5.

The structure contains two molecules per asymmetric unit with all 391 residues in both chains and one additional serine (S0) in chain B, originating from the expression tag. 281 water molecules were added and other solvent molecules such as glycerol, 1,4-dioxane and sulphate were identified and added as well. In addition to the solvent molecules electron density could be assigned for the co-crystallized ligand. In chain A, density for acetyl-CoA and CoA was present in the same binding pocket. Both ligands were therefore modelled with reduced occupancy and refined. The CoA-binding pocket of chain B also showed additional, positive difference electron density and CoA was modelled in two alternative conformations, with one at higher occupancy (0.4) than the other (0.3). Another remarkable aspect is the partial acetylation of C93 in both chains. The data were fully refined to R_{work} and R_{free} values of 19.3% and 23.1% respectively. Validation of the data set was performed with program tools in Coot and the independent MolProbity server (Chen et al., 2010), where clashing atoms, deviations from bond geometry or bond length or rotamer and Ramachandran outliers are reported. The final analysis shows bond and angle deviations of 1.3581° and 0.0114 Å and an overall clash score of 13.55, indicating that this structure is better than 62% of all PDB-released entries of the same resolution concerning clashes. The clash score is defined as the number of serious steric

Results

overlaps ($>0.4 \text{ \AA}$) per 1000 atoms. The clash score was comparably high, as the program was not able to distinguish between an acetyl-CoA ligand and a CoA ligand that occupy the same binding pocket with reduced occupancy. The final analysis of the Ramachandran plot shows that 97.9% of all residues follow the most favored backbone geometry and the remaining residues assume a position which is not optimal but still in the allowed region of the Ramachandran plot. The complete data collection and refinement statistics are shown in **Table 35**.

4.2.4.2.2 *Structure overview and ligand binding*

The overall fold of the structure follows the fold of apo FadA5 as described in 4.2.3.3, this is also supported by the analysis with PDBeFold that shows an RMSD of only 0.32 \AA , between the $C\alpha$ atoms of the chains, between the two models.

Both monomers showed prominent additional electron density at the location of the catalytic residue C93. This indicated two alternative, partially occupied conformations of this residue but as still additional density was visible for one of the conformations an acetylation (SCY93), caused by transfer of an acetyl group from acetyl-CoA to the enzyme, was modeled. Active site acetylation of thiolases has been described by Modis et al. and ligand-binding in the *Z. ramigera* biosynthetic thiolase has been investigated in depth and it was therefore not unanticipated to observe this modification. (Kursula et al., 2002; Modis and Wierenga, 1999, 2000) In addition to the partially acetylated C93 both monomers interact with a ligand that is also bound in different orientations. Chain A interacts with a mixture of CoA and acetyl-CoA which have been refined with a combined occupancy of 90%, 40% CoA and 50% acetyl-CoA respectively (**Figure 53A**). Chain B contains CoA only, but again, the ligand is present in two alternative conformations with occupancies of 40% and 30% respectively (not shown). The electron densities of all ligands are not very well defined and further conformations might be possible. The ligand densities were not the only difficult to interpret part of this structure, in addition some diffuse electron density can be observed in a region, later identified as the steroid-binding pocket, which could not be explained by modeling solvent molecules or ligands into it. Hence this density remained uninterpreted.

The active site residues are connected to each other by direct and water-mediated interactions. C93 is in direct contact to H347, but C377 is only interacting via water molecule Z253, as it is located farther away. Water molecule Z253 is of importance, as it is described in the literature to be part of the oxyanion hole I, which is especially important for the biosynthetic reaction, and it is annotated as a unique feature of the thiolase active site. (Haapalainen et al., 2006; Kursula et al., 2005) SCY93 does not interact with other residues within the active site.

Results

The described higher occupancy of a ligand compared to an alternative conformation can readily be explained by a higher number of direct and indirect hydrogen bond contacts between the enzyme and the ligand in comparison to the other ligand or its other orientation. Acetyl-CoA is thus embedded in the protein by three direct (Q151, R221 and H347) and twelve water-mediated contacts. The contact formed by R221 is a typical parallel π -cation stacking interaction with a distance of 3.7 Å between the guanidinyll function of R221 and the CoA nucleotide heterocycle (**Figure 53B**, orange ellipse). A remarkable fact of the ligand binding is that half of the interactions are provided by residues of sub-domain II (**Figure 53B**, purple residues), i.e. the active site occluding domain. Acetyl-CoA is also interacting with the catalytic triad, C93 and C377, via the Z253 water molecule. In contrast to this well interacting ligand, CoA of chain A is only connected by the weaker, π -cation stacking interaction with R221 and five water-mediated hydrogen bond bridges to the protein whereas the other interactions are missing due to geometry reasons. The higher occupied CoA of chain B forms 14 water-mediated or direct contacts, including indirect contacts to the active site, with the protein. The lower occupied CoA is more weakly connected, as only six protein-ligand interactions can be identified, without involvement of the catalytic triad.

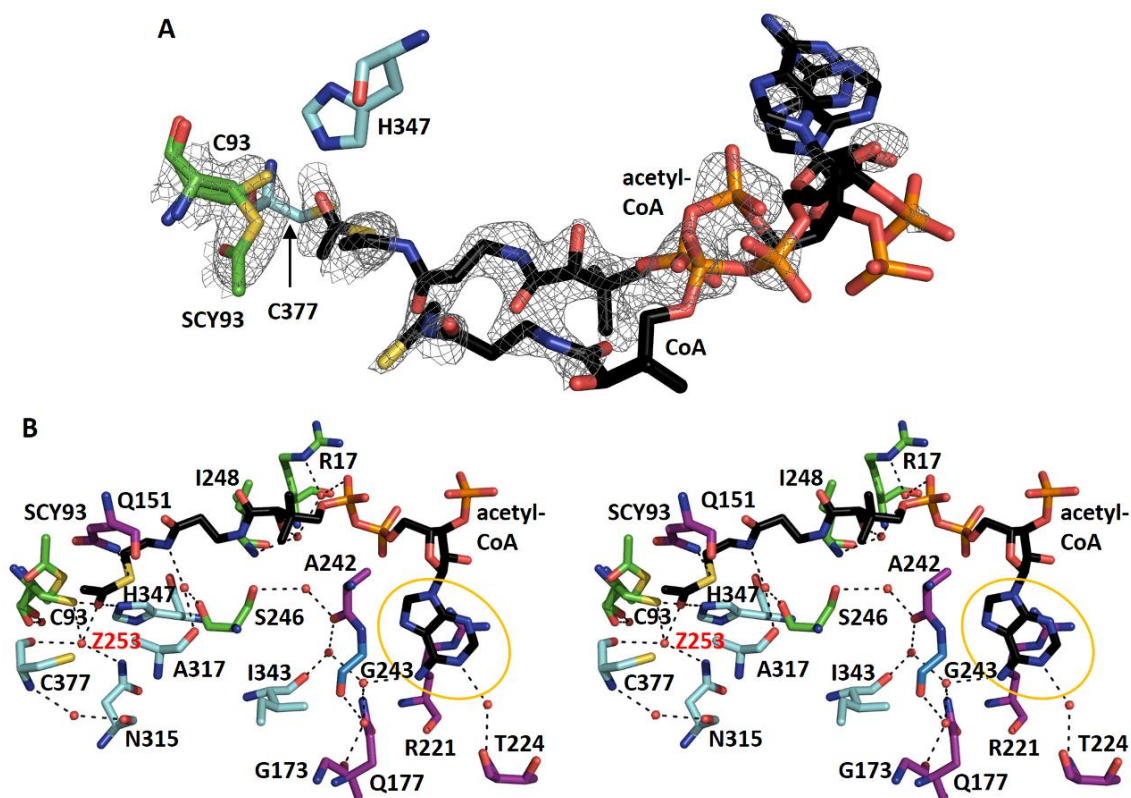


Figure 53 - Electron density and interaction network in the acetylated FadA5 WT(-acetyl)-CoA complex

A) SIGMAA-weighted $2F_o-F_c$ electron density omit map ($\sigma = 0.9$) of acetyl-CoA and CoA (both black) as well as the native (C93, green) and the acetylated C93 (SCY, green) residue. The active site residues are colored in the respective sub-domain color code. **B)** Interaction network of acetyl-CoA (black) and FadA5 shown in stereo. Direct ligand-protein contacts, water-mediated (red spheres) hydrogen bonds (dashed lines) and a π -cation-stacking interaction (orange ellipse) of the adenosine moiety of CoA with R221 can be observed. The catalytic triad interacts with each other via direct and water mediated contacts, a central position is assumed by water molecule Z253.

Results

4.2.4.3 Acetylated FadA5 C93S-CoA complex

4.2.4.3.1 Data collection, structure solution and validation

For data collection, a crystal plate was removed from the crystal bundle, shown in **Figure 52C**, and was fished with a nylon loop. The crystal was transferred into its cryoprotectant solution (0.1 M Bis-tris pH 6.5, 0.3 M NaCl, 27% (v/v) PEG 3350, 25% (v/v) glycerol + 10x ACO) and flash-frozen in liquid nitrogen. A data set with space group $P12_11$ was collected at the ESRF beamline BM14U at a wavelength of 0.9537 Å. The obtained data diffracted to 3.0 Å and the unit cell dimension were 127.61 Å x 105.16 Å x 147.66 Å and 90°, 107.33°, 90°. The data were processed with iMosflm, scaled with Scala and molecular replacement (MR) was performed with phaser. The model for phasing was a monomer of the acetylated FadA5 WT(-acetyl)-CoA complex structure. The refinement quality was significantly improved by the definition of a thin shell at low resolution. Subsequent model building was performed by alternating manual model building in Coot (Emsley and Cowtan, 2004) and refinement steps in Refmac5 (Murshudov et al., 1997). The application of NCS refinement with fixed isotropic B-factors proved to be a valuable tool in the refinement process of this structure. Eight protein chains were present in the asymmetric unit, and thus NCS averaging improved the quality of the overall electron density significantly. Fitting of the ligands into the electron density was facilitated with ligand and geometry files from the Coot library. The geometry file contains the restraints necessary for real space refinement by Coot and during refinement cycles in Refmac5.

All 391 residues in every chain were modelled and chains E, G and H in addition contain one or more residues of the expression tag. At this resolution (3 Å) only 39 water and some glycerol molecules could be modelled, but in addition, positive difference density could be assigned to bound CoA ligands, although the quality of the electron density differed among the subunits. The CoA electron density was refined with the help of NCS averaging of this specific ligand. Another remarkable aspect is the acetylation of every S93 residue (OAS93) in all chains. The data were fully refined to R_{work} and R_{free} values of 24.7% and 28.7% respectively. Validation of the data set was performed with program tools in Coot and the independent MolProbity server (Chen et al., 2010). The final analysis shows bond and angle deviations of 1.4556° and 0.0123 Å and an overall clash score of 6.1, which results in the fact that this structure is better than 99% of all PDB-released entries at the same resolution concerning clashes. The final analysis of the Ramachandran plot shows that 95.6% of all residues followed the most favored backbone geometry and only four residues (0.1%) are in the outlier region. Half of the outliers are located in regions at the expression tag with less defined electron density, which hindered correct geometry mapping. The complete data collection and refinement statistics are summarized in **Table 35**.

Results

4.2.4.3.2 *Structure overview and ligand binding*

The overall fold of the structure follows the fold of apo FadA5 as described in 4.2.3.3, which is also supported by the analysis with PDBeFold that showed an RMSD of only 0.64 Å, between the C α atoms of the chains, between the two models. However, this FadA5 structure displays the highest discrepancies compared to the apo structure.

The model revealed, similar to the acetylated FadA5 WT(-acetyl)-CoA structure, additional density near the oxygen of S93 in all eight chains. This density is interpreted as an acetylation of the serine, which results in the residue descriptor OAS93, again caused by transfer of the acetyl group from acetyl-CoA to the enzyme (**Figure 54**). In contrast to the acetylated FadA5 WT(-acetyl)-CoA complex only the reaction product of the enzyme's acetylation, CoA, is bound to FadA5. In general, the obtained electron densities for CoA are of mixed quality, also due to the low resolution, but NCS averaging of the ligand permitted modeling of CoA in all chains (**Figure 54**). The ligand is present in one conformation but hardly any contacts with the enzyme can be observed (not shown). In none of the eight chains, CoA and the catalytic triad form interactions or interact amongst each other. The lack of visible interactions is probably caused by the low resolution and the therefore low number of resolved water molecules, which are responsible for the most of the interactions in the acetylated FadA5 WT(-acetyl)-CoA complex. However, a proper π -cation stacking interaction of CoA with R221 is also not observed. For all enzyme chains, the interaction planes are not parallel, which at least causes a direct hydrogen-bond interaction between CoA and R221 but prevents the stacking interaction.

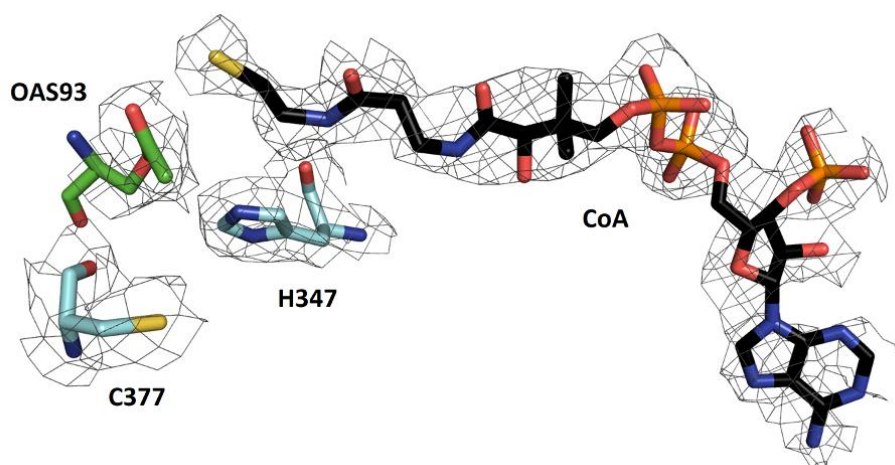


Figure 54 - OAS93 and CoA electron density in the FadA5 C93S-CoA-complex

SIGMAA-weighted $2F_o - F_c$ NCS averaged electron density maps ($\sigma = 1.0$; grey) of the bound CoA (black) and the active site residues for the enzyme chain A. Maps were generated for chain A with NCS averaging in COOT (Emsley and Cowtan, 2004). The catalytic triad is depicted in the sub-domain color code.

Results

4.2.4.4 FadA5 C93S-OPC complex

4.2.4.4.1 Data collection, structure solution and validation

For data collection, a trapezoid crystal, shown in **Figure 52E**, was fished with a nylon loop, transferred into the cryoprotectant solution (0.1 M MES pH 6.7, 27% (v/v) PEG 4000, 25% (v/v) PEG 400 + 10x 4-OPC-CoA) and flash-frozen in liquid nitrogen. A data set with space group P12₁1 was collected at the BESSY II beamline MX 14.1 at a wavelength of 0.9184 Å. The obtained data diffracted to 1.7 Å and the unit cell dimension were 76.82 Å x 100.39 Å x 107.98 Å and 90°, 100°, 90°. The data were processed with iMosflm, scaled with Scala and molecular replacement (MR) was performed with phaser. The model used for phasing was a monomer of the FadA5 WT(-acetyl)-CoA complex structure. Subsequent model building was performed by alternating manual model building in Coot (Emsley and Cowtan, 2004) and refinement steps in Phenix (Adams et al., 2010). Fitting of the ligands into the electron density maps was facilitated with ligand PDB and geometry files generated by the PRODRG server (Schuttelkopf and van Aalten, 2004). The geometry file contains the restraints necessary for real space refinement by Coot and during refinement cycles in Refmac5.

The model comprises four chains and in three of them all 391 residues could be modelled. Chains A and B contain in addition one or more residues of the expression tag. Chain D lacks residues G133 to D147 of sub-domain II, similar as observed for chain A of apo structure I. 1161 water molecules and several glycerol and PEG molecules as well as chloride and sodium ions were added. Additional electron density was present and assigned to Coenzyme A, which is a hydrolysis product of the co-crystallized 3-OPC-CoA. No density was identified for the complete 3-OPC-CoA ligand (**Figure 10**) but in addition to CoA, the other hydrolysis product 3-oxo-pregn-4-ene-20-carboxylic acid (OPC) was assigned to difference electron density in chains A to C. The data were fully refined to R_{work} and R_{free} values of 14.5% and 17.4% respectively. Validation of this data set was performed with program tools in Coot and the independent MolProbity server (Chen et al., 2010). The final analysis shows bond and angle deviations of 1.096° and 0.007 Å and an overall clash score of 5.0, which results in the fact that this structure is better than 93% of all PDB-released entries of the same resolution concerning clashes. The final analysis of the Ramachandran plot shows that 97.3% of all residues follow the most favored backbone geometry and only three residues (0.2%) are located in the outlier region. These outliers can be found in regions with less defined electron density, as the expression tag, which hinders correct geometry mapping. The complete data collection and refinement statistics are summarized in **Table 35**.

Results

4.2.4.4.2 *Structure overview and ligand binding*

The overall fold of the structure follows the fold of apo FadA5 as described in 4.2.3.3. This is also supported by the analysis with PDBeFold which reveals an RMSD of only 0.32 Å, between the C α atoms of the chains, between the two models.

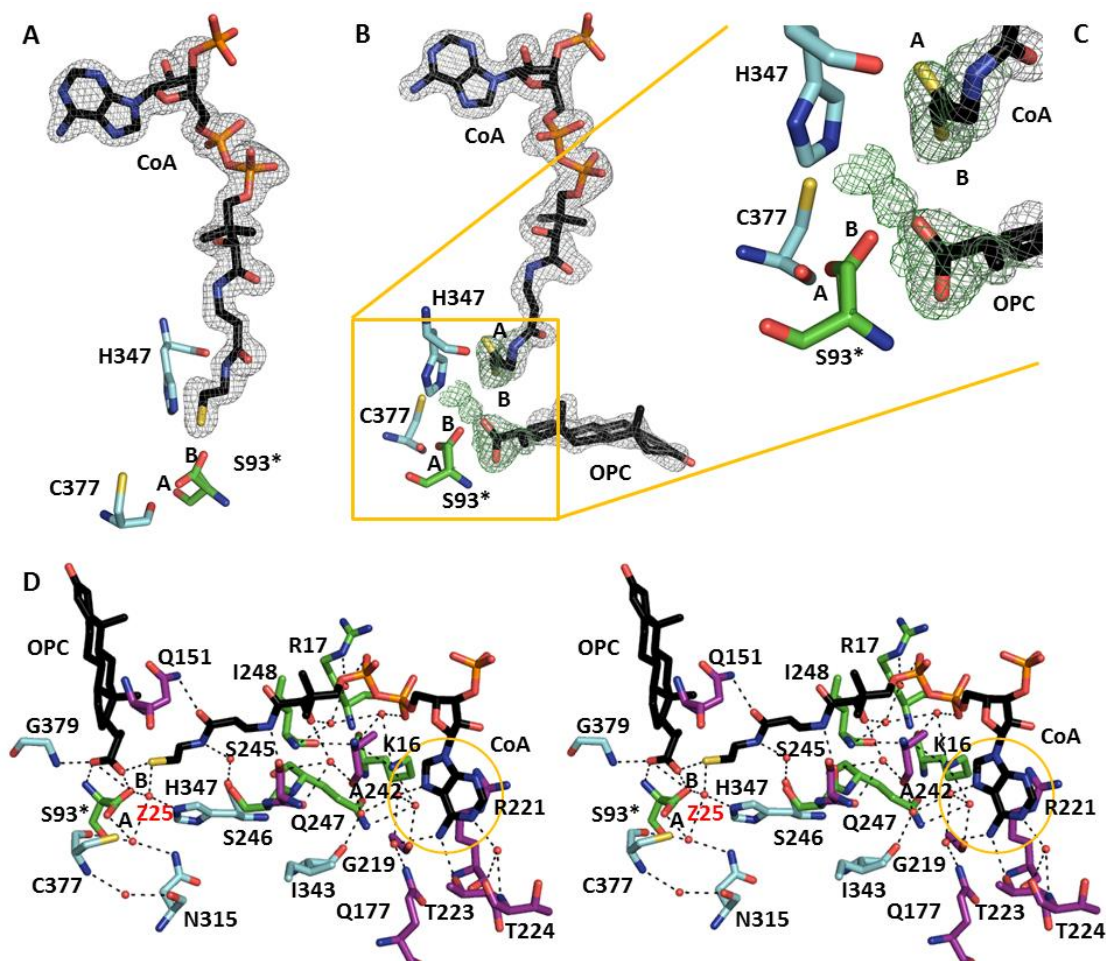


Figure 55 - Ligand binding in the OPC-complex:

A) SIGMAA-weighted $2F_o-F_c$ electron density omit map ($\sigma = 1.0$; grey) of CoA (black). The ligand is shown with the active site residues in the respective sub-domain color code. **B)** SIGMAA-weighted $2F_o-F_c$ electron density omit maps ($\sigma = 1.0$; grey) and F_o-F_c electron density omit maps ($\sigma = 3.0$; green) of the bound ligand (CoA and OPC in black) and the space around the active site residues for chain B of the enzyme. In close proximity to S93-B difference density pointing towards OPC is visible, indicating some occupancy for 3-OPC-CoA, the non-hydrolyzed product. **C)** Zoom into the catalytic triad of **B** **D)** Stereo view of the CoA (black) and FadA5 interaction via water (red spheres) mediated hydrogen bonds (dashed lines) as well as direct contacts and a π -cation-stacking interaction (orange ellipse) of the adenosine moiety of CoA with R221. The active site residues interact with each other via direct and water mediated contacts, a central position is assumed by water molecule **Z25**. The position of OPC (black) is also shown. Interacting residues are shown in stick representation, and colored in the sub-domain color code. * indicates residues with alternative conformations.

In all four protein chains (A-D) of the asymmetric unit additional electron density was assigned to a CoA ligand (Figure 55A). This was at first surprising, since the protein was co-crystallized with 3-OPC-CoA (Figure 10, compound 3), a product of the FadA5 thiolysis reaction, and not CoA alone. In chains A to C electron density remained unassigned after CoA was included but as the remaining electron

Results

density was not continuous modeling of the entire 3-OPC-CoA molecule was not successful. The positive difference electron density in close proximity to the CoA electron density was finally assigned to 3-oxo-pregn-4-ene-20-carboxylic acid (OPC) and it was assumed, that during the crystallization process the 3-OPC-CoA was hydrolysed into the free acid and the thiol. Chain D, however, is not only lacking residues 133 to 147 of sub-domain II, but here the OPC could also not be modeled.

The obtained electron density of CoA (**Figure 55A**) is continuous and much better defined than for the other above described complexes. The electron density for CoA in chains A and C clearly shows a single conformation for this ligand, whereas the electron density of the CoA moiety interacting with chains B and D is showing alternative conformations, which therefore were modeled. The flexibility of CoA in chain D is observed for larger parts of the ligand, reaching from the thiol to the C_{9P} atom. In chain B, only the thiol points into two different directions and therefore alternative conformations for the sulfur and the neighboring C_{2P} were assigned (**Figure 55B**). The angle between the two sulfur orientations is 96°, orientation B points towards the OPC molecule and superimposes with the thiol in chain A whereas orientation A points away from OPC. The distance between the thiol (SH-B) and OPC is larger, compared to chain A, where a hydrogen bond interaction is observed. The distance in chain B (3.4 Å) would still suffice for a weak interaction, but the geometry prevents this. The most intriguing feature of the active site in chain B is an extended region of not assigned difference electron density between CoA, OPC and S93 (**Figure 55C**). This electron density could be due to the presence of a very minor fraction of uncleaved 3-OPC-CoA but modeling with the original compound did not succeed, most likely due to the low occupancy. The same remaining electron density in this region is also observed for chain A but there it is less prominent.

Since the position of the OPC molecule and the CoA were very well defined in the A chain, the following analysis of the active site and ligand binding will be described with respect to this subunit. Interestingly, the active site serine 93 occupies two alternative conformations, where the oxygen atoms are pointing into opposite directions. Binding of CoA led to similar contacts with the active site residues as described for the acetyl-CoA of chain A in the acetylated WT(-acetyl)-CoA complex structure (**Figure 53B**). Three direct hydrogen bonding interactions of CoA with residues Q151, T223 and S246 and one extra water-mediated interaction to Q247 were observed. The last interaction is exclusively observed in the OPC-complex structure, whereas the first three are partially also present in the acetylated FdaA5 WT(-acetyl)-CoA structure, depending on the ligand orientation (**Figure 55C**). No interaction was observed, for the initially covalently-connected atoms of CoA and OPC. The distance between S_{1P} and C_{AQ} increased to 3.9 Å from formerly 1.8 Å.

Results

OPC is a steroid and therefore it is not surprising, that a hydrophobic binding pocket, surrounding it, could be identified. The ligand electron density for the modeled OPC and CoA molecules, in chain A to C, is continuous, (Figure 56A). The steroid-binding pocket is mainly formed by residues of sub-domain II, ranging from L128 to I139 and N150 to Q151 (Figure 56A). Not only hydrophobic amino acids contribute to the binding pocket, but polar or charged amino acids such as R136, N150 or Q151 contribute their main chain, C β or C γ atoms. In addition to the above mentioned residues, N68 from the adjacent monomer also provides its side chain to the binding pocket. Due to the hydrophobicity of OPC it hardly forms any direct or water-mediated interactions within the binding pocket or the

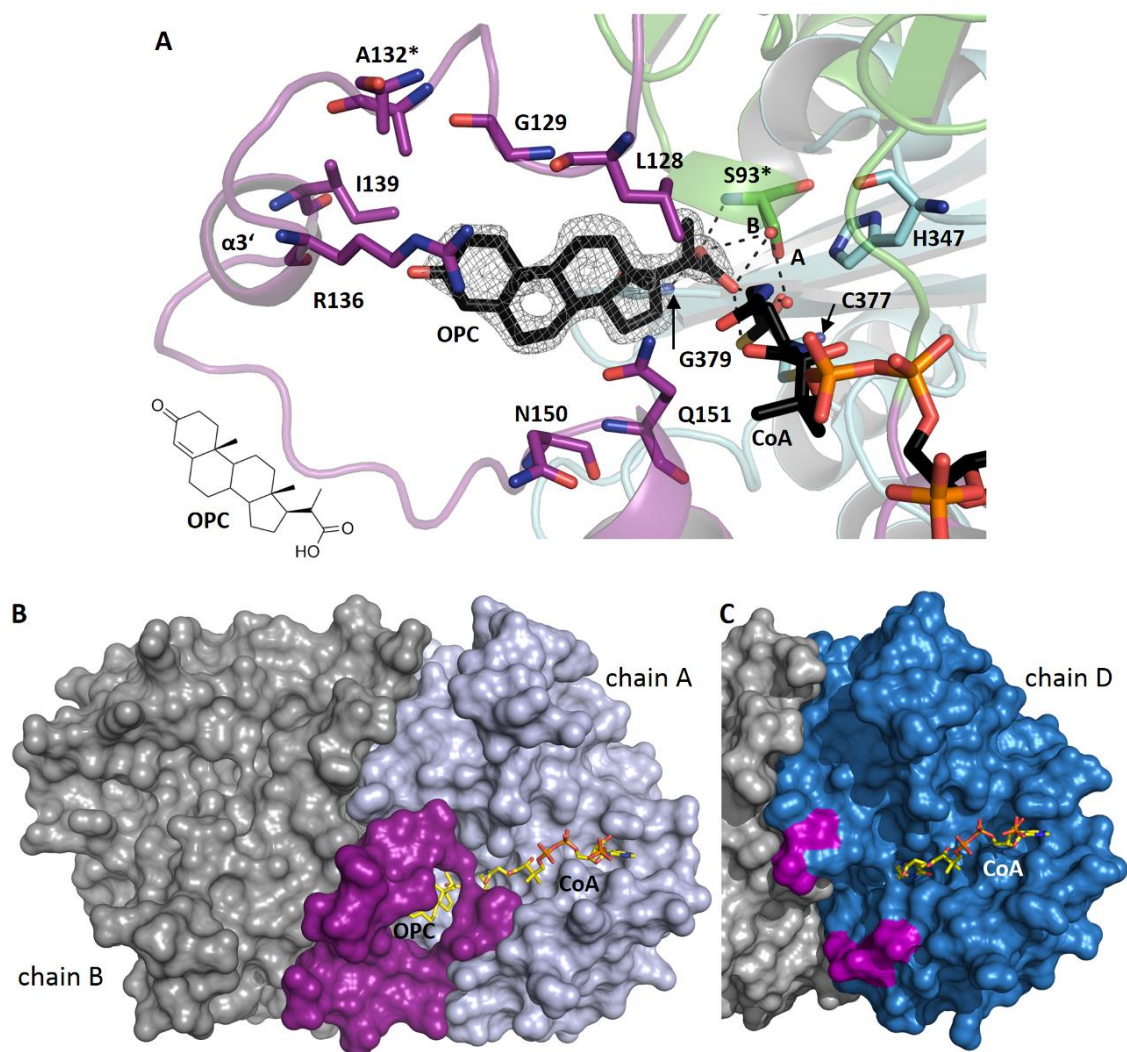


Figure 56 - Steroid-binding pocket of the OPC ligand:

A) SIGMAA-weighted $2F_o - F_c$ electron density omit map ($\sigma = 1.0$; gray) of OPC (black; chemical structure on the left). OPC is shown with respect to the residues of the lid sub-domain (purple) and active site residues. The residues of sub-domain II are forming a hydrophobic pocket to bind the ligand; the protein (chain A) is shown in cartoon representation in the respective sub-domain color code. *indicates residues with alternative conformations **B)** The surface presentation of the enzyme dimer (chain B, gray) illustrates the solvent exposed position of CoA and the occluded location of OPC (both yellow) with respect to chain A (light blue). Moving residues of sub-domain II are highlighted in purple. **C)** The surface presentation of the dimer formed by chains C (gray) and D (blue) illustrates that the residues, which are forming the steroid-binding pocket are missing and only the CoA (yellow) ligand is bound. The purple coloring indicates the beginning and the end point of the not resolved steroid-binding loop.

Results

active site. The carboxyl group directly interacts with S93(B) and G379 (Figure 55C and Figure 56A), whereas no interaction of the C3 ketone with its environment can be observed. Importantly, the steroid-binding pocket is located at the dimer interface with residues from both monomers contributing to the hydrophobic pocket (Figure 56B).

Prior to our structural analysis it was unclear how the steroid-CoA ligands enter and exit the enzyme. One option was a channeling of the complete steroid-CoA through the CoA-binding pocket, which would include passage via the catalytic triad into the steroid-binding pocket. The other option was an opening and closing mechanism of the so far not clearly identified binding-pocket to enable entry and release. A comparison of the A/B and C/D dimers (Figure 56B/C) clearly shows, that residues G133 to D147 which are not well defined in the electron density map of chain D are necessary to form a shielding loop for the steroid-binding pocket (purple surface Figure 56B). Upon closure, this loop most likely prevents the release of the OPC. The CoA ligand of chain D also displays a higher flexibility, indicated by slightly increased B-factors in comparison to the other CoA ligands and by the presence of a larger stretch of atoms in an alternative conformation. It is thus very likely that chain D of the OPC-complex represents FadA5 in motion where a region of 15 amino acids is reoriented to enable substrate entry and product release after the thiolysis reaction. This scenario is also reasonable from a different point of view, as channeling of a highly hydrophobic compound along a hydrophilic CoA-binding pocket and the active site is not very likely.

Table 35 - Data collection and refinement statistics of the FadA5 WT and C93S complex structures

	acetylated FadA5 WT(-acetyl)-CoA complex	acetylated FadA5 C93S-CoA complex	FadA5 C93S OPC-complex
Description	Acetylated thiolase (SCY93) in complex with acetyl-CoA or CoA	Acetylated thiolase (OAS93) in complex with CoA	thiolase in complex with cleaved product fragments (CoA and OPC)
Relevant buffers			
Protein buffer	20 mM Bicine-NaOH pH 8.5, 250 mM NaCl		
Crystallization condition	0.1 M MES pH 6.5 4% (v/v) 1,4-dioxan 1.8 M (NH ₄) ₂ SO ₄	0.1 M Bis-tris pH 6.5 0.2 M NaCl 25% (v/v) PEG 3350	0.1 M MES pH 6.7 23% (v/v) PEG 4000
Cryoprotectant solution	0.1 M MES pH 6.5 1.6 M (NH ₄) ₂ SO ₄ 10% (v/v) 1,4-dioxan 25% (v/v) glycerol + 10x ACO	0.1 M Bis-tris pH 6.5 0.3 M NaCl 27% (v/v) PEG 3350 25% (v/v) glycerol + 10x ACO	0.1 M MES pH 6.7 27% (v/v) PEG 4000 25% (v/v) PEG 400 + 10x 4-OPC-CoA
Data collection			
Wavelength (Å)	0.9184	0.9537	0.9184
Temperature (K)	100	100	100
Space group	P4 ₃ 2 ₁ 2	P12 ₁ 1	P12 ₁ 1
Unit cell parameters			
a/b/c (Å)	124.68 / 124.68 / 124.71	127.61 / 105.16 / 147.66	76.82 / 100.39 / 107.98
α/β/γ	90° / 90° / 90°	90° / 107° / 90°	90° / 100° / 90°
Resolution (Å)	36.0–1.95	58.8–3.0	47.5–1.7
Number of subunits	2	8	4
Total reflections	1,051,278 (64,269)	144,763 (21,071)	1,080,779 (154,493)
Unique reflections	72,009 (4,358)	70,941 (10,279)	175,728 (25,549)
Completeness (%)	100.0 (100.0)	95.0 (95.0)	99.6 (99.7)

Results

∅ Redundancy	14.6 (14.7)	2.0 (2.0)	6.0 (6.2)
$\langle I / \sigma(I) \rangle$	13.8 (2.5)	4.6 (1.3)	12.2 (2.7)
R_{merge}^a (%)	13.3 (108.0)	17.1 (63.4)	7.5 (62.3)
Wilson B-factor(\AA^2) ^b	23.7	26.5	19.3
Refinement			
Total number of atoms	6,504	23,784	13,496
$R_{\text{work}} / R_{\text{free}}$ (%)	19.3 / 23.1	24.7 / 28.7	14.5 / 17.4
RMSD			
bond angle (°)	1.3581	1.4556	1.096
bond length (\AA)	0.0114	0.0123	0.007
Estimated coordinate error (\AA) (Refmac / Phenix)	0.09	0.38	0.15
Average B factor (\AA^2) ^c	29.9	43.9	31.7
Ramachandran plot (MolProbity) ^d			
Most favored (%)	97.9	96.3	97.3
Allowed (%)	2.1	3.5	2.5
Disallowed (%)	0	0.2	0.2
PDB code	4ubv	4ubu	4ubt

Values in parentheses refer to the highest-resolution shell

$$^a R_{\text{merge}} = \sum_{hkl} \sum_i |I_i - \langle I \rangle| / \sum_{hkl} \sum_i I_i$$

^b truncate (French and Wilson, 1978); ^c baverage (Winn et al., 2011); ^d (Chen et al., 2010)

4.2.5 Comparison of the FadA5 structures

To obtain insights into the structural differences upon ligand binding the following comparisons were performed. The analysis focuses on the differences between FadA5 apo I and the complex structures, as these structures assume the same dimer whereas FadA5 apo II might be a crystallization artifact.

4.2.5.1 Active site

The RMSD calculations mentioned above already indicated a high structural similarity between WT apo I and the different complex structures. A structural superposition (Figure 57A) and cross structure calculations between the single FadA5 structures, which indicated a maximum RMSD of 0.64 \AA , between the C α atoms of the chains, shows that not only the apo structure is almost identical to the ligand bound structures but also the size of the bound ligand does not lead to large overall structural reorientations. This, however, was surprising, as the binding of a steroid was thought to lead to domain movements or to a volume change of the binding pocket. Differences among the structures are primary observed within sub-domain II. An additional helix, $\alpha 3'$ (R136 to I139), is formed between $\beta 5$ and $\alpha 4$ in the complex structures and a flexible region close to hairpins $\beta 6$ and $\beta 7$ is observed. However, the structural comparison suggests a rigidity of FadA5 which may not be in place when the enzyme undergoes catalysis. The structure of chain D in the OPC-complex provides a glimpse towards this flexibility (Figure 57C) as the steroid-binding loop is disordered and a movement of this loop may be required for steroid binding.

Results

A superposition of the FadA5 WT apo I chain A with every chain A of the complex structures, revealed that H347 and C377 hardly diverge in the different structures (Figure 57B). The observed minor movements are within the coordinate error of the models. Residue 93, on the other hand, displays a different orientation in every structure and a comparison of the coordinate error and the standard deviation between apo I and the different complex structures clearly shows that these movements are significant (Figure 57B). Importantly, residue 93 adopts non-acetylated and acetylated conformations, which is not only observed in different protein chains but even within the same monomer. For example, C93 and SCY93 were observed in alternative conformations in the acetylated WT(-acetyl)-CoA complex but none of these assume the same orientation as observed for WT C93 (Figure 57C, left). The acetylated S93 in the acetylated C93S-CoA complex does not assume alternative conformations, but also diverges from the apo orientation (Figure 57C, middle).

In the acetylated WT(-acetyl)-CoA structure, both sulfur atoms, C93 and SCY93, are shifted relative to

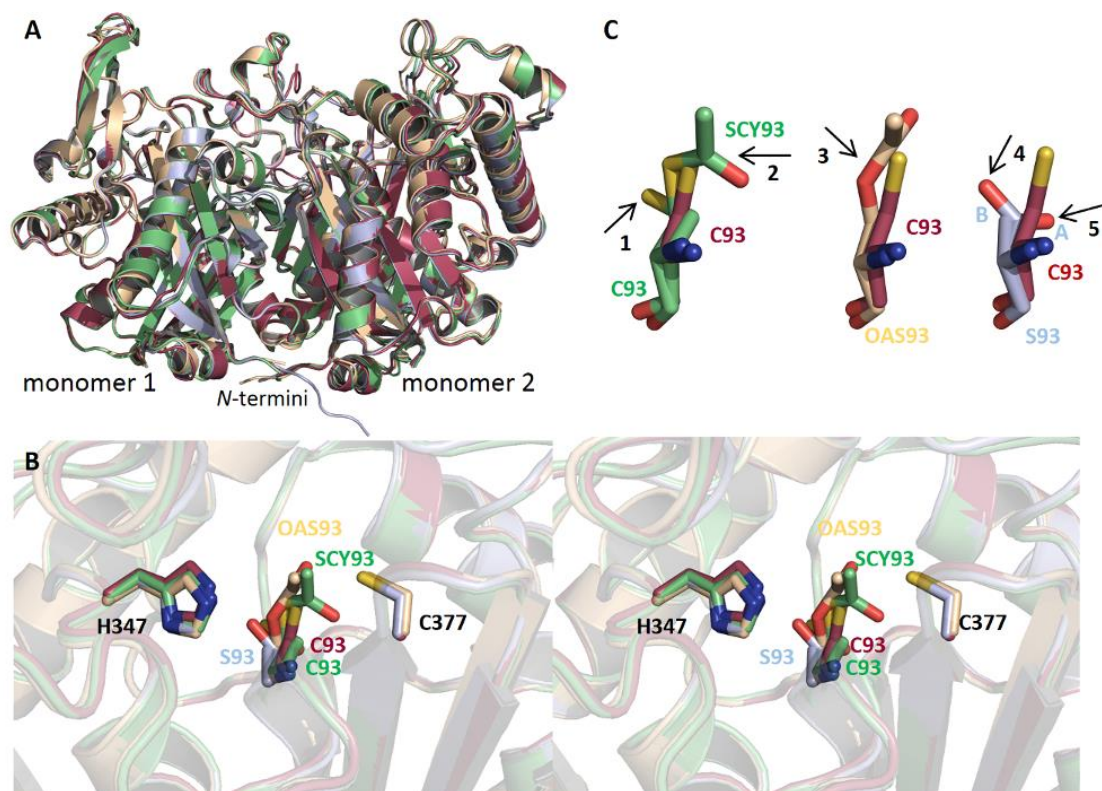


Figure 57 - Structure and active site comparison of the FadA5 structures:

A) Dimer superposition of all FadA5 WT apo I and the complex structures. The superposition in cartoon design shows a high structural similarity between the structures which is further supported by the analysis with the program PDBFold revealing low interstructure RMSD values (Krissinel and Henrick, 2004). (WT apo I, dark red; acetylated WT(-acetyl)CoA-complex, green; acetylated C93S-CoA-complex, beige; C93S-OPC-complex, light blue)

B) Detailed stereo view of an active site residue superposition, C93/S93/SCY93/OAS93, H347 and C377, indicates that only residue 93 assumes different orientations. The superimposed proteins are shown in cartoon presentation. Color code as above.

C) Detail of the superposition of C93 (WT I) with one of the residue 93 of the complex structures. Arrows 1 and 2 illustrate the movement of C93 and SCY93 respectively, in comparison to C93 (left). Arrow 3 indicates the movement of OAS93 in comparison to the WT residue (middle). Arrows 4 and 5 point to the altered conformations of S93(B/A) with respect to WT C93 (right). Color code as above.

Results

the sulfur of the apo C93. The unmodified cysteine moves by 95° in comparison to the WT C93 apo I structure and the acetylated sulfur by 25° towards the catalytic histidine. A similar, but minor conformational change is observed for the OAS93 in the acetylated C93S-CoA structure, where the oxygen is shifted by 39° in comparison to the sulfur of the apo C93. The acetylation, or more correctly acylation during catalysis of one catalytic cysteine, the nucleophile, is an important factor of the thiolase reaction. During thiolysis, the nucleophile is attacking the β -carbonyl of the β -ketoester substrate, compound 2 (**Figure 10**), by forming a tetrahedral transition state. In the following, the steroid is transferred to the cysteine and acetyl-CoA can be released (**Figure 11**, state II). This acylated enzyme state is mimicked by the two acetylated FadA5 structures, where the large steroid moiety (R) is presented by the methyl group. The movement of the oxygen and the sulfur, positions the respective acetyl-group directly towards the steroid-binding pocket and a substitution of the methyl group of the acetyl by the steroid could be readily accommodated. Similar alternative conformations of the catalytic cysteine and covalent modifications were observed for the biosynthetic thiolase from *Z. ramigera*. (Kursula et al., 2002; Modis and Wierenga, 2000)

S93 of the OPC-complex is not acetylated but present in two alternative conformations, which both differ from the apo C93 orientation (**Figure 57B/C**, right). The S93-A orientation is shifted by 79° and the S93-B by -80° relative to the sulfur of the apo C93, this enables S93-B to directly interact with OPC and to interact with CoA via the catalytic water Z25 (**Figure 55D**), which we also observed for the acetylated WT(-acetyl)-CoA complex.

Results

4.2.5.2 Ligand binding

The overall structural rigidity is maintained upon acetyl-CoA and CoA binding. The overall orientation of the CoA ligands is similar in all complex structures (Figure 58A) but the quality of the ligand electron density and the occupancies differ strongly. The best ligand density for CoA is observed in

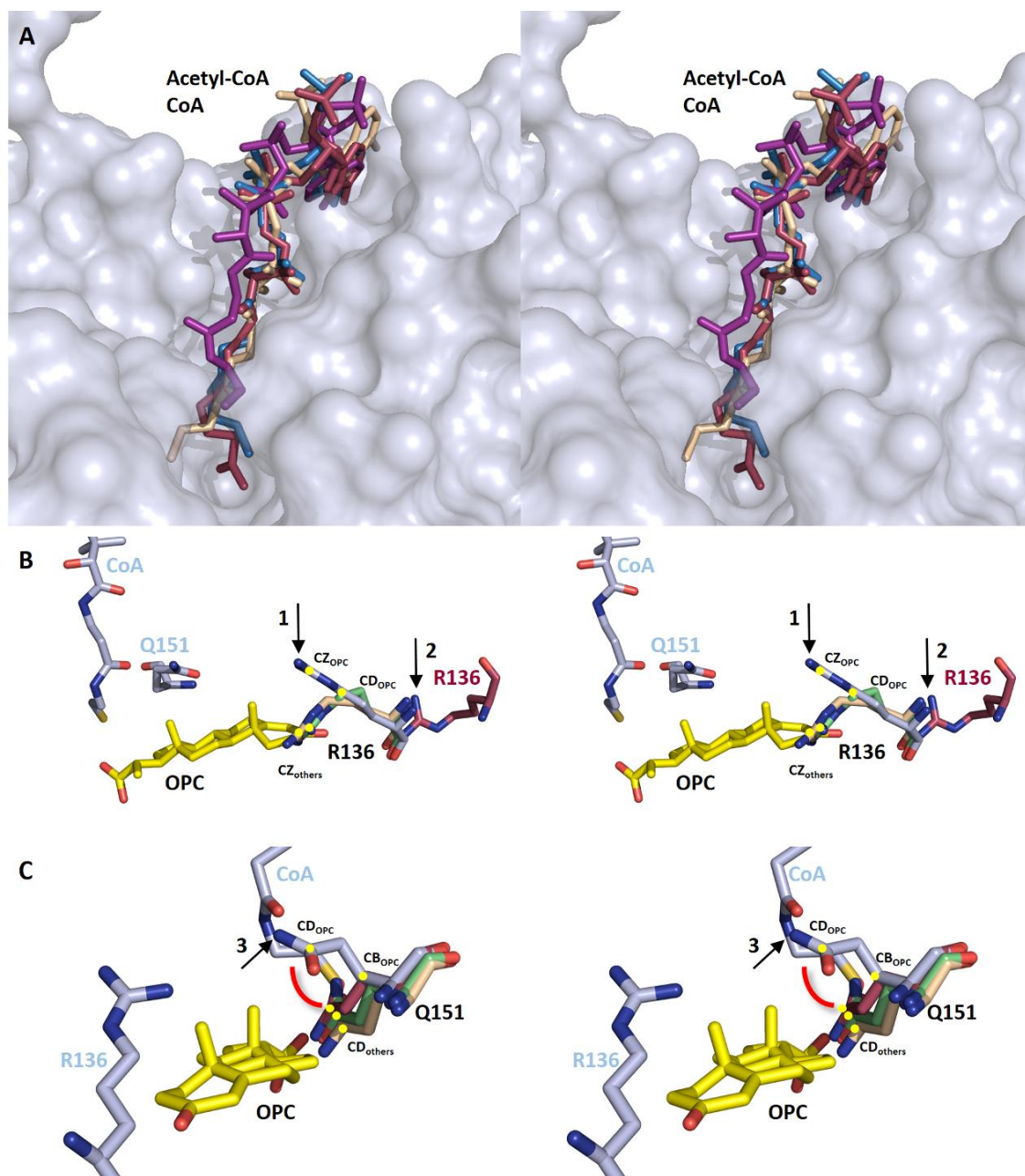


Figure 58 - Differences in ligand binding:

A) Stereo view of the superimposed acetyl-CoA and CoA binding pockets in the complex structures. Acetyl-CoA (dark red) and CoA (purple) belong to the acetylated WT(-acetyl)-CoA complex, they are both present in monomer A. CoA (beige) of the acetylated C93S-CoA complex and CoA (blue) of the OPC-complex are also shown. The protein is depicted as surface in grey. **B/C)** Detailed stereo view of the reorientations that were necessary by OPC (yellow) binding in the steroid-binding pocket. The movements of residues Q151 (light blue, arrow 1) and R136 (light blue, arrows 2/3) in comparison to the other structures are shown. The atoms used for the angle determination are labeled and indicated by yellow dots (residue color code: acetylated WT(-acetyl)-CoA complex, green; apo, dark red; acetylated C93S-CoA complex, beige; CoA, light blue).

Results

the OPC-complex (1.7 Å) and it is remarkably better compared to the structure of the acetylated FadA5 WT(-acetyl)-CoA complex which was solved at 1.95 Å resolution. This increase in ligand density quality might be due to the presence of the steroid in the OPC-complex. The electron density for CoA in chain D of this complex is, despite the absence of the steroid, still very well defined but shows a higher structural disorder indicated by a pronounced alternative conformation and slightly enhanced B-factors. The question therefore arises, if high occupancy ligand-binding is coupled to the necessity that both binding pockets are occupied. Whereas the 3'-phosphate-ADP part of the CoA moiety is for all FadA5 complex structures similarly oriented, the remaining parts of the ligand were found in different orientations and interacted differently, or not at all, with the catalytic triad and other surrounding residues.

The steroid-binding pocket is present as empty cavity in all OPC-free FadA5 structures, with exception of chain A in FadA5 WT apo I. Although the enzyme obviously always provides this cavity, the question arises, if any residue movements are necessary to enable OPC-binding. The superposition of selected chains in the four FadA5 structures show that the residues arginine 136 and glutamine 151 followed an alternative conformation in the OPC-complex, caused by displacement due to the steroid-binding (**Figure 58B/C**). The guanidiny-group of R136 (light blue) in the OPC bound structure shows a shift of 84° to 90°, when compared to the other complex structures and so the arginine is in a more extended orientation (**Figure 58B** arrow 1). The bend angle is defined as $\sphericalangle(CZ_{OPC}, CD_{OPC}, CZ_{others})$ (**Figure 58B**, yellow dots).

Apo structure I (dark red) differs in the position of R136 for both polypeptide chains, this residue is located approx. 6 Å (chain A, **Figure 58B** arrow 2) away from the steroid binding site with respect to R136 in the OPC-complex. This large difference is most likely caused by the disorder after amino acid S137, where residues up to S143 are not resolved.

Whereas R136 assumes an extended conformation in the OPC-complex structure, a bend of Q151 is observed (**Figure 58B** arrow 3). The bend angle between the glutamine in the steroid bound structure and the extended glutamine orientations of the other structures lies between 90° and 110°. An interesting case is here Q151 in chain D of the OPC-complex structure. Although the OPC molecule could not be modeled, this residue adopts two alternative conformations, one in the extended and one in the bent conformation, thus representing an OPC bound and an OPC free residue state within the same subunit.

The number of resolved water molecules is extremely different in all FadA5 structures, most likely due to the resolution differences of the structures which range from 1.7 Å to 3.0 Å. Nevertheless, a

Results

thiolase-specific water molecule can be identified, located near the catalytic triad of the OPC-complex (Z25) and the acetylated FadA5 WT(-acetyl)-CoA complex (Z253) (**Figure 55D**, **Figure 53B**).

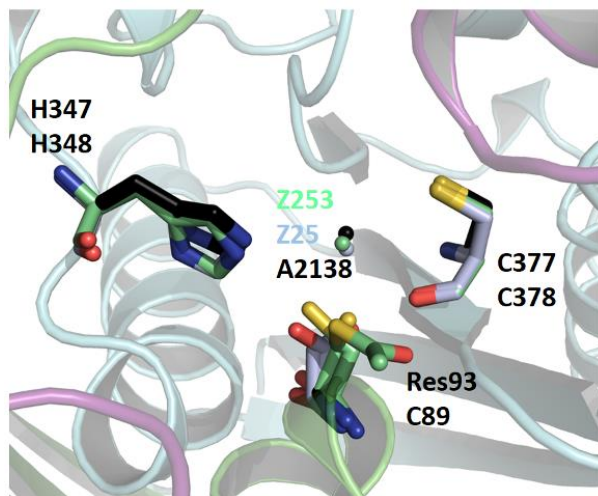


Figure 59 - Location of the thiolase specific water: Superposition of the biosynthetic thiolase (black) from *Z. ramigera* with the catalytic water containing structures of FadA5 (acetylated WT(-acetyl)-CoA, green; OPC-complex, light blue).

This water molecule was described in other thiolase structures and it seems to be highly conserved and of importance especially for the biosynthetic enzymes. (Haapalainen et al., 2006; Kursula et al., 2002; Kursula et al., 2005; Modis and Wierenga, 2000) An active site superposition of the two FadA5 complex structures (the OPC-complex and the acetylated FadA5 WT(-acetyl)-CoA complex) with a biosynthetic thiolase from *Z. ramigera* confirms the conserved location of this water molecule (**Figure 59**). In comparison to randomly bound surface water molecules, two important water

hot spots were described for the above mentioned biosynthetic thiolase. (Kursula et al., 2002) On the one hand, a water channel from the catalytic residues to the bulk solvent water was assigned and in addition a water reservoir near the catalytic triad was identified. Although the structural fold of biosynthetic and degradative enzymes is highly conserved, only the described water reservoir is present in the FadA5 structures, whereas the channel cannot be formed due to the blocking amino acids I314 and V322. FadA5 is in this respect more closely related to the degradative thiolase from *Saccharomyces cerevisiae*, where also only the reservoir is. (Mathieu et al., 1997)

4.2.6 Modeling and comparison with human thiolases

The probably most important reason for the in depth structural characterization of FadA5 is the question about the druggability of this thiolase. In the previous chapters it was shown, that although FadA5 is by sequence not very similar to thiolases of other organisms, the structural similarity in this enzyme class might prove it difficult to specifically target FadA5.

Structural analysis of the FadA5 structures and the comparison with the biosynthetic thiolase from *Z. ramigera* shows that the active site and the CoA-binding pocket are conserved, which renders them unlikely areas for the development of FadA5 inhibitors. Whereas CoA binding is a general feature, the nature of the substrate differs among the thiolases and thus an inhibitor targeting the steroid binding

Results

pocket of FadA5 could confer specificity if it differs significantly in comparison to the human thiolases.

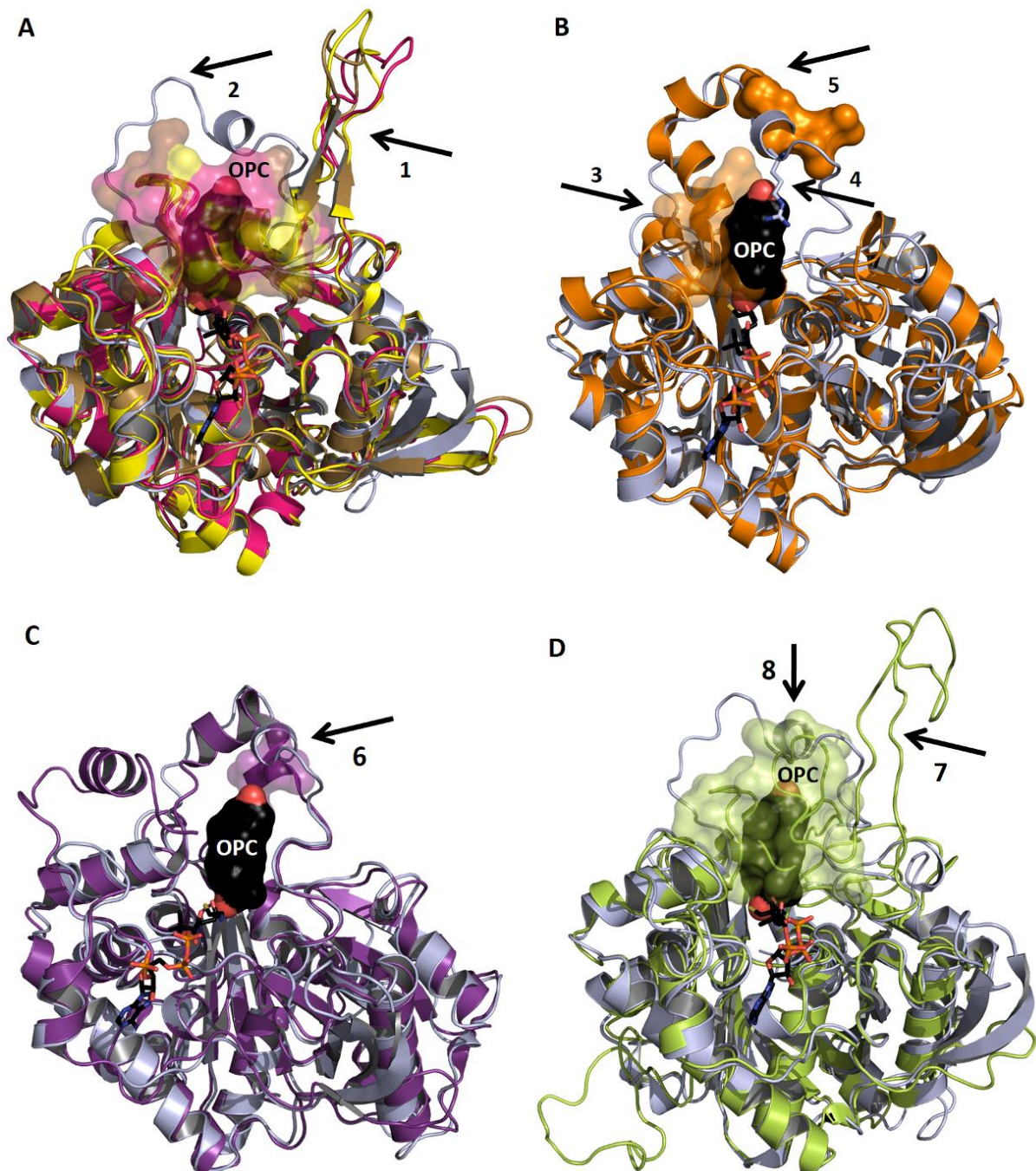


Figure 60 - Comparison of the FadA5 OPC-complex with (modeled) human thiolases

Superposition of the FadA5 OPC-complex monomer (light blue) with models or structures of different human thiolases and with the MtbTFE. Amino acids clashing with or in close proximity to the OPC (black/red surface) are shown in transparent surface presentation. **A)** Superposition with the model of the T1-thiolase (pink) and structures of CT- (1wl4, yellow) and T2-thiolases (2ib7, brown). Arrow 1 points to the tetramerization motif and arrow 2 to the steroid-binding pocket. **B)** Superposition of FadA5 with the AB-thiolase (2iik, orange) structure. Arrow 3 indicates the clashing region and arrow 5 a stretch of amino acids that cannot shelter OPC as observed for R136 (arrow 4) in FadA5. **C)** FadA5 superimposed with the MtbTFE (purple), which is also forming a steroid-binding pocket (arrow 6). **D)** Superposition with the model of the human TFE thiolase (green) that also forms a tetramerization motif (arrow 7) and seriously clashes with OPC (arrow 8).

Results

Six different thiolases have been identified in humans, but structural information is available for only three of them. To compare not only the available structures with the steroid binding pocket in FadA5, modeling experiments with the help of the I-TASSER server (Roy et al., 2010) were performed to generate models for the so far structurally uncharacterized thiolases. Modeling was pursued by two different strategies, either by defining an available structure as primary starting point for the modeling process or without template definition and modeling against the complete PDB. In the first approach the template structure and the Protein Data Bank (PDB) have to be included. In the other strategy no template structures were defined, the search process was completely automated and based on the complete PDB. The template approach was performed with the FadA5 OPC-complex structure, or one of the available human structures, 2iik (human peroxisomal AB-thiolase (AB)), or 1wl4 (human cytosolic thiolase (CT)), as the predominant template. Although a third human thiolase was characterized structurally, the mitochondrial T2-thiolase (T2), it was not used as an additional search template, as it is structurally nearly identical with the human CT-thiolase. Modeling was performed using the amino acid sequences of the so far structurally un-characterized human thiolases, namely the mitochondrial trifunctional enzyme thiolase (TFE), the mitochondrial T1-thiolase (T1) and the peroxisomal SCP2-thiolase (SCP2). A special focus was set on the TFE thiolase, as this thiolase was recently described in a phylogenetic relationship study to be the human thiolase, which is the closest relative to FadA5 (Anbazhagan et al., 2014). As described in the introduction, FadA5 is classified to belong to the trifunctional enzyme like thiolases which are highly related to the human TFE (1.3.4.3). The *Mycobacterium tuberculosis* trifunctional enzyme was also structurally characterized and therefore included as a template for modeling studies of the human TFE. (Venkatesan and Wierenga, 2013) Every modeling process with the I-TASSER server normally releases five models per search sequence, which resulted in 20 models for the human T1 and SCP2 enzymes and 25 models for the TFE protein. Modeling of the T1-thiolase led to only five of the 20 expected models. This is caused by the high structural similarity to the CT and the T2 enzyme and is clearly seen in the superposition of the five models with these two enzymes. Interestingly, the high structural similarity is not related to a high sequence identity, which is only between 40 and 43%. **Figure 60A** shows the superposition of one model structure of the T1-thiolase with CT, T2 and the FadA5 OPC-complex. Although the T1 enzyme was characterized to be a degradative thiolase, it belongs most likely to the thiolases that form a tetramerization loop (**Figure 60A**, arrow 1). Thiolases that form the tetramerization motif are most likely not able to form a steroid-binding pocket (arrow 2), since all of these human thiolases contain a stretch of seven amino acids emerging from the tetramerization loop which cut through or clash with the theoretical position of the steroid-binding pocket. In addition, it is known, that these thiolases have a preference for (branched) short chain and unbranched long chain substrates, which would not require the presence of such a large steroid-

Results

binding pocket (Dedkova et al., 2013; Liang and Schulz, 1998; Pokhrel et al., 2012) leading to the conclusion, that steroid-like FadA5 inhibitors would not inhibit these enzymes.

The SCP2-thiolases are sub-classified into type 1 and 2, but for neither type structural information was available. These thiolases are more than 150 amino acids longer than the other human enzymes or FadA5, which complicated modeling approaches and yielded 20 models displaying nearly 20 different orientations and structural elements for this additional stretch of residues. FadA5 is not very closely related to the SCP2 thiolases. Other *M. tuberculosis* or *M. smegmatis* enzymes have been identified to belong into one of the SCP2 classes but they haven't been structurally characterized so far. (Anbazhagan et al., 2014) 75% of the generated models show clashes with OPC or are not forming a steroid-binding pocket which was surprising since the SCP2-thiolases are degradative enzymes, which utilize bile acid synthesis intermediates or long chain branched or unbranched 3-ketoacyl-CoA compounds as substrates. These enzymes therefore clearly have to accommodate large substrates and the respective binding pockets might possibly also bind steroid-like inhibitors targeting FadA5. The fold of the additional 150 amino acids remains completely unclear and further complicates a final conclusion but these additional residues might also provide a different fold of the ligand binding pocket than observed for FadA5. This might facilitate specific targeting of FadA5. Due to these ambiguities, no superposition is shown here.

The AB-thiolases, which are degradative enzymes and located in the peroxisomes are also difficult to evaluate. The deposited structures are lacking a stretch of five amino acids, which are not resolved, but would be in close proximity to the steroid-binding pocket. Residue Q151, that is reoriented in the OPC-FadA5 complex structure, is replaced in the AB-thiolase by a methionine (M181), which would be clashing with the OPC in its current position. In addition, neighboring clashing regions (**Figure 60B**, arrow 3) can be identified, whereas residues (FadA5 R136; **Figure 60B**, arrow 4) responsible for shielding OPC from the solvent are smaller and may lead to a more solvent accessible pocket (**Figure 60B**, arrow 5). The overall orientation of the residues that could probably contribute to the formation of a steroid-binding pocket, would have to be altered significantly in comparison to the current structure and it is currently unclear if a similar binding pocket as observed in FadA5 can be formed. It cannot be excluded, that FadA5 specific inhibitors would also target the AB enzymes, but to accommodate a steroid-based inhibitor, similar to OPC, extensive reorientations of this enzyme seem to be necessary.

Modeling of the last human enzyme, the degradative TFE thiolase, included the *M. tuberculosis* TFE, which also forms a pocket that probably is capable of harboring a steroid (**Figure 60C**, arrow 6). Modeling yielded 25 different model structures out of which 80% contain a tetramerization loop (**Figure 60D**, arrow 7) but no steroid-binding pocket. These models follow a very similar fold as the

Results

structurally characterized T2- and CT-thiolases, especially with respect to the tetramerization loop and the thereby occurring clashes with the OPC. In all these models, a varying number of amino acids are clashing with the OPC molecule from the OPC-complex structure (Figure 60D, arrow 8). The remaining 20% of the models assume a similar fold as the AB-thiolase, for which a distinct statement is difficult to provide. The TFE thiolases also accommodate long chain branched or unbranched 3-ketoacyl-CoA compounds which require a larger binding pocket compared to the OPC binding pocket. Inhibition of the TFE's with FadA5 inhibitors cannot be excluded, though the structure prediction hints to the formation of a tetramerization loop instead of a steroid-binding pocket. This might allow the application of FadA5-tailored inhibitors without harming this human enzyme.

5 Conclusion and perspectives

5.1 Mrr1 – a transcription factor from *C. albicans*

The aims of this project were the successful recombinant expression of a soluble homogeneous Mrr1 construct that is suitable for biochemical characterization and crystallographic studies. The biophysical experiments should indicate proper folding and homogeneity of the sample and ideally reveal differences in functional assays between the wild type and the gain of function mutant constructs. Structural characterization of the wild type protein and the variants would then have pinpointed to the specific role of these amino acids towards their influence in terms of gain of function i.e. hyperactivity of the transcription factor.

5.1.1 Mrr1 constructs – expression and purification

The initial steps of this project were cloning, successful expression and purification trials. For this purpose, three different Mrr1 constructs were cloned into different vectors and were subsequently tested for proper expression and purification properties.

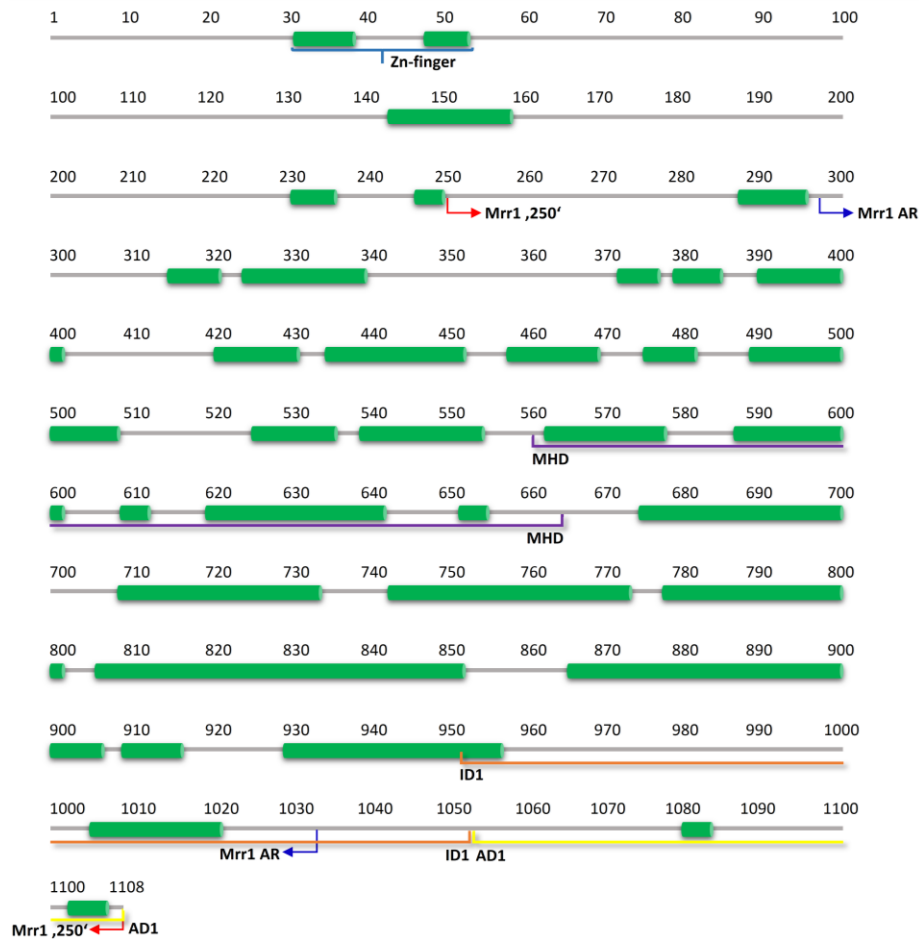
Prior to choosing the sequences for cloning, secondary structure predictions with the complete Mrr1 sequence were performed. An exemplary secondary structure prediction performed with the Phyre² server (Kelley and Sternberg, 2009) is shown in [Figure 61A](#), but results from all performed predictions contributed to the amino acid sequence selection to generate the truncation variants. A comparison of the results from the shown secondary structure prediction with the dissection experiments performed by Sabrina Schubert (Schubert et al., 2011b) some interesting suggestions or explanations for the obtained results might be provided. Residues 951 to 1050 of Mrr1 were assigned to the inhibitory domain 1 (ID1) ([Figure 61A](#)). The random deletion of amino acids 951 to 1000, which leads to hyperactivity of Mrr1, excises the last two turns of a predicted 30 amino acids long α -helix together with a stretch of largely disordered amino acids. This removal fuses the remainder of the 30 amino acids long helix nearly directly to another helix consisting of 18 amino acids. One could thus speculate that the excision of the disordered region leads to a more compact C-terminus and thereby constitutive activity of the transcription factor. The deletion of amino acids 1001 to 1050 would remove the 18 amino acids long α -helix completely and additional 30, predicted to be disordered, amino acids. In this way a long stretch of disordered amino acids would be removed, but the overall structure would be less affected since the long α -helix would stay intact. As this construct still appeared hyperactive it can be hypothesized that the disordered region may have a decisive role in activation. Another explanation might be provided by taking a closer look at the very C-terminal 58

Conclusion and perspectives

amino acids (Figure 61A). They represent the activation domain 1 and are also predicted to be largely disordered with the exception of two small very C-terminal α -helices. One could thus also speculate that the deletion of amino acids 1001 to 1050 positions these C-terminal α -helices in close proximity to the armadillo repeat α -helices and thereby activation is achieved.

The Armadillo repeat construct containing amino acids T298 to T1033 could not be expressed in *E.*

A



B

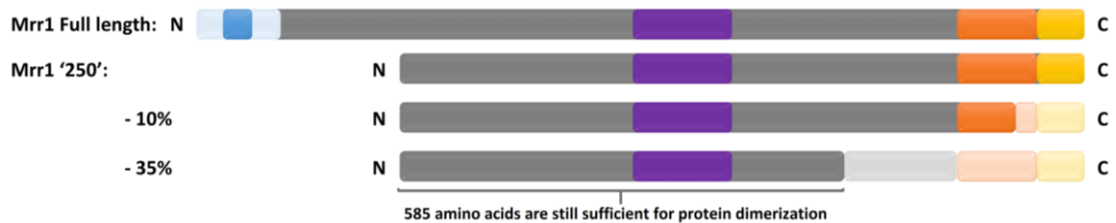


Figure 61 - Schematic illustration of secondary structure motifs, regulating domains and the degree of degradation of Mrr1:

A: Secondary structure prediction of Mrr1 Full length based on the Phyre² server (Kelley and Sternberg, 2009). The amino acids are indicated by the numbering above the secondary structure elements. The helices are shown as green cylinders. The truncation variants of Mrr1 are indicated by the blue (Armadillo repeat) or red ('250') arrows. Zn-finger domain (blue), middle homology region (MHD, purple), inhibitory domain 1 (ID1, orange) and activation domain 1 (AD1, yellow) are illustrated by brackets. **B:** The domain structure of Mrr1 Full length and the Mrr1 '250' variant is shown. 10% or respectively 35% C-terminal degradation are illustrated. The color code is as above and degraded regions are shown in transparent color.

Conclusion and perspectives

coli or at least not in amounts, which would be necessary for crystallization experiments. A reason might be that soluble recombinant protein expression, with proper folding, is only successful when the C-terminus, or at least parts of it, are present in the construct. It is very likely, that *E. coli* was not able to fold the AR protein construct properly, and therefore it was not solubly expressed. However, experiments in *C. albicans* showed that a C-terminal truncation and variants thereof can be expressed after induction with benomyl which subsequently lead to promoter activity at Mdr1. A deletion of amino acids 951 to 1000 or 1001 to 1050 even led to constitutively active protein as the inhibitory domain was disturbed by these removals. (Schubert et al., 2011b)

Mrr1 Full length and the truncated '250' construct, containing amino acids S250 to N1108 with an N- or C-terminal His₆-tag, could be recombinantly expressed in *E. coli* and for both variants purification protocols were established. The Mrr1 FL construct was purified in a two-step process, applying metal-affinity and size-exclusion chromatography. This construct could be obtained in reasonable purity but degradation and a high potential for disordered oligomerization, i.e. aggregation could not be overcome. The '250' construct was purified via a three-step protocol, including a TEV cleavage step to remove the N-terminal His₆-tag. First affinity chromatography with Ni-NTA resin was performed, then the protein was applied to, anion-exchange chromatography, followed by the TEV cleavage, and the purification was completed by a final size-exclusion chromatography. Four selected gain of function mutations were introduced into the '250' (N-terminal His₆-tag) construct and were expressed and purified as described for the WT '250' protein. The different Mrr1 '250' constructs, similar to the FL protein, showed degradation to various degrees, but in contrast to FL Mrr1 no aggregation was observed.

Since the FL protein could only be expressed and purified as an aggregate and the AR construct couldn't at all be expressed in *E. coli*, a eukaryotic expression host was chosen to increase the likelihood of obtaining soluble and functional protein. Unfortunately, recombinant expression of Mrr1 in *P. pastoris* was unsuccessful with all of the cloned constructs in any of the available strains. The reasons for this lack of success are not entirely clear. It is possible that the obtained colonies were not comprehensively expressed and analyzed, but several expression tests with different integrants in different strains were performed that did not yield a trace of the desired protein. A likely cause for the problems with the expression of the Mrr1 protein might be the fact that a synthetic gene construct that was tailored for gene expression in *E. coli* was used as cloning template for the *pichia* constructs. The *E. coli* optimization of the genomic code might prevent a successful recombinant protein expression in yeast. This hypothesis could be tested in the future optimization of the Mrr1 gene for expression in yeast.

Conclusion and perspectives

5.1.2 Biophysical characterization and crystallization

The protein constructs of Mrr1, which could be successfully expressed and purified, were analyzed by circular dichroism (CD) spectroscopy, dynamic (DLS) as well as multiangle laser light scattering (MALLS), to determine the oligomeric state, homogeneity and folding characteristics.

The FL protein was characterized by DLS clearly showing that only a minor fraction of the protein forms monomeric or dimeric complexes and the majority is present in higher oligomeric unordered clusters of at least 15 or even more subunits. DLS thus confirmed that the FL protein could not be expressed in a reasonable oligomeric state (likely a dimer (MacPherson et al., 2006)), utilizing *E. coli* as a host and with the described purification protocol.

CD experiments were performed with all Mrr1 '250' constructs, with the exception of the G997V variant, which was not available in sufficient and pure amounts. The CD spectra clearly showed that the proteins containing the *N*-terminal His₆-tag assume a different fold, compared to the proteins where the tag was removed. Neither removal nor retaining of the tag was necessarily detrimental. The protein construct containing a *C*-terminal tag displayed a similar spectrum as the protein samples for which the His₆-tag was removed. All CD-spectra of the Mrr1 constructs indicated a predominant α -helical fold but displayed different CD minima, the cause for this difference is not known. Although the CD spectra of the Mrr1 '250' proteins with a removed *N*-terminal tag or with a *C*-terminal tag followed more the ideal α -helical fold, there are also proteins described in the literature, which display CD-spectra that are very comparable to the one observed for the *N*-terminal tagged Mrr1. These proteins are also predominantly α -helical and certainly not disordered since they were successfully crystallized (e.g. PDB entry 1mbn). (Dodero et al., 2011)

SEC-MALLS experiments provided important information for all Mrr1 '250' variants. Every '250' variant with *N*- or *C*-terminal or without His₆-tag was analyzed with the help of this technique. The degradation problems were confirmed by the experiments. Every refractogram indicated at least some degradation but the wild type protein with the *N*-terminal tag or with the cleaved tag so far seemed to be the most stable variant, with hardly any degradation. The WT construct with the *C*-terminal tag displayed strong degradation, which was already apparent during the purification process. This construct was intended as an alternative to avoid the observed degradation, which was shown to originate from the *C*-terminus, but the heavy degradation in combination with the low yields renders this construct as an unlikely alternative for the *N*-terminally tagged version. All gain of function variants show different rates of degradation from only 3% (P683S) up to 35% (G878E) but as already mentioned in the results section, the degradation of the individual variants varied throughout the different purifications. Furthermore, the degradation behavior does not differ between analyzed tagged or untagged proteins. Besides the apparent degradation processes, it was

Conclusion and perspectives

remarkable that the protein was present in a predominant dimeric state. At least one part of the dimerization motif is located in the very *N*-terminal DNA-binding domain (DBD), including residues 1 to 106. As this domain is not present in the '250' constructs it must be assumed that the remainder of the polypeptide also contributes to dimer formation. This is not very surprising, as a stable dimer interface of such a large protein is very unlikely to be formed by a single 17 amino acids long helix.

The western blot experiments suggested that the degradation process is occurring *C*-terminally. This information together with the knowledge that degradation of up to 10% took place would mean that the protein is still able to dimerize when the complete activation domain 1 and parts of the inhibitory domain are missing (**Figure 61B**). If MALLS experiment, with the G878E variant where 35% degradation was suggested, is not an artifact, even one third of the protein could be missing without affecting dimer formation (**Figure 61B**). It is thus very likely that AD1 and ID1 do not play a role in dimer formation, and dimerization is probably also mediated by parts of the Armadillo repeat motif.

Crystallization experiments were performed with the Mrr1 '250' variants only. Initial experiments with protein containing the tag did not lead to any successful crystallization. The cleavage of the His₆-tag finally led to some crystals, which were very small and difficult to reproduce. Diffraction experiments confirmed that they were protein crystals but improvement of the crystal size or the resolution was so far not successful. The crystallization experiments therefore seem to confirm that the *N*-terminal tag somehow has a detrimental, hindering effect, at least for the crystallization process, thus supporting the first assumption from the results obtained by CD spectroscopy that the constructs containing an *N*-terminal His₆-tag may lead to an at least partially misfolded protein. The main goal of this work, the structural characterization and comparison of the WT protein versus gain of function variants was so far not successful but a first step towards the characterization of this intriguing protein has been achieved.

Interestingly recombinantly expressed wild type protein appears to be more stable than the chosen gain of function variants. A study has shown, that *C. albicans* strains containing homozygous P683S Mrr1 variants, showed clear fitness deficits in the absence of the drug selection pressure. (Sasse et al., 2012) One could thus speculate, that the observed instability of the recombinantly expressed gain of function variants is not an artifact of the expression system but is valid *in vivo* as well and thus leads to the reported fitness deficit if the selection pressure is removed.

Conclusion and perspectives

5.1.3 Modeling

Due to the absence of an Mrr1 structure, modeling was performed with the Mrr1 sequences of the full length protein and the two truncation variants. The I-TASSER server (Roy et al., 2010; Zhang, 2008) was used and no templates were assigned, since no structural information of similar transcription factors is available with the exception of a small part of the DNA-binding domain (e.g. (Marmorstein et al., 1992)). Five models per construct were obtained and all models of the truncated versions, '250' and AR, are nearly strictly α -helical and hardly differ from each other. Interestingly, also the predicted to be disordered C-terminus is mainly modeled to form an α -helical structure and thereby seems to be a continuation of the Armadillo repeat motif.

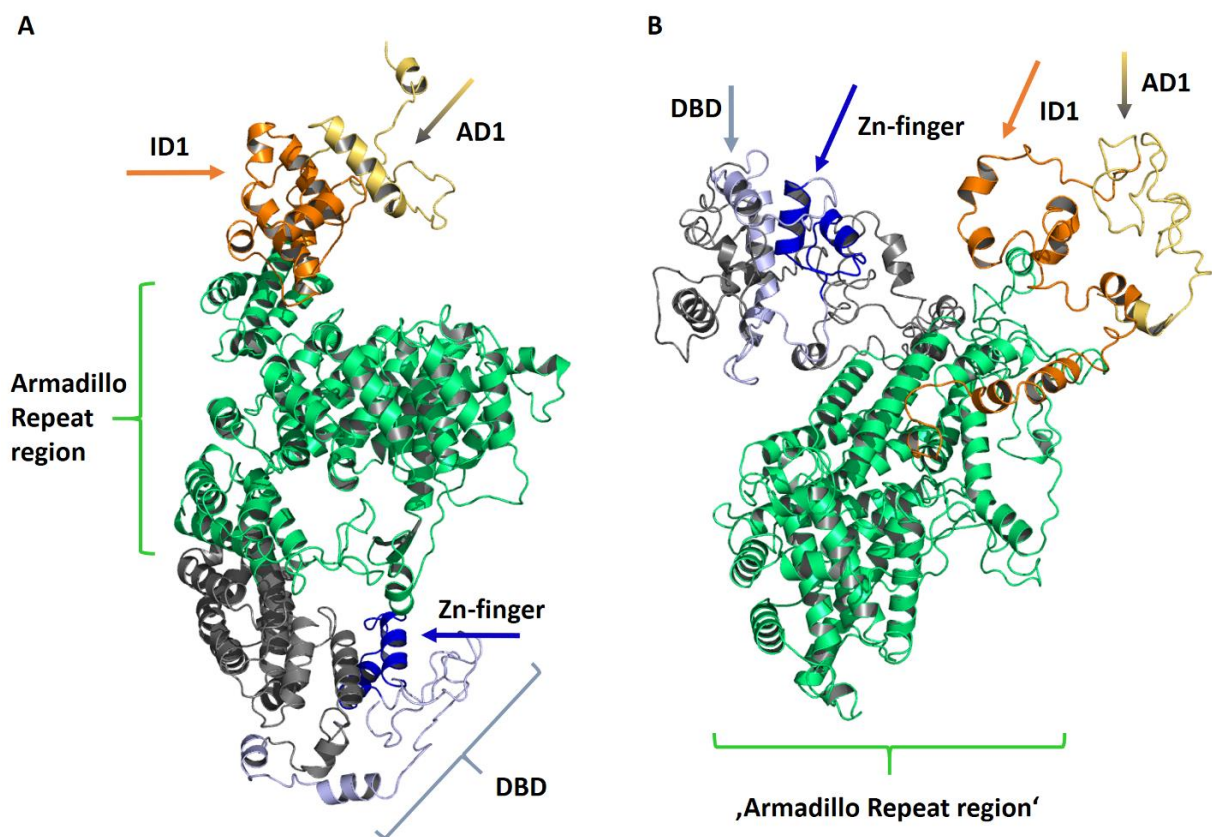


Figure 62 - Models of the Mrr1 full length protein:

Both models comprise the DBD (light blue) containing the Zn-finger (blue), a so far unassigned gray region, the (pseudo) Armadillo repeat motif (green) and the C-terminal inhibitory domain (ID1, orange) and activation domain (AD1, yellow). **A)** The model is elongated and consists of a central α -helical barrel (green bracket) with the N- and C-termini at opposite sides to each other. **B)** The second model comprises a globular helical region (green bracket) with the N- and C-termini being oriented on the same side. The potential dimerization helix is indicated by arrows 1. Bracket 2 marks the first part of ID1, which is connecting the pseudo Armadillo repeat to the C-terminal part of ID1 (arrow 3) and the activation domain. The models were generated with I-TASSER

Modeling of FL Mrr1 resulted in one model that sticks out due to its completely differing β -sheet dominated structure. This model was rejected, as it contradicted every secondary structure prediction. The template for this model was the structure of the human α 2-macroglobulin, which

Conclusion and perspectives

forms a cage like structure built by β -sheets. (Marrero et al., 2012) The other FL protein models were mainly dominated by the α -helical motif, however, two versions can be differentiated.

One elongated model can be described (**Figure 62A**), where the α -helices assume an elongated Armadillo repeat. In this model the α -helices (**green**) are forming a large helix of helices or a kind of barrel (**green** bracket). The *N*-terminus is quite disordered, but arranged at one end of the repeat indicated by the light blue and blue residues. Residues between the DBD and the beginning of the predicted repeat, S107 to Y297, are also forming α -helices (dark gray). The predicted to be disordered *C*-terminal region, including ID1 (**orange**) was modeled comparably ordered with α -helices, and therefore only the very *C*-terminal activation domain 1 (yellow) contains disordered regions.

The other model is far less elongated (**Figure 62B**) and the Armadillo repeat region (**green**) is more globular and is therefore strictly spoken not forming an Armadillo repeat. The *N*-terminus is a well separated extension from the globular core and contains more α -helices than predicted in the secondary structure analysis, but comprises more loops than described for the above mentioned model. The Zn-finger (**blue**) can be nicely identified (**Figure 62A**) and appears similar to the one described for Gal4 in the introduction (1.2.6). A helix spanning residues V82 to K105 might belong to the dimerization motif and was not modeled as an α -helix in the other model (**Figure 62B**, arrows 1). The *C*-terminus, orange and yellow, is also less well ordered than compared to the first model. ID1 can be subdivided into a linker (**orange** bracket 2) that is part of the Armadillo repeat motif but reaching from there to the more exposed second part of ID1 (**Figure 62B**, arrow 3) and the following activation domain. The second half of the inhibitory domain is partially comprised of α -helices, but is extending into a disordered region finally leading to disordered AD1 region (yellow).

The missing two models look very similar to each other and are also similar to the model depicted and described by **Figure 62A**. These two models shall not be discussed in detail, as both do not seem to be very realistic, when compared to the secondary structure prediction. In both models the Zn-binding α -helices are succeeded by 50 to 70 disordered amino acids and the remaining, nearly 1000, amino acids are again shown as α -helices. Especially this pronounced Armadillo repeat is not very likely and in contrast to every secondary structure prediction, where large disordered protein regions are predicted (**Figure 61A**).

Unfortunately it is not possible to draw any conclusions based on these models since they were modeled with structures that are barely identical to Mrr1 or are modeled without reasonable sequence coverage. All models differ from each other to a certain extent and disagree with the secondary structure predictions, especially with respect to the predicted disordered protein domains.

Conclusion and perspectives

In addition, the locations of the gain of function mutations differ in their position within the different models and thus also differ with respect to their potential assigned interaction partners. Therefore it was not possible to reliably describe an effect of these mutations in view of the overall structure of the protein.

5.1.4 Perspective

To move this project forward and in the end eventually succeed with the structural characterization of the *C. albicans* transcription factor Mrr1 two different strategies should be pursued. On the one hand new constructs for recombinant expression in *E. coli* could be designed and analyzed for their expression, purification and crystallization behavior. In this context, constructs containing only the DBD of Mrr1, could be expressed and purified and subsequently used for DNA-binding studies. DNA-binding studies with distinct stretches of the known binding motifs could then be pursued to identify a small DNA fragment with a length of 15 to 20 DNA bases that could be applied to co-crystallization experiments with the DBD. However this would be an additional project, since the main goal was and should be the structural characterization and comparison of the wild type protein and the gain of function variants that lead to resistance.

Protein constructs lacking the AD1 and / or ID1 domain could also be investigated for their crystallization behavior. The protein construct lacking AD1 represents an inducible protein in which the positions of the gain of function mutations are still present. Protein constructs lacking parts of the ID1 domain, are hyperactive and therefore structural changes, normally induced by the gain of function mutations, have already been accomplished. Although such a structure would be interesting, it remains essential to solve the structure of the WT protein for comparison. A protein construct lacking AD1 and ID1 is still inducible with benomyl and contains the regions of the protein with most of the mutations. A crystal structure of this construct might be interesting to analyze the locations of the mutations. In addition, inducers such as benomyl or H₂O₂ might be interesting compounds for biochemical experiments. Changes in the overall fold could be monitored by CD after the addition of the inducers to the purified protein. The inducing agents could also be included as ligands in co-crystallization approaches. Other changes in the Mrr1 constructs could include an *N*-terminal elongation and or *C*-terminal shortening of the original '250' construct. The *C*-terminal attachment of a 3-fold HA epitope tag to WT Mrr1 led to *in vivo* Mdr1 promoter upregulation (Schubert et al., 2011a). A similar construct could be cloned and expressed recombinantly in *E. coli*; the '250' variant or maybe the FL protein would be the initial constructs to start with.

Conclusion and perspectives

Even though other constructs can be generated crystallization approaches with the established '250' constructs should be further pursued. The obtained crystallization conditions could so far only be optimized with the help of the crystallization robot where also different drop sizes were tested. These fine screens and in addition initial screening approaches containing crystallization seeds might either result in new crystallization conditions or in hopefully larger crystals in 'old' conditions. Commercially available additive screens should also be tested in combination with the obtained hit conditions. A promising approach could also be co-crystallization experiments with the inducing agents benomyl or H₂O₂.

5.2 FadA5 – a β -ketoacyl-CoA thiolase from *M. tuberculosis*

The aim of this project was the structural characterization of the mycobacterial thiolase FadA5 in the absence and the presence of ligands. As FadA5 was shown to participate in the side-chain degradation of cholesterol and displays a clear preference for β -keto-steroid-CoA thioester, co-crystallization experiments with steroid-CoA compounds were pursued as well (Schaefer et al.). The structural characterization of FadA5 should provide the basis for a better mechanistic understanding of this thiolase and will be the basis to identify important binding pockets and residues involved in ligand binding. Initial steps towards lead determination and developments should be pursued.

5.2.1 Purification and crystallization

Purification of the FadA5 wild type protein and its variants C93S and C93A, was highly successful and the Initial crystallization attempts were also highly successful but surprisingly many crystals were obtained that did not permit refinement. Crystallization therefore remains coincidental and not an established process. Octahedral crystals were the predominant crystal form of all crystallization experiments belonging to space group I4₁22, and leading to a dataset that prohibited refinement. No other crystallization condition was similarly reproducible or led to datasets of similar consistent high resolution. It is thus not possible to report a reliable crystallization condition for FadA5 WT or its variants, which could be used for upcoming co-crystallization experiments. Compared to the wild type protein crystallization of the FadA5 C93S variant in the presence of ligands was more readily resulting in crystal growth however unfortunately belonging to space group I4₁22. Hence the majority of the structures in complex with ligands described below were derived from this variant.

The FadA5 WT protein was successfully crystallized and data sets were obtained for two different apo crystals and for FadA5 WT in complex with (acetyl-)CoA. The latter structure contains an

Conclusion and perspectives

acetylation at the catalytic cysteine C93 and the protein is in complex, with both, acetyl-CoA and CoA. The C93S variant was co-crystallized with acetyl-CoA again resulting in an acetylated enzyme but this time in complex with CoA, only. The last obtained structure is a co-crystal structure of the FadA5 C93S variant with 3-oxo-pregn-4-ene-20-carboxylic acid (OPC) and CoA. The precursor of OPC is 3-oxo-pregn-4-ene-20-carboxyl-CoA (3-OPC-CoA), a steroid and product of the degradation process performed by this thiolase. During the crystallization process, the steroid-CoA was hydrolytically cleaved, thus resulting in a complex with the single components of 3-OPC-CoA. In addition to the 3-OPC and CoA containing structure, attempts were made to obtain structures of the FadA5 C93S or C93A variants in complex with 3-oxo-cholesterol-4,22-diene-24-oyl-CoA (3-OCD-CoA), which is also a steroid and a substrate analog for the degradative direction. These crystallization trials so far only led to small needle or plate like crystals that could not be significantly improved.

Two different dimer forms have been obtained. The dominant dimer was observed in all ligand bound structures and one of the apo structures. This dimer assumes the typical thiolase fold as described in the literature and employs the commonly known dimer interface, whereas a tetramerization motif, as known from biosynthetic enzymes and also described for some degradative enzymes, was not identified. (Haapalainen et al., 2006; Mathieu et al., 1997) The catalytic triad consists of two cysteines, C93 and C377, and one histidine, H347, which is also conserved in the structure with the different interface. The second dimer form was only observed in one of the apo structures and comprises large reorientations of distinct regions in the polypeptide, which are most likely caused by the formation of disulfide bonds between C59 and C91 as well as C93 and C377. When the enzyme assumes this conformation it is likely to be inactive and a similar conformation was also described for a 3-ketoacyl-CoA thiolase from *Arabidopsis thaliana* but so far the physiological relevance has hardly been investigated. (Pye et al., 2010; Sundaramoorthy et al., 2006)

5.2.2 Structure analysis and comparison of the enzyme in the prominent dimeric form

In 2002, Kursula *et al.* described the catalytic cycle of a biosynthetic thiolase from *Z. ramigera* after obtaining several complexes containing varying ligands. (Kursula et al., 2002) In contrast, no comprehensive study of a degradative enzyme was performed so far. In this thesis, four different structures of the degradative thiolase FadA5 were obtained, that revealed significant insights into the degradation process performed by this thiolase. (Schaefer et al.)

The different structures represent nearly every step of the thiolase reaction cycle (Figure 11) and thus permit detailed mechanistic insights. For the first step of the cleavage reaction, the β -keto thioester substrate (Figure 10, compound 2) would be in complex with the apo enzyme. The apo

Conclusion and perspectives

FadA5 structure was solved in a most likely active and an inactive form. The active form was used for comparison with the ligand bound structures. This comparison showed that the apo enzyme is structurally highly similar to the complex structures with bound ligands. It was therefore hypothesized that the only visible movement upon binding of the substrate in state I (**Figure 11**), would be a shift of C93 to permit entry to the steroid-binding pocket. Movement of this residue was observed in all obtained complex structures of FadA5.

State II of this cycle is represented by the acetylated WT(-acetyl)-CoA complex structure where the active site cysteine 93 of the acylated enzyme is located in close proximity to an acetyl-CoA that theoretically was just cleaved off the substrate. In the acetylated WT structure the acylation with a bulky steroid is mimicked by the much shorter acetyl, but even in the presence of this shortened compound the orientation of the “real” substrate is unambiguous and is confirmed by the OPC-complex structure. The OPC-complex structure and the acetylated C93S-CoA complex mimic state III (**Figure 11**) of the degradation process. The latter structure is of minor importance and shall not be discussed in detail because it does not provide additional information in comparison to the acetylated WT complex or the OPC-complex. The OPC-complex, in contrast, provides unique information since the occupation of the steroid-binding pocket with the required residue movements, i.e. Q151 and R136, for proper binding could be observed and analyzed for the first time. The distinct movements of the arginine and the glutamine (**Figure 58**) hold the steroid, which is shortened by two carbons, in place until it is transferred to CoA. The mimic is not perfect, as a covalent bond between the active site residue S93 and OPC could not be observed. In the case of a covalent complex, one would expect a single conformation of S93 in the B orientation of the two observed conformations and an OPC molecule that is in slightly closer proximity to the CoA than observed in our structure. In addition to OPC, the complex contains a bound CoA molecule that can be activated by deprotonation with the help of the C377-thiolate. The activated CoA reacts with the covalently bound product that is thereby released from the enzyme. Finally, the enzyme is in complex with a more loose, non-covalently, bound intermediate (**Figure 11**, state IV), for which no structural information is available so far.

A general question that arises and cannot completely be addressed here is the binding and release mechanism of the steroids. They are located in a hydrophobic binding pocket during catalysis, but due to the rigidity of the enzyme in the apo or ligand-bound state, it is not clear whether the compound is transferred along the CoA-binding site and the catalytic triad or if a larger reorientation of the steroid-binding loop occurs. The fact that a hydrophobic compound would have to pass along hydrophilic binding regions makes a reorientation or opening of the binding loop more plausible. This hypothesis is also supported by chain D of the OPC-complex structure where no OPC binding was observed and a highly disordered binding loop (residues G133 – D147, **Figure 56C**) and a more

Conclusion and perspectives

flexible bound CoA ligand are present. In this chain, a snapshot with an opened binding loop that may just have released its product is observed and a solvent exposed steroid-binding site is displayed. The lower occupied CoA or acetyl-CoA ligands (**Figure 53A**) observed in the other complex structures together with the more flexible bound CoA in chain D of the OPC complex suggest that high occupancy CoA binding in FadA5 only occurs when the steroid-binding pocket is occupied. However, structural differences in the CoA-binding pocket of the steroid-bound and steroid-free structures were not observed and thus this phenomenon cannot be explained. Based on the different enzyme snapshots it is hypothesized that despite the fact that the overall fold of the protein is rigid, FadA5 may undergo larger reorientations to allow substrate entry and product release and that an occupation of both binding sites is essential for tight ligand binding, and catalysis.

A successful co-crystallization with the above mentioned substrate analog, 3-OCD-CoA, would provide distinct information about enzyme state I (**Figure 11**) and thereby the positioning of the steroid which is longer by two carbon atoms.

5.2.3 Steps towards the determination of a lead compound

As cholesterol metabolism became a focus of being an anti-TB drug target, the question of FadA5's potential and the feasibility of drug development arises. Despite their low sequence homology, thiolases display a highly conserved fold throughout the entire enzyme class, which leads to the question whether FadA5 can be specifically targeted without inhibiting human thiolases as well?

To address this question, structural comparisons between FadA5 and the human thiolases were performed. For this purpose, the structurally so far not characterized human enzymes, the T1-thiolase, the peroxisomal SCP2-thiolase and mitochondrial trifunctional enzyme thiolase were modeled as described in 4.2.6. Human thiolases can be subdivided into at least four different branches with respect to their phylogenetic relationship, but also with respect to their structural similarities. (Anbazhagan et al., 2014; Schaefer et al.) T1, T2 and CT thiolase are clearly closely related to each other and share a similar fold. These three enzymes form a tetramerization motif instead of the, for FadA5 described, steroid-binding pocket (**Figure 60A**). Modeling of the SCP2-thiolase was ambiguous and therefore also not straightforward to evaluate. This enzyme, for which different isoforms are described, can be more than 150 amino acids longer than FadA5 and these residues were difficult to model. However, 15 of the 20 released model structures do not contain a steroid-binding pocket and it can therefore be speculated that binding of a steroid-based inhibitor is not likely to take place. In addition, this thiolase is also only distantly related to FadA5 and more closely related mycobacterial enzymes, without structural information so far, have been reported.

Conclusion and perspectives

(Anbazhagan et al., 2014) Inhibition of the AB thiolase, a closer relative to FadA5, cannot be excluded, because this enzyme might also be targeted by a steroid-based lead compound. Although structural information is available for this enzyme, a comparison is difficult, as crucial regions, near the steroid-binding pocket are missing. However, it seems not likely for this enzyme, to accommodate steroid-compounds without larger structural rearrangements. On the one hand rearrangements seem necessary, to avoid clashes of the enzyme with the steroid and on the other a compensation for missing shielding residues, as R136 in FadA5, would be expected. The human TFE thiolase is the closest relative to FadA5. One fifth of the models proposed the formation of a steroid-binding pocket whereas all other models lacked this pocket. The likelihood that TFE assumes the steroid binding pocket fold is thus not extremely high. Although FadA5 is classified to be a TFE like enzyme, another *M. tuberculosis* thiolase has been described, that is more closely related to the human enzyme and these differences might enable specific FadA5 targeting.

Deduced from the models and the comparison, an inhibition of the human thiolases cannot be excluded, but important differences, which are confirmed for the published human structures, could be highlighted. In addition, crucial moieties, as the steroid-binding pocket and rearranged residues, are presented that might be important factors to specifically target FadA5 without affecting the human enzymes. The kinetic data on FadA5 showed a 400-fold higher specificity for a steroid, the 3,22-dioxo-chol-4-ene-24-oyl-CoA, as a substrate compared to acetoacetyl-CoA, which is due to a 40-fold reduction in K_{mA} and a 10-fold increase in k_{cat} with the steroid substrate further confirming the preference for steroid-CoA substrates for this thiolase. (Schaefer et al.)

To address *de novo* and *in silico* lead structure determination the FadA5 OPC-complex was also examined by a hotspot analysis to identify binding hotspots.

The hotspot analysis of the steroid- and CoA-binding pocket, using HotspotsX (Neudert and Klebe, 2011), was performed by Dr. Johannes Schiebel and revealed the most interesting hotspots in the region of the active site and near the adenine moiety of CoA. Targeting solely the latter hotspot cluster seems not very reasonable, as especially the orientation of the CoA ligand is widely conserved throughout the thiolase class, independent from the substituent at the thiol end. Therefore, targeting the active site pocket instead, seems to be the much better strategy to develop FadA5 specific inhibitors. The hotspot analysis indicates that a potential FadA5 inhibitor should contain a hydrophobic core, which might be derived from a steroid that is especially elongated into the active site to include identified binding hotspots.

Based on the OPC-complex and the acetylated FadA5 WT(-acetyl)-CoA structure as well as the comparison with the human thiolases and the performed hotspot analysis, suggestions to optimize binding can be made.

Conclusion and perspectives

In **Figure 63** the chemical structure of 3-OPC-CoA is shown, which will be used to discuss modification options and identified important functional groups that interact with the protein. The 3'-phosphate ADP will be neglected as a modification site since, this moiety is common to all thiolases. It could also be difficult to include the pantetheine as Meriläinen *et al.* showed that not the 3'-phospho-ADP is crucial for the correct positioning of the substrate but rather the pantetheine and especially the thioester or thiol are extremely important. (Meriläinen *et al.*, 2008) If the 3'-phospho-ADP is missing, the K_m is only increased from 24 μM to 73 μM and the k_{cat} is affected to a very minor extent in case of the biosynthetic thiolase from *Z. ramigera*. Hence, the pantetheine would have to be modified in a way, which would lead to higher affinities to FadA5 than to other thiolases. The orange arrows in **Figure 63** indicate atoms of this moiety, which interact with FadA5 in the OPC complex. The green

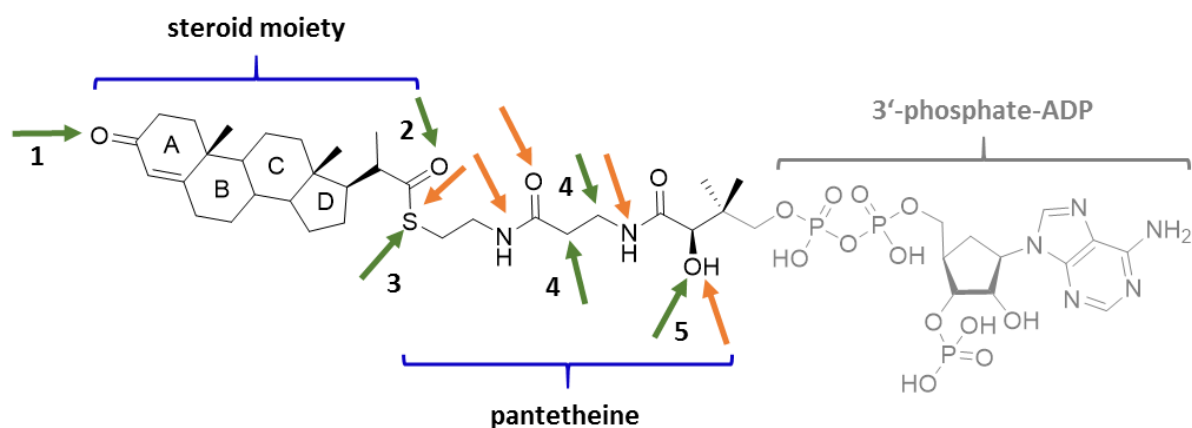


Figure 63 - Lead compound determination based on the FadA5 product 3-OPC-CoA:

3-OPC-CoA can be used as basis for lead determination. The development of a lead structure could focus on the steroid and pantetheine moieties, with alterations and or substitutions of already interacting functional groups (orange arrows). Hotspot analysis as well as structural comparison and analysis of additional functional groups revealed locations (green arrows) in the compound, where a modification might enhance the affinity and specificity to FadA5.

arrows depict targetable or changeable atoms, which might enhance the specificity to FadA5 and thereby reduce the affinity to other thiolases and these modifications might improve the overall binding in general. For example, in position 1 an elongation of the keto-function to an ester might enable the interaction of the steroid in the pocket with the peptide backbone. A similar elongation might be achieved, if not a product analog but a substrate analog, with two additional carbons, is chosen, which expands the linkage between the pantetheine and the actual sterol. Position 2 is interacting with the protein in the acetylated WT(-acetyl)-CoA structure and should be addressed as well (**Figure 53B**). Position 3, the thioester sulfur atom is crucial for correct substrate positioning in the active site (Meriläinen *et al.*, 2008) and it is already interacting via the catalytic water with S93 and H347 (**Figure 55D**). As thioesters are far more reactive than oxoesters, compounds, where the sulfur is replaced by an oxygen could be tested, but alterations might be more efficient at position 2, so that the correct positioning of the compound is not affected. The arrows at position 4 indicate an

Conclusion and perspectives

important binding hotspot. Substitution of the hydrogen of the hydrocarbon, indicated by arrows 4, with an acceptor could generate a buried hydrogen bond, which is positioned between H347 and S246. The hydroxyl, indicated by arrow 5, is already interacting with FadA5 via a water molecule. This end of the pantetheine could be elongated, to enable a direct contact between the ligand and the protein, or the compound could be terminated by a specifically to FadA5 tailored 5- or 6-membered ring.

Functionalization of the steroid compound thus seems to be reasonable and feasible to create an initial lead structure. In a broader approach, the various FadA5 structures, obtained during this work, could be used for a virtual screening approach e.g. using the ZINC database (Irwin and Shoichet, 2005), which contains commercially available compounds, to identify initial hits. These hits could subsequently be tested for their binding to FadA5 and in co-crystallization approaches. In the following, promising compounds could be used for lead optimization.

The here described experiments, the structural data, the homology analysis with the human enzymes, the hotspot analysis and the kinetic data of FadA5 lead to the following conclusion: Due to the low similarity of FadA5 compared to the human enzymes and the preference for steroid-CoA substrates over acetoacetyl-CoA substrates it may be feasible to target FadA5 specifically. Therefore steps in the direction of lead determination should be pursued.

5.2.4 The other dimer - a redox switch?

As reported in the results section, FadA5 WT crystallized in the absence of ligands in two different forms. Not only the crystal shape, octahedral versus hexagonal, space group symmetry or cell dimensions were differing but also the thiolase dimer interface. The interface, observed for one apo structure and all ligand bound structures is conserved throughout the thiolase class, whereas the hexagonal dimer interface is clearly diverging. Both protein purifications were performed without any reducing agents and also the final protein buffers did not contain any such additives. A similar shift in a thiolase dimer was so far only described for one 3-ketoacyl-CoA thiolase in *A. thaliana* (AtKAT) and there suggestions were made about a peroxisomal redox switch which may control the thiolytic cleavage reaction of this degrading enzyme. (Pye et al., 2010; Sundaramoorthy et al., 2006) The two different crystal structures of the AtKAT enzyme were obtained under reducing and non-reducing conditions. The authors also describe the formation of a disulfide bond, involving the important nucleophilic cysteine of the catalytic triad, and specify this oxidized enzyme to be inactive. Apo structure II of FadA5 revealed the formation of two disulfide bonds, involving both catalytic cysteines and in addition C59 and C91. Most likely FadA5 is inactive in this form as well. Pye *et al.*

Conclusion and perspectives

investigated the thiolase activity dependent on oxidizing (cystamine, CSSC) and reducing (cysteamine, CSH) agents and identified a sensitive redox switch for this enzyme.

Although similar experiments, performed by our collaboration partners, Rui Lu from Prof. Nicole Sampson laboratory (Stony Brook University, NY, USA) are still preliminary, FadA5 activity seems to be influenced by CSSC or CSH addition. The influence of these compounds was examined in the FadA5 activity assay, where the conversion of 3,22-dioxo-chol-4-ene-24-oyl-CoA (**Figure 10**, compound 2) into 3-oxo-pregn-4-ene-20-carboxyl-CoA (3-OPC-CoA, **Figure 10**, compound 3) is analyzed by mass spectrometry. FadA5 was thus oxidized, reduced, or just analyzed in buffer, without the presence of additives. A 10-fold reduced activity for the oxidized enzyme in comparison to the buffer control was observed, whereas the presence of a reducing agent led to a 2-fold increase of enzyme activity. Future experiments will further target the redox properties of FadA5, including the determination of FadA5's reduction potential to evaluate if it is physiologically relevant. In addition, FadA5 should be crystallized after incubation with CSSC or CSH to confirm the structural reorientations caused by these agents and compare them to the so far obtained crystal structures.

These experiments should lead to a more definitive conclusion whether a redox switch is present in a bacterial thiolase.

6 References

- Selman A. Waksman - Banquet Speech. (Nobelprize.org).
- (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50, 760-763.
- (2006). Purification of Polyhistidine-Tagged Proteins (Machery-Nagel).
- (2008). Protein production and purification. *Nat Meth* 5, 135-146.
- (2010). The PyMOL Molecular Graphics System. (Schrodinger, LLC).
- Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66, 213-221.
- Ahmad, S. (2011). Pathogenesis, immunology, and diagnosis of latent *Mycobacterium tuberculosis* infection. *Clinical & developmental immunology* 2011, 814943.
- Alarco, A.M., and Raymond, M. (1999). The bZip transcription factor Cap1p is involved in multidrug resistance and oxidative stress response in *Candida albicans*. *J Bacteriol* 181, 700-708.
- Almeida Da Silva, P.E., and Palomino, J.C. (2011). Molecular basis and mechanisms of drug resistance in *Mycobacterium tuberculosis*: classical and new drugs. *J Antimicrob Chemother* 66, 1417-1430.
- Anbazhagan, P., Harijan, R.K., Kiema, T.R., Janardan, N., Murthy, M.R., Michels, P.A., Juffer, A.H., and Wierenga, R.K. (2014). Phylogenetic relationships and classification of thiolases and thiolase-like proteins of *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*. *Tuberculosis (Edinb)*.
- Antononkov, V.D., Van Veldhoven, P.P., Waelkens, E., and Mannaerts, G.P. (1997). Substrate specificities of 3-oxoacyl-CoA thiolase A and sterol carrier protein 2/3-oxoacyl-CoA thiolase purified from normal rat liver peroxisomes. Sterol carrier protein 2/3-oxoacyl-CoA thiolase is involved in the metabolism of 2-methyl-branched fatty acids and bile acid intermediates. *J Biol Chem* 272, 26023-26031.
- Baillie, G.S., and Douglas, L.J. (2000). Matrix polymers of *Candida* biofilms and their possible role in biofilm resistance to antifungal agents. *J Antimicrob Chemother* 46, 397-403.
- Balganesh, T.S., Alzari, P.M., and Cole, S.T. (2008). Rising standards for tuberculosis drug development. *Trends Pharmacol Sci* 29, 576-581.
- Banachowicz, E. (2006). Light scattering studies of proteins under compression. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics* 1764, 405-413.
- Barry, C.E., 3rd, Boshoff, H.I., Dartois, V., Dick, T., Ehrst, S., Flynn, J., Schnappinger, D., Wilkinson, R.J., and Young, D. (2009). The spectrum of latent tuberculosis: rethinking the biology and intervention strategies. *Nat Rev Microbiol* 7, 845-855.
- Battye, T.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr* 67, 271-281.
- Behnen, J., Koster, H., Neudert, G., Craan, T., Heine, A., and Klebe, G. (2012). Experimental and computational active site mapping as a starting point to fragment-based lead discovery. *ChemMedChem* 7, 248-261.
- Berg, J.M., Tymoczko, J.L., and Stryer, L. (2007). *Biochemistry*, 6th edn (New York: W.H. Freeman).
- Bergfors, T.M. (1999). Protein crystallization : techniques, strategies, and tips : a laboratory manual (La Jolla, Calif.: International University Line).
- Berman, J., and Sudbery, P.E. (2002). *Candida Albicans*: a molecular revolution built on lessons from budding yeast. *Nature reviews. Genetics* 3, 918-930.
- Beverly, P.C., Sridhar, S., Lalvani, A., and Tchilian, E.Z. (2014). Harnessing local and systemic immunity for vaccines against tuberculosis. *Mucosal immunology* 7, 20-26.
- Brennan, P.J. (2003). Structure, function, and biogenesis of the cell wall of *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* 83, 91-97.
- Brower, C.M. (2013). This illustration depicts a three-dimensional (3D) computer-generated image of a cluster of rod-shaped drug-resistant *Mycobacterium tuberculosis* bacteria, the pathogen responsible for causing the disease tuberculosis (TB). The artistic recreation was based upon scanning electron micrographic imagery. (Public Health Image Library (PHIL)),
- Brunger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472-475.
- Calderone, R.A. (2002). *Candida and candidiasis* (Washington, D.C.: ASM Press).
- Calderone, R.A., and Fonzi, W.A. (2001). Virulence factors of *Candida albicans*. *Trends Microbiol* 9, 327-335.

References

- Camirero, J.A., Sotgiu, G., Zumla, A., and Migliori, G.B. (2010). Best drug treatment for multidrug-resistant and extensively drug-resistant tuberculosis. *The Lancet Infectious Diseases* 10, 621-629.
- Casabon, I., Crowe, A.M., Liu, J., and Eltis, L.D. (2013). FadD3 is an acyl-CoA synthetase that initiates catabolism of cholesterol rings C and D in actinobacteria. *Mol Microbiol* 87, 269-283.
- Chang, J.C., Miner, M.D., Pandey, A.K., Gill, W.P., Harik, N.S., Sasseti, C.M., and Sherman, D.R. (2009). *igr* Genes and Mycobacterium tuberculosis cholesterol metabolism. *J Bacteriol* 191, 5232-5239.
- Chayen, N.E. (2004). Turning protein crystallisation from an art into a science. *Curr Opin Struct Biol* 14, 577-583.
- Chayen, N.E., and Saridakis, E. (2008). Protein crystallization: from purified protein to diffraction-quality crystal. *Nat Meth* 5, 147-153.
- Chen, I., and Dubnau, D. (2004). DNA uptake during bacterial transformation. *Nat Rev Microbiol* 2, 241-249.
- Chen, V.B., Arendall, W.B., 3rd, Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica. Section D, Biological crystallography* 66, 12-21.
- Cheng, J., Randall, A.Z., Sweredoski, M.J., and Baldi, P. (2005). SCRATCH: a protein structure and structural feature prediction server. *Nucleic acids research* 33, W72-76.
- Clontech (2009). TALON® Metal Affinity Resins - User Manual (Clontech).
- Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., *et al.* (1998). Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* 393, 537-544.
- Coste, A., Selmecki, A., Forche, A., Diogo, D., Bounoux, M.E., d'Enfert, C., Berman, J., and Sanglard, D. (2007). Genotypic evolution of azole resistance mechanisms in sequential *Candida albicans* isolates. *Eukaryot Cell* 6, 1889-1904.
- Coste, A., Turner, V., Ischer, F., Morschhauser, J., Forche, A., Selmecki, A., Berman, J., Bille, J., and Sanglard, D. (2006). A mutation in Tac1p, a transcription factor regulating CDR1 and CDR2, is coupled with loss of heterozygosity at chromosome 5 to mediate antifungal resistance in *Candida albicans*. *Genetics* 172, 2139-2156.
- Coste, A.T., Karababa, M., Ischer, F., Bille, J., and Sanglard, D. (2004). TAC1, transcriptional activator of CDR genes, is a new transcription factor involved in the regulation of *Candida albicans* ABC transporters CDR1 and CDR2. *Eukaryot Cell* 3, 1639-1652.
- Cowen, L.E. (2008). The evolution of fungal drug resistance: modulating the trajectory from genotype to phenotype. *Nat Rev Microbiol* 6, 187-198.
- da Silva, P.E., Von Groll, A., Martin, A., and Palomino, J.C. (2011). Efflux as a mechanism for drug resistance in Mycobacterium tuberculosis. *FEMS immunology and medical microbiology* 63, 1-9.
- de Micheli, M., Bille, J., Schueller, C., and Sanglard, D. (2002). A common drug-responsive element mediates the upregulation of the *Candida albicans* ABC transporters CDR1 and CDR2, two genes involved in antifungal drug resistance. *Mol Microbiol* 43, 1197-1214.
- Dedkova, E.N., Seidlmayer, L.K., and Blatter, L.A. (2013). Mitochondria-mediated cardioprotection by trimetazidine in rabbit heart failure. *J Mol Cell Cardiol* 59, 41-54.
- Del Sorbo, G., Schoonbeek, H., and De Waard, M.A. (2000). Fungal transporters involved in efflux of natural toxic compounds and fungicides. *Fungal Genet Biol* 30, 1-15.
- DeLano, W.L. (2002). The PyMOL Molecular Graphics System on World Wide Web
- Diederichs, K., and Karplus, P.A. (1997). Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol* 4, 269-275.
- Dodero, V.I., Quirolo, Z.B., and Sequeira, M.A. (2011). Biomolecular studies by circular dichroism. *Front Biosci (Landmark Ed)* 16, 61-73.
- Dorland, W.A.N. (2007). *Dorland's Illustrated Medical dictionary*, 31 edn (Philadelphia, Pa. New York: Saunders Press ; distributed by Holt, Rinehart & Winston).
- Dorman, S.E., and Chaisson, R.E. (2007). From magic bullets back to the magic mountain: the rise of extensively drug-resistant tuberculosis. *Nat Med* 13, 295-298.
- Douglas, L.J. (2003). *Candida* biofilms and their role in infection. *Trends Microbiol* 11, 30-36.
- Doyle, M.L. (1997). Characterization of binding interactions by isothermal titration calorimetry. *Current opinion in biotechnology* 8, 31-35.
- Dubnau, E., Chan, J., Mohan, V.P., and Smith, I. (2005). Responses of mycobacterium tuberculosis to growth in the mouse lung. *Infect Immun* 73, 3754-3757.

References

- Dubnau, E., Fontan, P., Manganelli, R., Soares-Appel, S., and Smith, I. (2002). Mycobacterium tuberculosis genes induced during infection of human macrophages. *Infect Immun* 70, 2787-2795.
- Dunkel, N., Blass, J., Rogers, P.D., and Morschhauser, J. (2008a). Mutations in the multi-drug resistance regulator MRR1, followed by loss of heterozygosity, are the main cause of MDR1 overexpression in fluconazole-resistant *Candida albicans* strains. *Mol Microbiol* 69, 827-840.
- Dunkel, N., Liu, T.T., Barker, K.S., Homayouni, R., Morschhauser, J., and Rogers, P.D. (2008b). A gain-of-function mutation in the transcription factor Upc2p causes upregulation of ergosterol biosynthesis genes and increased fluconazole resistance in a clinical *Candida albicans* isolate. *Eukaryot Cell* 7, 1180-1190.
- Eddouzi, J., Lohberger, A., Vogne, C., Manai, M., and Sanglard, D. (2013a). Identification and antifungal susceptibility of a large collection of yeast strains isolated in Tunisian hospitals. *Med Mycol* 51, 737-746.
- Eddouzi, J., Parker, J.E., Vale-Silva, L.A., Coste, A., Ischer, F., Kelly, S., Manai, M., and Sanglard, D. (2013b). Molecular mechanisms of drug resistance in clinical *Candida* species isolated from Tunisian hospitals. *Antimicrob Agents Chemother* 57, 3182-3193.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60, 2126-2132.
- Ericsson, U.B., Hallberg, B.M., DeTitta, G.T., Dekker, N., and Nordlund, P. (2006). Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry* 357, 289-298.
- Erinc Sahin, C.J.R. (2012). Size-Exclusion Chromatography with Multi-angle Light Scattering for Elucidating Protein Aggregation Mechanisms. In *Therapeutic Proteins Methods and Protocols* (Springer), p. 516.
- Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62, 72-82.
- Evans, P.R. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr* 67, 282-292.
- Fair, R.J., and Tor, Y. (2014). Antibiotics and bacterial resistance in the 21st century. *Perspectives in medicinal chemistry* 6, 25-64.
- Feldmann, H. *Yeast : molecular and cell biology*, 2nd, completely revised and greatly enlarged edition. edn.
- Fould, B., Garlatti, V., Neumann, E., Fenel, D., Gaboriaud, C., and Arlaud, G.J. (2010). Structural and functional characterization of the recombinant human mitochondrial trifunctional protein. *Biochemistry* 49, 8608-8617.
- French, S., and Wilson, K. (1978). On the treatment of negative intensity observations. *Acta Crystallographica Section A* 34, 517-525.
- Gandhi, N.R., Nunn, P., Dheda, K., Schaaf, H.S., Zignol, M., van Soolingen, D., Jensen, P., and Bayona, J. (2010). Multidrug-resistant and extensively drug-resistant tuberculosis: a threat to global control of tuberculosis. *Lancet* 375, 1830-1843.
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., and Bairoch, A. (2003). ExpASY: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31, 3784-3788.
- GEHealthcare Gel Filtration - Principles and Methods (GE Healthcare).
- Ghannoum, M.A., and Rice, L.B. (1999). Antifungal agents: mode of action, mechanisms of resistance, and correlation of these mechanisms with bacterial resistance. *Clinical microbiology reviews* 12, 501-517.
- Gilbert, H.F. (1981). Proton transfer from acetyl-coenzyme A catalyzed by thiolase I from porcine heart. *Biochemistry* 20, 5643-5649.
- Gilbert, H.F., Lennox, B.J., Mossman, C.D., and Carle, W.C. (1981). The relation of acyl transfer to the overall reaction of thiolase I from porcine heart. *The Journal of biological chemistry* 256, 7371-7377.
- Haapalainen, A.M., Merilainen, G., Pirila, P.L., Kondo, N., Fukao, T., and Wierenga, R.K. (2007). Crystallographic and kinetic studies of human mitochondrial acetoacetyl-CoA thiolase: the importance of potassium and chloride ions for its structure and function. *Biochemistry* 46, 4305-4321.
- Haapalainen, A.M., Merilainen, G., and Wierenga, R.K. (2006). The thiolase superfamily: condensing enzymes with diverse reaction specificities. *Trends Biochem Sci* 31, 64-71.
- Harry, J.B., Oliver, B.G., Song, J.L., Silver, P.M., Little, J.T., Choiniere, J., and White, T.C. (2005). Drug-induced regulation of the MDR1 promoter in *Candida albicans*. *Antimicrob Agents Chemother* 49, 2785-2792.
- Heilmann, C.J., Schneider, S., Barker, K.S., Rogers, P.D., and Morschhauser, J. (2010). An A643T mutation in the transcription factor Upc2p causes constitutive ERG11 upregulation and increased fluconazole resistance in *Candida albicans*. *Antimicrob Agents Chemother* 54, 353-359.
- Hershkovitz, I., Donoghue, H.D., Minnikin, D.E., Besra, G.S., Lee, O.Y., Germaey, A.M., Galili, E., Eshed, V., Greenblatt, C.L., Lemma, E., *et al.* (2008). Detection and molecular characterization of 9,000-year-old *Mycobacterium tuberculosis* from a Neolithic settlement in the Eastern Mediterranean. *PLoS One* 3, e3426.

References

- Hoehamer, C.F., Cummings, E.D., Hilliard, G.M., Morschhauser, J., and Rogers, P.D. (2009). Proteomic analysis of Mrr1p- and Tac1p-associated differential protein expression in azole-resistant clinical isolates of *Candida albicans*. *Proteomics Clin Appl* 3, 968-978.
- Hol, W.G. (1985). The role of the alpha-helix dipole in protein function and structure. *Progress in biophysics and molecular biology* 45, 149-195.
- Hoot, S.J., Smith, A.R., Brown, R.P., and White, T.C. (2011). An A643V amino acid substitution in Upc2p contributes to azole resistance in well-characterized clinical isolates of *Candida albicans*. *Antimicrob Agents Chemother* 55, 940-942.
- Hull, C.M., and Johnson, A.D. (1999). Identification of a mating type-like locus in the asexual pathogenic yeast *Candida albicans*. *Science* 285, 1271-1275.
- Invitrogen (2010). PichiaPink Expression System - User Manual, Vol Manual part no. A10984 (Invitrogen by life technologies).
- Irwin, J.J., and Shoichet, B.K. (2005). ZINC--a free database of commercially available compounds for virtual screening. *J Chem Inf Model* 45, 177-182.
- Kabsch, W. (2010). XDS. *Acta Crystallographica Section D* 66, 125-132.
- Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577-2637.
- Kaplan, C.D.W. (1969). This is an image of a SABHI agar plate culture of the fungus *Candida albicans* grown at 20 °C. (Public Health Image Library (PHIL)) Karkowska-Kuleta, J., Rapala-Kozik, M., and Kozik, A. (2009). Fungi pathogenic to humans: molecular bases of virulence of *Candida albicans*, *Cryptococcus neoformans* and *Aspergillus fumigatus*. *Acta biochimica Polonica* 56, 211-224.
- Karplus, P.A., and Diederichs, K. (2012). Linking crystallographic model and data quality. *Science* 336, 1030-1033.
- Kayser, F.H., Bienz, F.H., Eckert, K.A., and J., Z. (2001). *Medizinische Mikrobiologie*, 10. edn (Stuttgart ; New York, NY: Georg Thieme Verlag).
- Kelley, L.A., and Sternberg, M.J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* 4, 363-371.
- Kendall, S.L., Withers, M., Soffair, C.N., Moreland, N.J., Gurucha, S., Sidders, B., Frita, R., Ten Bokum, A., Besra, G.S., Lott, J.S., *et al.* (2007). A highly conserved transcriptional repressor controls a large regulon involved in lipid degradation in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis*. *Mol Microbiol* 65, 684-699.
- Kim, M.J., Wainwright, H.C., Locketz, M., Bekker, L.G., Walther, G.B., Dittrich, C., Visser, A., Wang, W., Hsu, F.F., Wiehart, U., *et al.* (2010). Caseation of human tuberculosis granulomas correlates with elevated host lipid metabolism. *EMBO molecular medicine* 2, 258-274.
- Koul, A., Arnoult, E., Lounis, N., Guillemont, J., and Andries, K. (2011). The challenge of new drug discovery for tuberculosis. *Nature* 469, 483-490.
- Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta crystallographica. Section D, Biological crystallography* 60, 2256-2268.
- Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372, 774-797.
- Krug, M., Weiss, M.S., Heinemann, U., and Mueller, U. (2012). XDSAPP: a graphical user interface for the convenient processing of diffraction data using XDS. *Journal of Applied Crystallography* 45, 568-572.
- Kubica, C.D.G. (1976). This is a close-up of a *Mycobacterium tuberculosis* culture revealing this organism's colonial morphology. (Public Health Image Library (PHIL)) Kursula, P., Ojala, J., Lambeir, A.M., and Wierenga, R.K. (2002). The catalytic cycle of biosynthetic thiolase: a conformational journey of an acetyl group through four binding modes and two oxyanion holes. *Biochemistry* 41, 15543-15556.
- Kursula, P., Sikkila, H., Fukao, T., Kondo, N., and Wierenga, R.K. (2005). High resolution crystal structures of human cytosolic thiolase (CT): a comparison of the active sites of human CT, bacterial thiolase, and bacterial KAS I. *J Mol Biol* 347, 189-201.
- Langer, G., Cohen, S.X., Lamzin, V.S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature protocols* 3, 1171-1179.
- Leibert, E., Danckers, M., and Rom, W.N. (2014). New drugs to treat multidrug-resistant tuberculosis: the case for bedaquiline. *Therapeutics and clinical risk management* 10, 597-602.
- Leslie, A.G.W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography*, No. 26.
- Li, H., Robertson, A.D., and Jensen, J.H. (2005). Very fast empirical prediction and rationalization of protein pKa values. *Proteins* 61, 704-721.
- Li, M.Z., and Elledge, S.J. (2007). Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nature methods* 4, 251-256.

References

- Liang, X., and Schulz, H. (1998). 4-bromotiglic acid, a novel inhibitor of thiolases and a tool for assessing the cooperation between the membrane-bound and soluble beta-oxidation systems of rat liver mitochondria. *Biochemistry* 37, 15548-15554.
- Lo, H.J., Kohler, J.R., DiDomenico, B., Loebenberg, D., Cacciapuoti, A., and Fink, G.R. (1997). Nonfilamentous *C. albicans* mutants are avirulent. *Cell* 90, 939-949.
- Lodi, P.J., and Knowles, J.R. (1993). Direct evidence for the exploitation of an alpha-helix in the catalytic mechanism of triosephosphate isomerase. *Biochemistry* 32, 4338-4343.
- MacPherson, S., Larochelle, M., and Turcotte, B. (2006). A fungal family of transcriptional regulators: the zinc cluster proteins. *Microbiol Mol Biol Rev* 70, 583-604.
- Magee, B.B., and Magee, P.T. (2000). Induction of mating in *Candida albicans* by construction of MTL α and MTL α strains. *Science* 289, 310-313.
- Marmorstein, R., Carey, M., Ptashne, M., and Harrison, S.C. (1992). DNA recognition by GAL4: structure of a protein-DNA complex. *Nature* 356, 408-414.
- Marrero, A., Duquerroy, S., Trapani, S., Goulas, T., Guevara, T., Andersen, G.R., Navaza, J., Sottrup-Jensen, L., and Gomis-Ruth, F.X. (2012). The crystal structure of human alpha2-macroglobulin reveals a unique molecular cage. *Angew Chem Int Ed Engl* 51, 3340-3344.
- Mathieu, M., Modis, Y., Zeelen, J.P., Engel, C.K., Abagyan, R.A., Ahlberg, A., Rasmussen, B., Lamzin, V.S., Kunau, W.H., and Wierenga, R.K. (1997). The 1.8 Å crystal structure of the dimeric peroxisomal 3-ketoacyl-CoA thiolase of *Saccharomyces cerevisiae*: implications for substrate binding and reaction mechanism. *Journal of Molecular Biology* 273, 714-728.
- Matthews, B.W. (1968). Solvent content of protein crystals. *Journal of Molecular Biology* 33, 491-497.
- McCoy, A.J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D Biol Crystallogr* 63, 32-41.
- McFerrin, M.B., and Snell, E.H. (2002). The development and application of a method to quantify the quality of cryoprotectant solutions using standard area-detector X-ray images. *Journal of Applied Crystallography* 35, 538-545.
- McRee, D.E., and David, P.R. (1999). *Practical protein crystallography*, 2nd edn (San Diego, Calif.: Academic Press).
- Merilainen, G., Schmitz, W., Wierenga, R.K., and Kursula, P. (2008). The sulfur atoms of the substrate CoA and the catalytic cysteine are required for a productive mode of substrate binding in bacterial biosynthetic thiolase, a thioester-dependent enzyme. *The FEBS journal* 275, 6136-6148.
- Middleton, B. (1974). The kinetic mechanism and properties of the cytoplasmic acetoacetyl-coenzyme A thiolase from rat liver. *Biochem J* 139, 109-121.
- Miller, M.G., and Johnson, A.D. (2002). White-opaque switching in *Candida albicans* is controlled by mating-type locus homeodomain proteins and allows efficient mating. *Cell* 110, 293-302.
- Miner, M.D., Chang, J.C., Pandey, A.K., Sasseti, C.M., and Sherman, D.R. (2009). Role of cholesterol in *Mycobacterium tuberculosis* infection. *Indian J Exp Biol* 47, 407-411.
- Modis, Y., and Wierenga, R.K. (1999). A biosynthetic thiolase in complex with a reaction intermediate: the crystal structure provides new insights into the catalytic mechanism. *Structure* 7, 1279-1290.
- Modis, Y., and Wierenga, R.K. (2000). Crystallographic analysis of the reaction pathway of *Zoogloea ramigera* biosynthetic thiolase. *Journal of Molecular Biology* 297, 1171-1182.
- Moen, M.D., Lyseng-Williamson, K.A., and Scott, L.J. (2009). Liposomal amphotericin B: a review of its use as empirical therapy in febrile neutropenia and in the treatment of invasive fungal infections. *Drugs* 69, 361-392.
- Mogavero, S., Tavanti, A., Senesi, S., Rogers, P.D., and Morschhauser, J. (2011). Differential Requirement of the Transcription Factor Mcm1 for Activation of the *Candida albicans* Multidrug Efflux Pump MDR1 by Its Regulators Mrr1 and Cap1. *Antimicrob Agents Chemother* 55, 2061-2066.
- Mohn, W.W., van der Geize, R., Stewart, G.R., Okamoto, S., Liu, J., Dijkhuizen, L., and Eltis, L.D. (2008). The actinobacterial *mce4* locus encodes a steroid transporter. *J Biol Chem* 283, 35368-35374.
- Morio, F., Pagniez, F., Besse, M., Gay-andrieu, F., Miegeville, M., and Le Pape, P. (2013). Deciphering azole resistance mechanisms with a focus on transcription factor-encoding genes TAC1, MRR1 and UPC2 in a set of fluconazole-resistant clinical isolates of *Candida albicans*. *Int J Antimicrob Agents* 42, 410-415.
- Morschhauser, J. (2010a). Regulation of multidrug resistance in pathogenic fungi. *Fungal Genet Biol* 47, 94-106.
- Morschhauser, J. (2010b). Regulation of white-opaque switching in *Candida albicans*. *Med Microbiol Immunol* 199, 165-172.
- Morschhauser, J., Barker, K.S., Liu, T.T., Bla, B.W.J., Homayouni, R., and Rogers, P.D. (2007). The transcription factor Mrr1p controls expression of the MDR1 efflux pump and mediates multidrug resistance in *Candida albicans*. *PLoS Pathog* 3, e164.

References

- Mülhardt, C. (2009). *Der Experimentator - Molekularbiologie / Genomics*, 6th edn (Heidelberg: Spektrum Akademischer Verlag).
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* *53*, 240-255.
- Nesbitt, N.M., Yang, X., Fontan, P., Kolesnikova, I., Smith, I., Sampson, N.S., and Dubnau, E. (2010). A thiolase of *Mycobacterium tuberculosis* is required for virulence and production of androstenedione and androstadienedione from cholesterol. *Infect Immun* *78*, 275-282.
- Neudert, G., and Klebe, G. (2011). DSX: a knowledge-based scoring function for the assessment of protein-ligand complexes. *J Chem Inf Model* *51*, 2731-2745.
- Niewerth, M., and Korting, H.C. (2000). The use of systemic antimycotics in dermatotherapy. *European journal of dermatology : EJD* *10*, 155-160.
- Olsson, M.H.M., Søndergaard, C.R., Rostkowski, M., and Jensen, J.H. (2011). PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *Journal of Chemical Theory and Computation* *7*, 525-537.
- Ouellet, H., Johnston, J.B., and de Montellano, P.R. (2011). Cholesterol catabolism as a therapeutic target in *Mycobacterium tuberculosis*. *Trends Microbiol* *19*, 530-539.
- Painter, J., and Merritt, E.A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta crystallographica. Section D, Biological crystallography* *62*, 439-450.
- Pandey, A.K., and Sasseti, C.M. (2008). Mycobacterial persistence requires the utilization of host cholesterol. *Proceedings of the National Academy of Sciences of the United States of America* *105*, 4376-4380.
- Pannu, N.S., Murshudov, G.N., Dodson, E.J., and Read, R.J. (1998). Incorporation of prior phase information strengthens maximum-likelihood structure refinement. *Acta Crystallogr D Biol Crystallogr* *54*, 1285-1294.
- Parish, T., and Brown, A.C. (2008). *Mycobacteria protocols*, 2nd edn (New York, NY: Humana Press).
- Parish, T., and Stoker, N.G. (1998). *Mycobacteria protocols* (Totowa, N.J.: Humana Press).
- Pathologists, C.A.S.o.C. Histopathology of *Candida albicans* infection. Methenamine silver stain. (Public Health Image Library (PHIL)), p. Pseudohyphae and true hyphae.
- Peifer, M., Berg, S., and Reynolds, A.B. (1994). A repeating amino acid motif shared by proteins with diverse cellular roles. *Cell* *76*, 789-791.
- Perlin, D.S. (2007). Resistance to echinocandin-class antifungal drugs. *Drug resistance updates : reviews and commentaries in antimicrobial and anticancer chemotherapy* *10*, 121-130.
- Peyron, P., Vaubourgeix, J., Poquet, Y., Levillain, F., Botanch, C., Bardou, F., Daffe, M., Emile, J.F., Marchou, B., Cardona, P.J., *et al.* (2008). Foamy macrophages from tuberculous patients' granulomas constitute a nutrient-rich reservoir for *M. tuberculosis* persistence. *PLoS Pathog* *4*, e1000204.
- Pierce, M.M., Raman, C.S., and Nall, B.T. (1999). Isothermal titration calorimetry of protein-protein interactions. *Methods* *19*, 213-221.
- Poce, G., Coccoza, M., Consalvi, S., and Biava, M. (2014). SAR analysis of new anti-TB drugs currently in pre-clinical and clinical development. *European journal of medicinal chemistry* *86C*, 335-351.
- Pokhrel, L., Maezawa, I., Nguyen, T.D., Chang, K.O., Jin, L.W., and Hua, D.H. (2012). Inhibition of Acyl-CoA: cholesterol acyltransferase (ACAT), overexpression of cholesterol transporter gene, and protection of amyloid beta (A β) oligomers-induced neuronal cell death by tricyclic pyrone molecules. *Journal of medicinal chemistry* *55*, 8969-8973.
- Polak, A., and Scholer, H.J. (1975). Mode of action of 5-fluorocytosine and mechanisms of resistance. *Chemotherapy* *21*, 113-130.
- Pollastri, G., and McLysaght, A. (2005). Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics* *21*, 1719-1720.
- Price, A.C., Rock, C.O., and White, S.W. (2003). The 1.3-Angstrom-resolution crystal structure of beta-ketoacyl-acyl carrier protein synthase II from *Streptococcus pneumoniae*. *J Bacteriol* *185*, 4136-4143.
- Pye, V.E., Christensen, C.E., Dyer, J.H., Arent, S., and Henriksen, A. (2010). Peroxisomal plant 3-ketoacyl-CoA thiolase structure and activity are regulated by a sensitive redox switch. *The Journal of biological chemistry* *285*, 24078-24088.
- Rhodes, G. (2006). *Crystallography made crystal clear : a guide for users of macromolecular models*, 3rd edn (Amsterdam ; Boston: Elsevier/Academic Press).
- Rognon, B., Kozovska, Z., Coste, A.T., Pardini, G., and Sanglard, D. (2006). Identification of promoter elements responsible for the regulation of MDR1 from *Candida albicans*, a major facilitator transporter involved in azole resistance. *Microbiology* *152*, 3701-3722.

References

- Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5, 725-738.
- Rupp, B. (2010). *Biomolecular crystallography : principles, practice, and application to structural biology* (New York: Garland Science).
- Sacchettini, J.C., Rubin, E.J., and Freundlich, J.S. (2008). Drugs versus bugs: in pursuit of the persistent predator *Mycobacterium tuberculosis*. *Nat Rev Microbiol* 6, 41-52.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239, 487-491.
- Sambrook, J., and Russell, D.W. (2001). *Molecular cloning : a laboratory manual*, 3rd edn (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).
- Sanglard, D. (2002a). Clinical relevance of mechanisms of antifungal drug resistance in yeasts. *Enfermedades infecciosas y microbiologia clinica* 20, 462-469; quiz 470, 479.
- Sanglard, D. (2002b). Resistance of human fungal pathogens to antifungal drugs. *Curr Opin Microbiol* 5, 379-385.
- Sanglard, D., Coste, A., and Ferrari, S. (2009). Antifungal drug resistance mechanisms in fungal pathogens from the perspective of transcriptional gene regulation. *FEMS yeast research* 9, 1029-1050.
- Sanglard, D., Ischer, F., Monod, M., and Bille, J. (1997). Cloning of *Candida albicans* genes conferring resistance to azole antifungal agents: characterization of CDR2, a new multidrug ABC transporter gene. *Microbiology* 143 (Pt 2), 405-416.
- Sanglard, D., Kuchler, K., Ischer, F., Pagani, J.L., Monod, M., and Bille, J. (1995). Mechanisms of resistance to azole antifungal agents in *Candida albicans* isolates from AIDS patients involve specific multidrug transporters. *Antimicrob Agents Chemother* 39, 2378-2386.
- Santos, M.A., and Tuite, M.F. (1995). The CUG codon is decoded in vivo as serine and not leucine in *Candida albicans*. *Nucleic Acids Res* 23, 1481-1486.
- Sasse, C., Dunkel, N., Schafer, T., Schneider, S., Dierolf, F., Ohlsen, K., and Morschhauser, J. (2012). The stepwise acquisition of fluconazole resistance mutations causes a gradual loss of fitness in *Candida albicans*. *Mol Microbiol* 86, 539-556.
- Schaefer, C.M. (2010). Grundlegende Schritte zur strukturellen Charakterisierung eines neu identifizierten, potenziellen Wirkstoffziels im Kampf gegen die Tuberkulose - Analyse des Proteins FadA5 von *Mykobakterium tuberculosis*. In Department of Chemistry (Würzburg, Julius Maximilians University Würzburg).
- Schaefer, C.M., Lu, R., Nesbitt, N.M., Schiebel, J., Sampson, N.S., and Kisker, C. FadA5 a thiolase from *Mycobacterium tuberculosis* - a unique steroid-binding pocket reveals the potential for drug development against tuberculosis. *Structure*.
- Schubert, S., Barker, K.S., Znaidi, S., Schneider, S., Dierolf, F., Dunkel, N., Aid, M., Boucher, G., Rogers, P.D., Raymond, M., et al. (2011a). Regulation of Efflux Pump Expression and Drug Resistance by the Transcription Factors Mrr1, Upc2, and Cap1 in *Candida albicans*. *Antimicrob Agents Chemother* 55, 2212-2223.
- Schubert, S., Popp, C., Rogers, P.D., and Morschhauser, J. (2011b). Functional dissection of a *Candida albicans* zinc cluster transcription factor, the multidrug resistance regulator Mrr1. *Eukaryot Cell* 10, 1110-1121.
- Schubert, S., Rogers, P.D., and Morschhauser, J. (2008). Gain-of-function mutations in the transcription factor MRR1 are responsible for overexpression of the MDR1 efflux pump in fluconazole-resistant *Candida dubliniensis* strains. *Antimicrob Agents Chemother* 52, 4274-4280.
- Schuttelkopf, A.W., and van Aalten, D.M. (2004). PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta crystallographica. Section D, Biological crystallography* 60, 1355-1363.
- Silver, L.L. (2011). Challenges of antibacterial discovery. *Clinical microbiology reviews* 24, 71-109.
- Singh, A.K., and Reyrat, J.M. (2009). Laboratory maintenance of *Mycobacterium smegmatis*. *Curr Protoc Microbiol Chapter* 10, Unit10C 11.
- Slutsky, B., Staebell, M., Anderson, J., Risen, L., Pfaller, M., and Soll, D.R. (1987). "White-opaque transition": a second high-frequency switching system in *Candida albicans*. *J Bacteriol* 169, 189-197.
- Smith, I. (2003). *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clinical microbiology reviews* 16, 463-496.
- Smith, T., Wolff, K.A., and Nguyen, L. (2013). Molecular biology of drug resistance in *Mycobacterium tuberculosis*. *Curr Top Microbiol Immunol* 374, 53-80.
- Snapper, S.B., Melton, R.E., Mustafa, S., Kieser, T., and Jacobs, W.R., Jr. (1990). Isolation and characterization of efficient plasmid transformation mutants of *Mycobacterium smegmatis*. *Mol Microbiol* 4, 1911-1919.
- Stiller, R.L., Bennett, J.E., Scholer, H.J., Wall, M., Polak, A., and Stevens, D.A. (1983). Correlation of in vitro susceptibility test results with in vivo response: flucytosine therapy in a systemic candidiasis model. *J Infect Dis* 147, 1070-1077.

References

- Stover, C.K., Warrenner, P., VanDevanter, D.R., Sherman, D.R., Arain, T.M., Langhorne, M.H., Anderson, S.W., Towell, J.A., Yuan, Y., McMurray, D.N., *et al.* (2000). A small-molecule nitroimidazopyran drug candidate for the treatment of tuberculosis. *Nature* *405*, 962-966.
- Sundaramoorthy, R., Micossi, E., Alphey, M.S., Germain, V., Bryce, J.H., Smith, S.M., Leonard, G.A., and Hunter, W.N. (2006). The crystal structure of a plant 3-ketoacyl-CoA thiolase reveals the potential for redox control of peroxisomal fatty acid beta-oxidation. *Journal of Molecular Biology* *359*, 347-357.
- Taylor, R.G., Walker, D.C., and McInnes, R.R. (1993). *E. coli* host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Res* *21*, 1677-1678.
- Technology, W. (2014a). Chromatography mode. (Wyatt Technology).
- Technology, W. (2014b). Light scattering.
- Thompson, S., Mayerl, F., Peoples, O.P., Masamune, S., Sinskey, A.J., and Walsh, C.T. (1989). Mechanistic studies on beta-ketoacyl thiolase from *Zoogloea ramigera*: identification of the active-site nucleophile as Cys89, its mutation to Ser89, and kinetic and thermodynamic characterization of wild-type and mutant enzymes. *Biochemistry* *28*, 5735-5742.
- Tischler, A.D., and McKinney, J.D. (2010). Contrasting persistence strategies in *Salmonella* and *Mycobacterium*. *Curr Opin Microbiol* *13*, 93-99.
- Tonge, P.J. (2000). Another brick in the wall. *Nat Struct Mol Biol* *7*, 94-96.
- Tsao, S., Rahkhoodaee, F., and Raymond, M. (2009). Relative contributions of the *Candida albicans* ABC transporters Cdr1p and Cdr2p to clinical azole resistance. *Antimicrob Agents Chemother* *53*, 1344-1352.
- Ulrichs, T., and Kaufmann, S. (2006). Immunologie der Tuberkulose und neue Impfstoffansätze. *Monatsschrift Kinderheilkunde* *154*, 133-141.
- Van der Geize, R., Yam, K., Heuser, T., Wilbrink, M.H., Hara, H., Anderton, M.C., Sim, E., Dijkhuizen, L., Davies, J.E., Mohn, W.W., *et al.* (2007). A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)* *104*, 6.
- Vandeputte, P., Ferrari, S., and Coste, A.T. (2012). Antifungal resistance and new strategies to control fungal infections. *International journal of microbiology* *2012*, 713687.
- Velec, H.F., Gohlke, H., and Klebe, G. (2005). DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J Med Chem* *48*, 6296-6303.
- Venkatesan, R., and Wierenga, R.K. (2013). Structure of mycobacterial beta-oxidation trifunctional enzyme reveals its altered assembly and putative substrate channeling pathway. *ACS chemical biology* *8*, 1063-1073.
- Villemagne, B., Crauste, C., Flipo, M., Baulard, A.R., Deprez, B., and Willand, N. (2012). Tuberculosis: the drug development pipeline at a glance. *European journal of medicinal chemistry* *51*, 1-16.
- WHO (2013). Global tuberculosis report 2013. (World Health Organisation).
- WHO (2014). Antimicrobial resistance: global report on surveillance 2014. (World Health Organisation), p. 257.
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., *et al.* (2011). Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography* *67*, 235-242.
- Winn, M.D., Murshudov, G.N., and Papiz, M.Z. (2003). Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods Enzymol* *374*, 300-321.
- Wipperman, M.F., Sampson, N.S., and Thomas, S.T. (2014). Pathogen roid rage: Cholesterol utilization by *Mycobacterium tuberculosis*. *Critical reviews in biochemistry and molecular biology*.
- Wirsching, S., Michel, S., Kohler, G., and Morschhauser, J. (2000a). Activation of the multiple drug resistance gene MDR1 in fluconazole-resistant, clinical *Candida albicans* strains is caused by mutations in a trans-regulatory factor. *Journal of bacteriology* *182*, 400-404.
- Wirsching, S., Michel, S., and Morschhauser, J. (2000b). Targeted gene disruption in *Candida albicans* wild-type strains: the role of the MDR1 gene in fluconazole resistance of clinical *Candida albicans* isolates. *Mol Microbiol* *36*, 856-865.
- Woodruff, H.B. (2014). Selman A. Waksman, winner of the 1952 Nobel Prize for physiology or medicine. *Appl Environ Microbiol* *80*, 2-8.
- Wyatt, P.J. (1993). Light scattering and the absolute characterization of macromolecules. *Analytica Chimica Acta* *272*, 1-40.
- Yam, K.C., D'Angelo, I., Kalscheuer, R., Zhu, H., Wang, J.X., Snieckus, V., Ly, L.H., Converse, P.J., Jacobs, W.R., Jr., Strynadka, N., *et al.* (2009). Studies of a ring-cleaving dioxygenase illuminate the role of cholesterol metabolism in the pathogenesis of *Mycobacterium tuberculosis*. *PLoS Pathog* *5*, e1000344.

References

- Yan, C., Lee, L.H., and Davis, L.I. (1998). Crm1p mediates regulated nuclear export of a yeast AP-1-like transcription factor. *EMBO J* 17, 7416-7429.
- Yang, Y.L. (2003). Virulence factors of *Candida* species. *Journal of microbiology, immunology, and infection = Wei mian yu gan ran za zhi* 36, 223-228.
- Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9, 40.
- Zhang, Y., and Yew, W.W. (2009). Mechanisms of drug resistance in *Mycobacterium tuberculosis*. *The international journal of tuberculosis and lung disease : the official journal of the International Union against Tuberculosis and Lung Disease* 13, 1320-1330.
- Znaidi, S., Barker, K.S., Weber, S., Alarco, A.M., Liu, T.T., Boucher, G., Rogers, P.D., and Raymond, M. (2009). Identification of the *Candida albicans* Cap1p regulon. *Eukaryot Cell* 8, 806-820.
- Znaidi, S., Weber, S., Al-Abdin, O.Z., Bomme, P., Saidane, S., Drouin, S., Lemieux, S., De Deken, X., Robert, F., and Raymond, M. (2008). Genomewide location analysis of *Candida albicans* Upc2p, a regulator of sterol metabolism and azole drug resistance. *Eukaryot Cell* 7, 836-847.
- Zulauf, M., and D'Arcy, A. (1992). Light scattering of proteins as a criterion for crystallization. *Journal of Crystal Growth* 122, 102-106.

7 Copyright

General:

Parts of this thesis were already published in the Cell Press Journal 'Structure'

see (Schaefer et al.)

Extracts from WHO website or publications

WHO exercises copyright over its website and publications. It does so not for financial gain but to ensure that its information is used in accordance with the organization's principles.

Extracts from WHO's website and publications may be reproduced for certain purposes and with certain restrictions, as indicated in the copyright notice.

If you wish to use an extract from the WHO website or from a WHO publication, please check first that the extract (paragraph, table, figure, photograph or other illustration) actually belongs to WHO. If the extract concerned is marked as reproduced from another published source, then please contact the publisher of that source for permission to reproduce it. In such a case, WHO cannot help you.

If the extract is not marked as being reproduced from somewhere else, then it belongs to WHO.

If you wish to use the extract for research, private study or in a noncommercial document with limited circulation (such as an **academic thesis or dissertation**), you may do so **without seeking permission**. Our only requirement is that the WHO source should be appropriately acknowledged.

I hereby thank the WHO for the permission to use website and publication contents. Wherever a provided content was reused in this thesis a proper citation was attached.

Figure 1:

Centers for Disease Control and Prevention – Public Health Image Library (PHIL)

Picture ID: 3192

Content Providers CDC / Dr. William Kaplan

Copyright restrictions: **None** - This image is in the public domain and thus free of any copyright restrictions. As a matter of courtesy we request that the content provider be credited and notified in any public or private usage of this image.

Figure 2:

Centers for Disease Control and Prevention – Public Health Image Library (PHIL)

Picture ID: 1215

Content Providers CDC / American Society of Clinical Pathologists

Any public or private use of this image is subject to prevailing copyright laws. Please contact the content provider of this image for permission requests.

Permission was granted by Mr. Brian Oliver from the ASCP on October 1st 2014 (see below)

Von: Oliver, Brian <Brian.Oliver@ascp.org>

An: christin.schaefer@virchow.uni-wuerzburg.de <christin.schaefer@virchow.uni-wuerzburg.de>

Betreff: RE: Permission for using an ASCP image in my PhD thesis

Hi Christin,

We have processed your request and we would like to extend permission to you so you can complete your thesis. This

Copyright

leads me to my question for you. Do you need a formal permission letter or will this email suffice for your committee? If not, that is not a problem. I will then draft a letter noting our conditions -- it would be the same conditions you mentioned in your request.

If you do not need the letter, consider my email as an early congratulations for completing your thesis. Best of luck! I hope you do not have to defend it, however if you do, it will go fine.

Best,
Brian

Brian Oliver, MLIS
Coordinator-Archives, Permissions and Rights
American Society for Clinical Pathology
33 West Monroe, Suite 1600
Chicago, Illinois 60603
Tele: 312-541-4745 Email: brian.oliver@ascp.org

-----Original Message-----

From: christin.schaefer@virchow.uni-wuerzburg.de
[mailto:christin.schaefer@virchow.uni-wuerzburg.de]
Sent: Thursday, September 25, 2014 4:11 PM
To: Oliver, Brian
Subject: Permission for using an ASCP image in my PhD thesis

Dear Mr. Oliver,

I was informed to contact you, to ask for the permission to use one of the ASCP images in my PhD thesis. I am a trained biochemist and I am working in the lab of Prof. Dr. Caroline Kisker at the Rudolf Virchow Center for Experimental Biomedicine in Würzburg, Germany. Currently I am writing my thesis with the working title 'Structural and functional characterization of potential new drug targets from human pathogens'. One of the investigated proteins is a *C. albicans* enzyme and I would like to include one of your pictures as descriptive depiction for the morphological dimorphism of the fungus into my introduction. The image is currently deposited on the PHIL homepage and is assigned to number #1215 with the description 'Histopathology of *Candida albicans* infection. Methenamine silver stain.' Could you provide the permission for using the image in my thesis? I would of course credit and notify the ASCP in the figure legend, the reference list and the acknowledgments for providing the picture. In case of any questions, please do not hesitate to contact me.
Thank you for your help.

Sincerely
Christin Schäfer

Figure 7:

Centers for Disease Control and Prevention – Public Health Image Library (PHIL)

Picture ID: 16881

Content Providers CDC / Melissa Brower

Copyright restrictions: **None** - This image is in the public domain and thus free of any copyright restrictions. As a matter of courtesy we request that the content provider be credited and notified in any public or private usage of this image.

Figure 8:

Centers for Disease Control and Prevention – Public Health Image Library (PHIL)

Picture ID: 1428

Content Providers CDC / Dr. George Kubica

Copyright restrictions: **None** - This image is in the public domain and thus free of any copyright restrictions. As a matter of courtesy we request that the content provider be credited and notified in any public or private usage of this image.

Figure 9:

Copyright

Van der Geize et al., 2007

Requests for permission should be made in writing. For the fastest response time, please send your request via e-mail to PNASPermissions@nas.edu. If necessary, requests may be faxed to 202-334-2739 or mailed to:

PNAS Permissions Editor
500 Fifth Street, NW
NAS 340
Washington, DC 20001 USA

Anyone may, without requesting permission, use original figures or tables published in PNAS for noncommercial and educational use (i.e., in a review article, in a book that is not for sale) provided that the original source and the applicable copyright notice are cited.

Figure 10 / 11:

Both figures are extracted from the publication (Schaefer et al.), where the author of this thesis is author, too.
<http://www.cell.com/structure/authors#prepub>

Author's rights:

As an author, you (or your employer or institution) may do the following:

- **Include the article in full or in part in a thesis or dissertation (provided that this is not to be published commercially)**

Figure 17:

ELSEVIER LICENSE
TERMS AND CONDITIONS
Oct 09, 2014

This is a License Agreement between **Christin Schäfer** ("You") and **Elsevier** ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, UK
Registered Company Number	1982084
Customer name	Christin Schäfer
Customer address	Rudolf Virchow Center Würzburg, 97080
License number	3484841245983
License date	Oct 09, 2014
Licensed content publisher	Elsevier
Licensed content publication	Current Opinion in Structural Biology
Licensed content title	Turning protein crystallisation from an art into a science
Licensed content author	Naomi E Chayen
Licensed content date	October 2004
Licensed content volume number	14
Licensed content issue number	5
Number of pages	7
Start Page	577
End Page	583
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic

Copyright

Are you the author of this Elsevier article?	No
Will you be translating?	No
Title of your thesis/dissertation	Approaching antimicrobial resistance - Structural and functional characterization of the fungal transcription factor Mrr1 from <i>Candida albicans</i> and the bacterial β -ketoacyl-CoA thiolase FadA5 from <i>Mycobacterium tuberculosis</i> .
Expected completion date	Oct 2014
Estimated size (number of pages)	200
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
VAT/Local Sales Tax	0.00 USD / 0.00 GBP
Total	0.00 USD

8 Appendix

8.1 Supplemental Figures

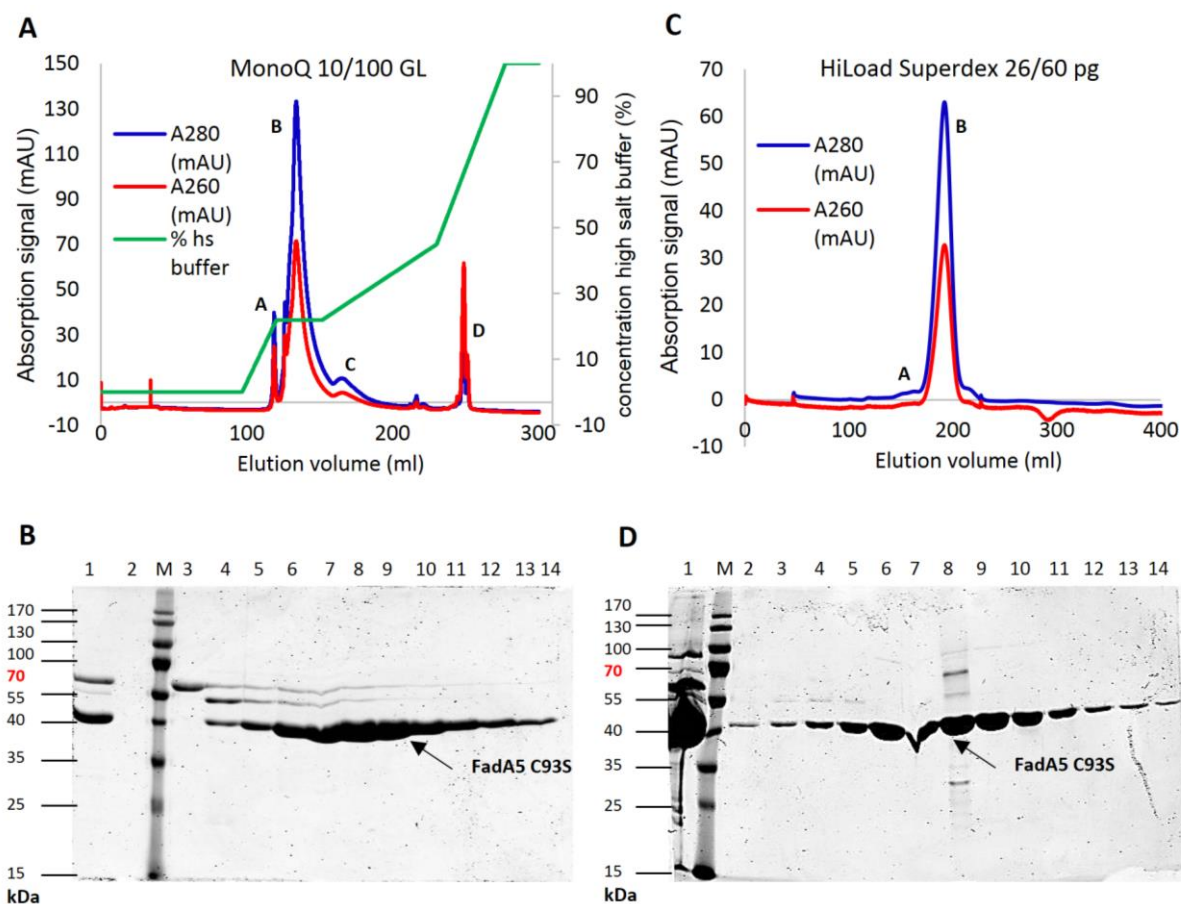


Figure S 1 - Purification of FadA5 C93S:

A) Chromatogram of the anion-exchange chromatography step: absorption at 280 nm (blue) and 260 nm (red); % high salt buffer (green) **B)** 15% SDS-PAGE of the anion-exchange step: M = marker, 1 = load; 2 = wash unbound sample; 3 = peak A; 4/5 = fractions between peak A and B; 6-13 = elution fractions peak B; 14 = elution fractions peak C; peak D not analyzed **C)** Chromatogram of the size-exclusion chromatography step: absorption at 280 nm (blue) and 260 nm (red) **D)** 15% SDS-PAGE of the size-exclusion step: M = marker; 1 = load, 2-4 = elution fractions peak A; 5-14 = elution fractions peak B with tailing

Appendix

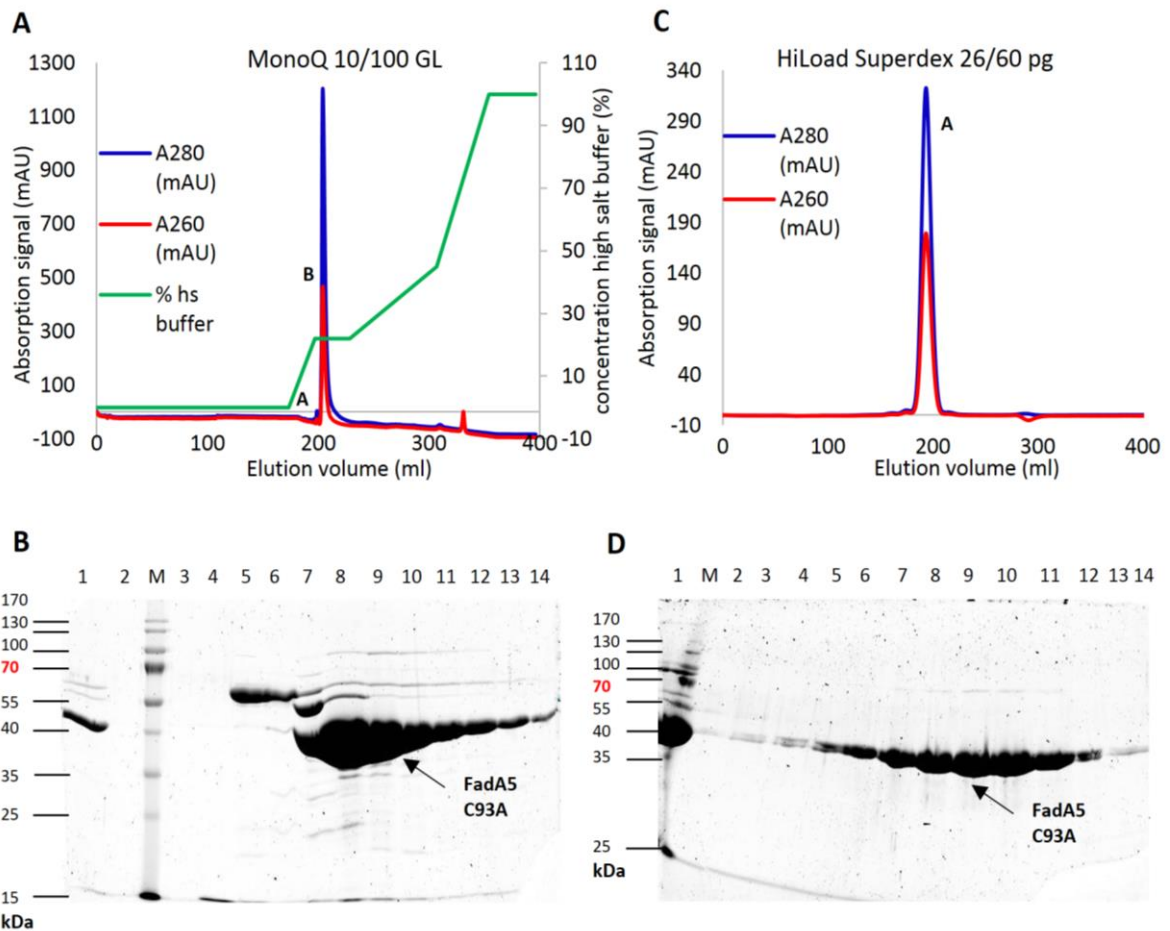


Figure S 2 - Purification of FadA5 C93A:

A) Chromatogram of the anion-exchange chromatography step: absorption at 280 nm (blue) and 260 nm (red); % high salt buffer (green) **B)** 15% SDS-PAGE of the anion-exchange step: M = marker; 1 = load; 2 = flow through; 3 = wash unbound sample; 4-7 = elution fractions peak A; 8-14 = elution fractions peak B **C)** Chromatogram of the size-exclusion chromatography step: absorption at 280 nm (blue) and 260 nm (red) **D)** 15% SDS-PAGE of the size-exclusion step: M = marker, 1 = load, 2-4 = elution fractions before peak A, 5-14 = elution fractions peak A

Appendix

8.2 Supplemental Tables

Table S 1 - Used primers

No	primer name	sequence (5'-->3')	construct	organism	DNA sequence	amino acid sequence
1	vector3' Insert for	gacaagcttgcgccgcactcgag	pBadm11 FL	<i>E. coli</i>	1 - 3324	M1 - N1108
2	vector5' Insert rev N-term	ggcgccctgaaaataaagattctc				
3	FL1 N-term 5' for	ctttattttcaggcgccatgagtatt gcaaccaccccgattgaaacc				
4	FL5 N-term 3' rev	gtcggccgcaagctgtcttagttgg taaacagctgatcaaacggcagatc				
1	vector 3' Insert for	gacaagcttgcgccgcactcgag	pBadm11 AR	<i>E. coli</i>	891 - 3099	T298 - T1033
2	vector 5' Insert rev N-term	ggcgccctgaaaataaagattctc				
5	AR5 Armadillo Repeat 3' N-termHis	aatctttattttcaggcgccaccagc atttgcgtgctgatttctgctgt				
6	AR6 Armadillo Repeat 5' N-terminalHis	gtcggccgcaagctgtcttaggtaa acggctggcttcttctgctggcgtg				
7	aa250fwd	catgccatggcaagcgaatgagccct	pBadm11 '250'	<i>E. coli</i>	751 - 3324	S250 - N1108
8	aa250rev	gggtaccttagttgtaaacag				
9	pPinkalpha-HC rev	ggccttttctcgagagatacccc	pPinkαHC FL	<i>P. pastoris</i>	1 - 3324	M1 - N1108
10	pPinkalpha-HC for	ttaaatacaggcccttttcttctgtcg				
11	His6-Tag	ggtatctctcgagaaaaggcctatgaa acatcaccatcaccatcacccc				
12	revFL_250	ggaaaagggcctgtatattaattagttg gtaaacagctgatcaaacggcag				
13	AR_del_1_fwd	caaccagcatttgcgtgctg	pPinkαHC ARlong	<i>P. pastoris</i>	891 - 3324	T298 - N1108
14	deletion rev	@-ccatggcgccctgaaaataa				
14	deletion rev	@-ccatggcgccctgaaaataa	pPinkαHC '250'	<i>P. pastoris</i>	751 - 3324	S250 - N1108
15	deletion forw	@-caagcgaatgagccctccg				
16	pPink-H/LC rev	cgg ttt cga ata att agt tg	pPinkHC/LC FL	<i>P. pastoris</i>	1 - 3324	M1 - N1108
17	pPink-H/LC fwd	cgg cgg gcc att taa ata cag gc				
11	His6-Tag	ggtatctctcgagaaaaggcctatgaa a catcaccatcaccatcacccc				
12	revFL_250	ggaaaagggcctgtatattaattagttg g taaacagctgatcaaacggcag				
18	G997V_fwd	ccaatgattttgttctggacctggtg	pBadm11 '250' G997V	<i>E. coli</i>	751 - 3324	S250 - N1108
19	G997V_rev	caccaggtccagaaacaaatcattgg				
20	P683S_fwd	ccaaaagcgtttttcgggttacttcagc	pBadm11 '250' P683S	<i>E. coli</i>	751 - 3324	S250 - N1108
21	P683S_rev	gctgaagtaaccgaaaaaacgctttt gg				
22	30-70_fus_fwd	ttattttcaggcgccatggcaaccagc a tttgcgtgctg	pPinkHC/LC ARlong	<i>P. pastoris</i>	891 - 3324	T298 - N1108
23	30-70_fus_rev	acgcacgcaaatgctggttccatggc gccctgaaaataa				
14	deletion rev	@-ccatggcgccctgaaaataa				
13	AR_del_1_fwd	caaccagcatttgcgtgctg				
24	30-31_fus_fwd	ttattttcaggcgccatggcaagcgaa atgagccctccg	pPinkHC/LC '250'	<i>P. pastoris</i>	751 - 3324	S250 - N1108
25	30-31_fus_rev	cggagggctcatttgccttggcattggc cctgaaaataa				
14	deletion rev	@-ccatggcgccctgaaaataa				
15	deletion forw	@-caagcgaatgagccctccg				
26	G878E_fwd_2	tgcagaagttgagattgcagcagtttag c	pBadm11 '250'	<i>E. coli</i>	751 - 3324	S250 - N1108

Appendix

27	G878E_rev_2	caacacgtcttcaactctaacgtcgtca	G878E			
28	K335N_fwd_2	catatcctgaaaaagaacgagaagaa g aac	pBadm11 '250' K335N	<i>E. coli</i>	751 - 3324	S250 - N1108
29	K335N_rev_2	ggacttttctgtccttcttcttgaacg				
30	pETm11_R	cat ggt ata tct cct tct taa agt taa atc				
31	pETm11_F	gtc gac aag ctt gcg g				
32	FL_Mrr1_R	tgcgccgcaagcttgcgacgttggtgta aacagctgatcaaac	pETm11 '250' (CT)	<i>E. coli</i>	751 - 3324	S250 - N1108
33	Mrr1_250_F	ttaagaaggagatataccatgagcgaa a tgagccc				

Table S 2 - Composition and list of all used buffers during this work

Protein / Organism	No	Purification technique	Buffer	Composition	Annotations
FadA5 WT and variants <i>Mycobacterium smegmatis</i>	1	Affinity chromatography Co-Talon resin	Binding buffer	50 mM Tris-HCl pH 7.4 300 mM NaCl 10 mM imidazole 10% (v/v) glycerol	---
	2		Elution buffer	50 mM Tris-HCl pH 7.4 300 mM NaCl 150 mM imidazole 10% (v/v) glycerol	---
	3	Anion-exchange chromatography MonoQ (MQ) 10/100 GL	MQ buffer low salt	50 mM Tris-HCl pH 7.6 50 mM NaCl	sterile /degased
	4		MQ buffer high salt	50 mM Tris-HCl pH 7.6 1.0 M NaCl	sterile /degased
	5	Size-exclusion (SE) chromatography HiLoad 26/60 Superdex 200 pg	SE buffer	20 mM Bicine-NaOH pH 8.5 250 mM NaCl	sterile /degased
5o	SE buffer (old)		60 mM Bicine-NaOH pH 8.5 250 mM NaCl	sterile /degased	
Mrr1 Full length and Armadillo Repeat constructs <i>Escherichia coli</i>	6	Affinity chromatography Co-Talon, Ni-IDA, Ni-NTA, Ni-TED resins	Binding buffer	100 mM Tris-HCl pH 8.0, 500 mM NaCl 5 mM imidazole	
	7		Wash buffers	Binding buffer with 15, 45, 60, 75, 100 mM imidazole	
	8		Wash buffer special	Binding buffer with 10 mM imidazole 2 mM ATP 10 mM MgCl ₂	
	9		Elution buffers	Binding buffer with 150 / 250 mM imidazole	
	10		Binding buffer	100 mM Pipes-NaOH pH 7.5 300 mM NaCl 5 mM imidazole +/- 10% (v/v) glycerol	
	11		Wash buffers	Binding buffer with 75 / 100 mM imidazole	
	12		Elution buffers	Binding buffer with 150 / 250 mM imidazole	
	13		Binding buffer	50 mM NaPhosphate pH 8.0 500 mM NaCl 10 mM imidazole	sterile /degased
	14		Elution buffer	Binding buffer with 500 mM imidazole	sterile /degased
	15		Binding buffer	100 mM Tris-HCl pH 8.0, 500 mM NaCl 5 mM imidazole 5 mM dithiotreitol (DTT)	
16	Wash buffers	Binding buffer with 30, 60, 75 mM imidazole			
17	Elution buffers	Binding buffer with 150 / 250 mM imidazole			
Mrr1 250 variants with N-terminal His ₆ -Tag <i>Escherichia coli</i>	18	Affinity chromatography Ni-MAC (1 ml) / HisTrap FF crude	Binding buffer	40 mM NaPhosphate pH 8.0 500 mM NaCl 10 mM imidazole	sterile /degased
	19		Elution buffer	Binding buffer with 500 mM imidazole	sterile /degased
	20	Affinity chromatography Ni-IDA, Ni-NTA, Ni-TED resin	Wash buffers	Binding buffer with 25, 35, 45, 50, 75, 100 mM	

Appendix

	V 21 I / II		Elution buffers	imidazole	
			Binding buffer with	150, 250 mM imidazole	
			Binding buffer	40 mM NaPhosphate pH 8.0 500 mM KCl 10 mM imidazole	
			Wash buffers	Binding buffer with 25, 35, 45, 50, 75, 100 mM imidazole	
	23 I - VI 24 I / II	Affinity chromatography Ni-NTA resin	Elution buffers	Binding buffer with 150, 250 mM imidazole	
			Binding buffer	20 mM Tris-HCl pH 8.0 300 mM KCl 5 mM imidazole	
			Wash buffers	Binding buffer with 25, 50, 75, 100 mM imidazole	
			Elution buffers	Binding buffer with 100, 250 mM imidazole	
Mrr1 250 with C-terminal His ₆ -Tag <i>E. coli</i>	42		Binding buffer	20 mM Tris-HCl pH 8.0 300 mM KCl 5 mM imidazole	
			Wash buffers	Binding buffer with 25, 50, 75, 100 mM imidazole	
			Elution buffers	Binding buffer with 100, 250 mM imidazole	
			Binding buffer	50 mM Tris-HCl pH 7.8 50 mM NaCl	sterile /degased
Mrr1 250 variant <i>Escherichia coli</i>	25 26 27 28 29 30 31 32	Anion-exchange chromatography MonoQ (MQ) 5/50 GL	MQ buffer low salt	50 mM Tris-HCl pH 7.8 50 mM NaCl	sterile /degased
			MQ buffer high salt	50 mM Tris-HCl pH 7.8 1 M NaCl	sterile /degased
			MQ buffer low salt	50 mM Hepes-NaOH pH 8.0 50 mM KCl	sterile /degased
			MQ buffer high salt	50 mM Hepes-NaOH pH 8.0 1 M KCl	sterile /degased
			MQ buffer low salt	50 mM Bicine-NaOH pH 8.5 50 mM KCl	sterile /degased
			MQ buffer high salt	50 mM Bicine-NaOH pH 8.5 1 M KCl	sterile /degased
			MQ buffer low salt	50 mM Bicine-NaOH pH 9.0 50 mM KCl	sterile /degased
			MQ buffer high salt	50 mM Bicine-NaOH pH 9.0 1 M KCl	sterile /degased
Mrr1 <i>Escherichia coli</i>	FL AR 250 FL 250 FL FL 250	Size-exclusion (SE) chromatography Superdex 200 columns different sizes	SE buffer	100 mM Tris-HCl pH 8.0 300 mM NaCl	sterile /degased
			SE buffer	50 mM ADA-NaOH pH 7.0 300 mM NaCl	sterile /degased
			SE buffer	50 mM Tris-HCl pH 8.0 300 mM NaCl 5 mM DTT	sterile /degased
			SE buffer	50 mM Tris-HCl pH 8.0 200 mM NaCl 5 mM DTT	sterile /degased
			SE buffer	20 mM Bicine-NaOH pH 8.0 300 mM KCl	sterile /degased
			TEV buffer	50 mM Tris-HCl pH 8.0 0.5 mM EDTA 1 mM DTT	
Mrr1 250 variants <i>E. coli</i>	38	TEV cleavage			
Mrr1 <i>Pichia pastoris</i>	39	Affinity chromatography Ni-NTA resin (Expression test)	Breaking buffer	50 mM NaPhosphate pH 7.4 1 mM PMSF 1 mM EDTA 5% (v/v) glycerol	Recipe Invitrogen PichiaPink manual
Mrr1 250 variants <i>E. coli</i>	40	CD spectroscopy	CD buffer	20 mM Bicine-NaOH pH 8.0 300 mM NaF	
FadA5 variants <i>M. smegmatis</i>	41	Isothermal titration calorimetry (ITC)	ITC buffer	50 mM MES pH 6.7 250 mM NaCl	

Appendix

Table S 3 - Composition of the standard thermofluor screen

<i>well</i>	<i>condition</i>	<i>well</i>	<i>condition</i>
A1	citrate pH 4.5	E1	cacodylate pH 6.0
A2	bis-tris pH 7.0	E2	bis-tris propan pH 7.0
A3	imidazole pH 6.5	E3	Mops pH 7.0
A4	Hepes pH 8.0	E4	Bicine-NaOH pH 9.0
A5	Tris-HCl pH 8.5	E5	glycylglycine pH 8.5
A6	reference (current sample buffer)	E6	negative sample (water)
A7	citrate pH 4.5, 0.3 M NaCl	E7	cacodylate pH 6.0, 0.3 M NaCl
A8	bis-tris pH 7.0, 0.3 M NaCl	E8	bis-tris propan pH 7.0, 0.3 M NaCl
A9	imidazole pH 6.5, 0.3 M NaCl	E9	Mops pH 7.0, 0.3 M NaCl
A10	Hepes pH 8.0, 0.3 M NaCl	E10	Bicine-NaOH pH 9.0, 0.3 M NaCl
A11	Tris-HCl pH 8.5, 0.3 M NaCl	E11	glycylglycine pH 8.5, 0.3 M NaCl
A12	reference (current sample buffer)	E12	negative sample (water)
B1	acetate pH 4.6	F1	cacodylate pH 6.5
B2	ADA pH 6.5	F2	Pipes pH 6.5
B3	imidazole pH 8.0	F3	Mops pH 7.5
B4	Hepes pH 8.5	F4	Tris-HCl pH 7.0
B5	Tris-HCl pH 9.0	F5	Ches pH 9.0
B6	reference (current sample buffer)	F6	negative sample (water)
B7	acetate pH 4.6, 0.3 M NaCl	F7	cacodylate pH 6.5, 0.3 M NaCl
B8	ADA pH 6.5, 0.3 M NaCl	F8	Pipes pH 6.5, 0.3 M NaCl
B9	imidazole pH 8.0, 0.3 M NaCl	F9	Mops pH 7.5, 0.3 M NaCl
B10	Hepes pH 8.5, 0.3 M NaCl	F10	Tris-HCl pH 7.0, 0.3 M NaCl
B11	Tris-HCl pH 9.0, 0.3 M NaCl	F11	Ches pH 9.0, 0.3 M NaCl
B12	reference (current sample buffer)	F12	negative sample (water)
C1	MES pH 5.5	G1	bis-tris pH 5.5
C2	ADA pH 7.0	G2	Pipes pH 7.0
C3	Na/K PO ₄ pH 6.8	G3	Hepes pH 7.0
C4	Bicine-NaOH pH 8.0	G4	Tris-HCl pH 7.5
C5	Taps pH 8.0	G5	Ches pH 9.5
C6	reference (current sample buffer)	G6	negative sample (water)
C7	MES pH 5.5, 0.3 M NaCl	G7	bis-tris pH 5.5, 0.3 M NaCl
C8	ADA pH 7.0, 0.3 M NaCl	G8	Pipes pH 7.0, 0.3 M NaCl
C9	Na/K PO ₄ pH 6.8, 0.3 M NaCl	G9	Hepes pH 7.0, 0.3 M NaCl
C10	Bicine-NaOH pH 8.0, 0.3 M NaCl	G10	Tris-HCl pH 7.5, 0.3 M NaCl
C11	Taps pH 8.0, 0.3 M NaCl	G11	Ches pH 9.5, 0.3 M NaCl
C12	reference (current sample buffer)	G12	negative sample (water)
D1	MES pH 6.5	H1	bis-tris pH 6.5
D2	bis-tris propan pH 6.0	H2	Pipes pH 7.5
D3	Na/K-PO ₄ pH 7.55	H3	Hepes pH 7.5
D4	Bicine-NaOH pH 8.5	H4	Tris-HCl pH 8.0
D5	Taps pH 9.0	H5	Caps pH 9.8
D6	reference (current sample buffer)	H6	negative sample (water)
D7	MES pH 6.5, 0.3 M NaCl	H7	bis-tris pH 6.5, 0.3 M NaCl
D8	bis-tris propan pH 6.0, 0.3 M NaCl	H8	Pipes pH 7.5, 0.3 M NaCl
D9	Na/K-PO ₄ pH 7.55, 0.3 M NaCl	H9	Hepes pH 7.5, 0.3 M NaCl
D10	Bicine-NaOH pH 8.5, 0.3 M NaCl	H10	Tris-HCl pH 8.0, 0.3 M NaCl
D11	Taps pH 9.0, 0.3 M NaCl	H11	Caps pH 9.8, 0.3 M NaCl
D12	reference (current sample buffer)	H12	negative sample (water)

Appendix

Table S 4 - Composition of the FadA5 crystallization screen

Well	conc.	buffer	conc.	precipitant	conc.	precipitant
A1	0.1 M	MES pH 6.5	1.0 M	Ammonium sulfate	4.0%	1,4-dioxane
A2	0.1 M	MES pH 6.5	1.2 M	Ammonium sulfate	4.0%	1,4-dioxane
A3	0.1 M	MES pH 6.5	1.4 M	Ammonium sulfate	4.0%	1,4-dioxane
A4	0.1 M	MES pH 6.5	1.6 M	Ammonium sulfate	4.0%	1,4-dioxane
A5	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	4.0%	1,4-dioxane
A6	0.1 M	MES pH 6.5	2.0 M	Ammonium sulfate	4.0%	1,4-dioxane
A7	0.1 M	MES pH 5.5	1.0 M	Ammonium sulfate		
A8	0.1 M	MES pH 5.5	1.2 M	Ammonium sulfate		
A9	0.1 M	MES pH 5.5	1.4 M	Ammonium sulfate		
A10	0.1 M	MES pH 5.5	1.6 M	Ammonium sulfate		
A11	0.1 M	MES pH 5.5	1.8 M	Ammonium sulfate		
A12	0.1 M	MES pH 5.5	2.0 M	Ammonium sulfate		
B1	0.1 M	MES pH 6.5	1.0 M	Ammonium sulfate	7.0%	1,4-dioxane
B2	0.1 M	MES pH 6.5	1.2 M	Ammonium sulfate	7.0%	1,4-dioxane
B3	0.1 M	MES pH 6.5	1.4 M	Ammonium sulfate	7.0%	1,4-dioxane
B4	0.1 M	MES pH 6.5	1.6 M	Ammonium sulfate	7.0%	1,4-dioxane
B5	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	7.0%	1,4-dioxane
B6	0.1 M	MES pH 6.5	2.0 M	Ammonium sulfate	7.0%	1,4-dioxane
B7	0.1 M	MES pH 6.0	1.0 M	Ammonium sulfate		
B8	0.1 M	MES pH 6.0	1.2 M	Ammonium sulfate		
B9	0.1 M	MES pH 6.0	1.4 M	Ammonium sulfate		
B10	0.1 M	MES pH 6.0	1.6 M	Ammonium sulfate		
B11	0.1 M	MES pH 6.0	1.8 M	Ammonium sulfate		
B12	0.1 M	MES pH 6.0	2.0 M	Ammonium sulfate		
C1	0.1 M	MES pH 6.5	1.0 M	Ammonium sulfate	10.0%	1,4-dioxane
C2	0.1 M	MES pH 6.5	1.2 M	Ammonium sulfate	10.0%	1,4-dioxane
C3	0.1 M	MES pH 6.5	1.4 M	Ammonium sulfate	10.0%	1,4-dioxane
C4	0.1 M	MES pH 6.5	1.6 M	Ammonium sulfate	10.0%	1,4-dioxane
C5	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	10.0%	1,4-dioxane
C6	0.1 M	MES pH 6.5	2.0 M	Ammonium sulfate	10.0%	1,4-dioxane
C7	0.1 M	MES pH 6.5	1.0 M	Ammonium sulfate		
C8	0.1 M	MES pH 6.5	1.2 M	Ammonium sulfate		
C9	0.1 M	MES pH 6.5	1.4 M	Ammonium sulfate		
C10	0.1 M	MES pH 6.5	1.6 M	Ammonium sulfate		
C11	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate		
C12	0.1 M	MES pH 6.5	2.0 M	Ammonium sulfate		
D1	0.1 M	MES pH 6.5	1.0 M	Ammonium sulfate	13.0%	1,4-dioxane
D2	0.1 M	MES pH 6.5	1.2 M	Ammonium sulfate	13.0%	1,4-dioxane
D3	0.1 M	MES pH 6.5	1.4 M	Ammonium sulfate	13.0%	1,4-dioxane
D4	0.1 M	MES pH 6.5	1.6 M	Ammonium sulfate	13.0%	1,4-dioxane
D5	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	13.0%	1,4-dioxane
D6	0.1 M	MES pH 6.5	2.0 M	Ammonium sulfate	13.0%	1,4-dioxane
D7	0.1 M	HEPES pH 7	1.0 M	Ammonium sulfate		
D8	0.1 M	HEPES pH 7	1.2 M	Ammonium sulfate		
D9	0.1 M	HEPES pH 7	1.4 M	Ammonium sulfate		
D10	0.1 M	HEPES pH 7	1.6 M	Ammonium sulfate		
D11	0.1 M	HEPES pH 7	1.8 M	Ammonium sulfate		
D12	0.1 M	HEPES pH 7	2.0 M	Ammonium sulfate		
E1	0.1 M	MES pH 6.5	1.6 M	Ammonium sulfate	4.0%	1,4-dioxane
E1					2.5%	DMSO
E2	0.1 M	MES pH 6.5	1.6 M	Ammonium sulfate	4.0%	1,4-dioxane
E2					5.0%	DMSO
E3	0.1 M	MES pH 6.5	1.6 M	Ammonium sulfate	4.0%	1,4-dioxane
E3					7.5%	DMSO
E4	0.1 M	MES pH 6.5	1.6 M	Ammonium sulfate	4.0%	1,4-dioxane
E4					10.0%	DMSO
E5	0.1 M	MES pH 6.5	1.6 M	Ammonium sulfate	4.0%	1,4-dioxane
E5					15.0%	DMSO
E6	0.1 M	MES pH 6.5	1.6 M	Ammonium sulfate	4.0%	1,4-dioxane
E6					20.0%	DMSO
E7	0.1 M	MES pH 6.5	1.6 M	Ammonium sulfate	4.0%	1,4-dioxane
E7					2.5%	DMSO
E8	0.1 M	MES pH 6.5	1.6 M	Ammonium sulfate	4.0%	1,4-dioxane

Appendix

<i>E8</i>					5.0%	DMSO
<i>E9</i>	0.1 M	MES pH 6.5	1.6 M	Ammonium sulfate	4.0%	1,4-dioxane
<i>E9</i>					7.5%	DMSO
<i>E10</i>	0.1 M	MES pH 6.5	1.6 M	Ammonium sulfate	4.0%	1,4-dioxane
<i>E10</i>					10.0%	DMSO
<i>E11</i>	0.1 M	MES pH 6.5	1.6 M	Ammonium sulfate	4.0%	1,4-dioxane
<i>E11</i>					15.0%	DMSO
<i>E12</i>	0.1 M	MES pH 6.5	1.6 M	Ammonium sulfate	4.0%	1,4-dioxane
<i>E12</i>					20.0%	DMSO
<i>F1</i>	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	4.0%	1,4-dioxane
<i>F1</i>					2.5%	DMSO
<i>F2</i>	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	4.0%	1,4-dioxane
<i>F2</i>					5.0%	DMSO
<i>F3</i>	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	4.0%	1,4-dioxane
<i>F3</i>					7.5%	DMSO
<i>F4</i>	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	4.0%	1,4-dioxane
<i>F4</i>					10.0%	DMSO
<i>F5</i>	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	4.0%	1,4-dioxane
<i>F5</i>					15.0%	DMSO
<i>F6</i>	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	4.0%	1,4-dioxane
<i>F6</i>					20.0%	DMSO
<i>F7</i>	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	4.0%	1,4-dioxane
<i>F7</i>					2.5%	DMSO
<i>F8</i>	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	4.0%	1,4-dioxane
<i>F8</i>					5.0%	DMSO
<i>F9</i>	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	4.0%	1,4-dioxane
<i>F9</i>					7.5%	DMSO
<i>F10</i>	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	4.0%	1,4-dioxane
<i>F10</i>					10.0%	DMSO
<i>F11</i>	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	4.0%	1,4-dioxane
<i>F11</i>					15.0%	DMSO
<i>F12</i>	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	4.0%	1,4-dioxane
<i>F12</i>					20.0%	DMSO
<i>G1</i>	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	7.0%	1,4-dioxane
<i>G2</i>	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	7.0%	1,4-dioxane
<i>G3</i>	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	7.0%	1,4-dioxane
<i>G4</i>	0.1 M	MES pH 6.5	2.0 M	Ammonium sulfate	7.0%	1,4-dioxane
<i>G5</i>	0.1 M	MES pH 6.5	2.0 M	Ammonium sulfate	7.0%	1,4-dioxane
<i>G6</i>	0.1 M	MES pH 6.5	2.0 M	Ammonium sulfate	7.0%	1,4-dioxane
<i>G7</i>	0.1 M	MES pH 6.5	2.0 M	Ammonium sulfate	7.0%	1,4-dioxane
<i>G8</i>	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	7.0%	1,4-dioxane
<i>G9</i>	0.1 M	MES pH 6.5	1.8 M	Ammonium sulfate	7.0%	1,4-dioxane
<i>G10</i>	0.1 M	MES pH 6.5	2.0 M	Ammonium sulfate	7.0%	1,4-dioxane
<i>G11</i>	0.1 M	MES pH 6.5	2.0 M	Ammonium sulfate	7.0%	1,4-dioxane
<i>G12</i>	0.1 M	MES pH 6.5	2.0 M	Ammonium sulfate	7.0%	1,4-dioxane
<i>H1</i>	0.1 M	MES pH 6.0	1.8 M	Ammonium sulfate		DMSO
<i>H2</i>	0.1 M	MES pH 6.0	1.8 M	Ammonium sulfate	2.5%	DMSO
<i>H3</i>	0.1 M	MES pH 6.0	1.8 M	Ammonium sulfate	5.0%	DMSO
<i>H4</i>	0.1 M	MES pH 6.0	2.0 M	Ammonium sulfate		DMSO
<i>H5</i>	0.1 M	MES pH 6.0	2.0 M	Ammonium sulfate	2.5%	DMSO
<i>H6</i>	0.1 M	MES pH 6.0	2.0 M	Ammonium sulfate	5.0%	DMSO
<i>H7</i>	0.1 M	MES pH 6.0	1.8 M	Ammonium sulfate		DMSO
<i>H8</i>	0.1 M	MES pH 6.0	1.8 M	Ammonium sulfate	2.5%	DMSO
<i>H9</i>	0.1 M	MES pH 6.0	1.8 M	Ammonium sulfate	5.0%	DMSO
<i>H10</i>	0.1 M	MES pH 6.0	2.0 M	Ammonium sulfate		DMSO
<i>H11</i>	0.1 M	MES pH 6.0	2.0 M	Ammonium sulfate	2.5%	DMSO
<i>H12</i>	0.1 M	MES pH 6.0	2.0 M	Ammonium sulfate	5.0%	DMSO

Appendix

Table S 5 - Composition of the Thiolase crystallization screen

Well	conc.	buffer	conc.	precipitate	conc.	precipitate
A1	0.1 M	Sodium acetate pH 5.0	0.7 M	Lithium sulfate	0.7 M	Ammonium sulfate
A2	0.1 M	Sodium acetate pH 5.0	0.8 M	Lithium sulfate	0.7 M	Ammonium sulfate
A3	0.1 M	Sodium acetate pH 5.0	0.9 M	Lithium sulfate	0.7 M	Ammonium sulfate
A4	0.1 M	Sodium acetate pH 5.0	1.1 M	Lithium sulfate	0.7 M	Ammonium sulfate
A5	0.1 M	Sodium acetate pH 5.0	1.2 M	Lithium sulfate	0.7 M	Ammonium sulfate
A6	0.1 M	Sodium acetate pH 5.0	1.3 M	Lithium sulfate	0.7 M	Ammonium sulfate
A7	0.1 M	Sodium citrate pH 5.5		Lithium sulfate	1.2 M	Ammonium sulfate
A8	0.1 M	Sodium citrate pH 5.5		Lithium sulfate	1.4 M	Ammonium sulfate
A9	0.1 M	Sodium citrate pH 5.5		Lithium sulfate	1.5 M	Ammonium sulfate
A10	0.1 M	Sodium citrate pH 5.5		Lithium sulfate	1.6 M	Ammonium sulfate
A11	0.1 M	Sodium citrate pH 5.5		Lithium sulfate	1.8 M	Ammonium sulfate
A12	0.1 M	Sodium citrate pH 5.5		Lithium sulfate	1.9 M	Ammonium sulfate
B1	0.1 M	Sodium acetate pH 5.0	0.7 M	Lithium sulfate	0.8 M	Ammonium sulfate
B2	0.1 M	Sodium acetate pH 5.0	0.8 M	Lithium sulfate	0.8 M	Ammonium sulfate
B3	0.1 M	Sodium acetate pH 5.0	0.9 M	Lithium sulfate	0.8 M	Ammonium sulfate
B4	0.1 M	Sodium acetate pH 5.0	1.1 M	Lithium sulfate	0.8 M	Ammonium sulfate
B5	0.1 M	Sodium acetate pH 5.0	1.2 M	Lithium sulfate	0.8 M	Ammonium sulfate
B6	0.1 M	Sodium acetate pH 5.0	1.3 M	Lithium sulfate	0.8 M	Ammonium sulfate
B7	0.1 M	Sodium acetate pH 5.0	0.75 M	Lithium sulfate	1.2 M	Ammonium sulfate
B8	0.1 M	Sodium citrate pH 5.5	0.75 M	Lithium sulfate	1.4 M	Ammonium sulfate
B9	0.1 M	Sodium citrate pH 5.5	0.75 M	Lithium sulfate	1.5 M	Ammonium sulfate
B10	0.1 M	Sodium citrate pH 5.5	0.75 M	Lithium sulfate	1.6 M	Ammonium sulfate
B11	0.1 M	Sodium citrate pH 5.5	0.75 M	Lithium sulfate	1.8 M	Ammonium sulfate
B12	0.1 M	Sodium citrate pH 5.5	0.75 M	Lithium sulfate	1.9 M	Ammonium sulfate
C1	0.1 M	Sodium acetate pH 5.0	0.7 M	Lithium sulfate	0.9 M	Ammonium sulfate
C2	0.1 M	Sodium acetate pH 5.0	0.8 M	Lithium sulfate	0.9 M	Ammonium sulfate
C3	0.1 M	Sodium acetate pH 5.0	0.9 M	Lithium sulfate	0.9 M	Ammonium sulfate
C4	0.1 M	Sodium acetate pH 5.0	1.1 M	Lithium sulfate	0.9 M	Ammonium sulfate
C5	0.1 M	Sodium acetate pH 5.0	1.2 M	Lithium sulfate	0.9 M	Ammonium sulfate
C6	0.1 M	Sodium acetate pH 5.0	1.3 M	Lithium sulfate	0.9 M	Ammonium sulfate
C7	0.1 M	Sodium citrate pH 5.5	0.85 M	Lithium sulfate	1.2 M	Ammonium sulfate
C8	0.1 M	Sodium citrate pH 5.5	0.85 M	Lithium sulfate	1.4 M	Ammonium sulfate
C9	0.1 M	Sodium citrate pH 5.5	0.85 M	Lithium sulfate	1.5 M	Ammonium sulfate
C10	0.1 M	Sodium citrate pH 5.5	0.85 M	Lithium sulfate	1.6 M	Ammonium sulfate
C11	0.1 M	Sodium citrate pH 5.5	0.85 M	Lithium sulfate	1.8 M	Ammonium sulfate
C12	0.1 M	Sodium citrate pH 5.5	0.85 M	Lithium sulfate	1.9 M	Ammonium sulfate
D1	0.1 M	Sodium acetate pH 5.0	0.7 M	Lithium sulfate	1 M	Ammonium sulfate
D2	0.1 M	Sodium acetate pH 5.0	0.8 M	Lithium sulfate	1 M	Ammonium sulfate
D3	0.1 M	Sodium acetate pH 5.0	0.9 M	Lithium sulfate	1 M	Ammonium sulfate
D4	0.1 M	Sodium acetate pH 5.0	1.1 M	Lithium sulfate	1 M	Ammonium sulfate
D5	0.1 M	Sodium acetate pH 5.0	1.2 M	Lithium sulfate	1 M	Ammonium sulfate
D6	0.1 M	Sodium acetate pH 5.0	1.3 M	Lithium sulfate	1 M	Ammonium sulfate
D7	0.1 M	Sodium citrate pH 5.5	0.9 M5	Lithium sulfate	1.2 M	Ammonium sulfate
D8	0.1 M	Sodium citrate pH 5.5	0.9 M5	Lithium sulfate	1.4 M	Ammonium sulfate
D9	0.1 M	Sodium citrate pH 5.5	0.9 M5	Lithium sulfate	1.5 M	Ammonium sulfate
D10	0.1 M	Sodium citrate pH 5.5	0.9 M5	Lithium sulfate	1.6 M	Ammonium sulfate
D11	0.1 M	Sodium citrate pH 5.5	0.9 M5	Lithium sulfate	1.8 M	Ammonium sulfate
D12	0.1 M	Sodium citrate pH 5.5	0.9 M5	Lithium sulfate	1.9 M	Ammonium sulfate
E1	0.1 M	MES pH 6.5	15%	(v/v) PEG 5000 MME		
E2	0.1 M	MES pH 6.5	16%	(v/v) PEG 5000 MME		
E3	0.1 M	MES pH 6.5	17%	(v/v) PEG 5000 MME		
E4	0.1 M	MES pH 6.5	18%	(v/v) PEG 5000 MME		
E5	0.1 M	MES pH 6.5	19%	(v/v) PEG 5000 MME		
E6	0.1 M	MES pH 6.5	20%	(v/v) PEG 5000 MME		
E7	0.1 M	Tris-HCl pH 8.0	1.9 M	Ammonium sulfate		
E8	0.1 M	Tris-HCl pH 8.0	2 M	Ammonium sulfate		
E9	0.1 M	Tris-HCl pH 8.0	2.1 M	Ammonium sulfate		
E10	0.1 M	Tris-HCl pH 8.5	1.9 M	Ammonium sulfate		
E11	0.1 M	Tris-HCl pH 8.5	2 M	Ammonium sulfate		
E12	0.1 M	Tris-HCl pH 8.5	2.1 M	Ammonium sulfate		
F1	0.1 M	HEPES pH 7.5	0.2 M5	Magnesium chloride	23%	(v/v) PEG 3350
F2	0.1 M	HEPES pH 7.5	0.3 M5	Magnesium chloride	23%	(v/v) PEG 3350
F3	0.1 M	HEPES pH 7.5	0.2 M5	Magnesium chloride	25%	(v/v) PEG 3350

Appendix

F4	0.1 M	HEPES pH 7.5	0.3 M5	Magnesium chloride	25%	(v/v) PEG 3350
F5	0.1 M	HEPES pH 7.5	0.2 M5	Magnesium chloride	27%	(v/v) PEG 3350
F6	0.1 M	HEPES pH 7.5	0.3 M5	Magnesium chloride	27%	(v/v) PEG 3350
F7	0.1 M	MES pH 5.6	1.8 M	Ammonium sulfate		
F8	0.1 M	MES pH 5.6	1.9 M	Ammonium sulfate		
F9	0.1 M	MES pH 5.6	2.9 M	Ammonium sulfate		
F10	0.1 M	Tris-HCl pH 7.0	0.1 M	Sodium iodide	20%	(v/v) PEG 3350
F11	0.1 M	Tris-HCl pH 7.0	0.2 M	Sodium iodide	20%	(v/v) PEG 3350
F12	0.1 M	Tris-HCl pH 7.0	0.3 M	Sodium iodide	20%	(v/v) PEG 3350
G1	0.1 M	Tris-HCl pH 8.5	0.1 M5	Ammonium sulfate	23%	(v/v) PEG 4000
G2	0.1 M	Tris-HCl pH 8.5	0.2 M5	Ammonium sulfate	23%	(v/v) PEG 4000
G3	0.1 M	Tris-HCl pH 8.5	0.1 M5	Ammonium sulfate	25%	(v/v) PEG 4000
G4	0.1 M	Tris-HCl pH 8.5	0.2 M5	Ammonium sulfate	25%	(v/v) PEG 4000
G5	0.1 M	Tris-HCl pH 8.5	0.1 M5	Ammonium sulfate	27%	(v/v) PEG 4000
G6	0.1 M	Tris-HCl pH 8.5	0.2 M5	Ammonium sulfate	27%	(v/v) PEG 4000
G7	0.15 M	Lithium acetate	18%	(v/v) PEG 3350		
G8	0.25 M	Lithium acetate	20%	(v/v) PEG 3350		
G9	0.15 M	Lithium acetate	22%	(v/v) PEG 3350		
G10	0.25 M	Lithium acetate	18%	(v/v) PEG 3350		
G11	0.15 M	Lithium acetate	20%	(v/v) PEG 3350		
G12	0.25 M	Lithium acetate	22%	(v/v) PEG 3350		
H1	0.1 M	Tris-HCl pH 8.5	0.05 M	Calcium chloride	19%	(v/v) PEG 4000
H2	0.1 M	Tris-HCl pH 8.5	0.1 M	Calcium chloride	19%	(v/v) PEG 4000
H3	0.1 M	Tris-HCl pH 8.5	0.2 M	Calcium chloride	19%	(v/v) PEG 4000
H4	0.1 M	Tris-HCl pH 8.5	0.05 M	Calcium chloride	21%	(v/v) PEG 4000
H5	0.1 M	Tris-HCl pH 8.5	0.1 M	Calcium chloride	21%	(v/v) PEG 4000
H6	0.1 M	Tris-HCl pH 8.5	0.2 M	Calcium chloride	21%	(v/v) PEG 4000
H7	0.05 M	Sodium acetate pH 5	0.8 M	Lithium sulfate	0.8 M	Ammonium sulfate
H7	0.05 M	Sodium citrate pH 5				
H8	0.05 M	Sodium acetate pH 5	0.8 M	Lithium sulfate	1 M	Ammonium sulfate
H8	0.05 M	Sodium citrate pH 5				
H9	0.05 M	Sodium acetate pH 5	0.9 M	Lithium sulfate	0.8 M	Ammonium sulfate
H9	0.05 M	Sodium citrate pH 5				
H10	0.05 M	Sodium acetate pH 5	0.9 M	Lithium sulfate	1 M	Ammonium sulfate
H10	0.05 M	Sodium citrate pH 5				
H11	0.05 M	Sodium acetate pH 5	1 M	Lithium sulfate	0.8 M	Ammonium sulfate
H11	0.05 M	Sodium citrate pH 5				
H12	0.05 M	Sodium acetate pH 5	1 M	Lithium sulfate	1 M	Ammonium sulfate
H12	0.05 M	Sodium citrate pH 5				

8.3 Abbreviations

1x TBS	Tris-buffered saline	HRP	horseradish peroxidase
1x TBST	Tris-buffered saline with Tween 20	HYD	hydrophobic probe
1x TE	1x Tris-HCl EDTA buffer	IDA	iminodiacetic acid
3,4-DHSA	3,4-dihydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione	IMAC	Immobilized Metal Affinity Chromatography
3-HSA	3-hydroxy-9,10-seconandrost-1,3,5(10)-triene-9,17-dione	IPTG	Isopropyl β-D-1-thiogalactopyranoside
3-ODC-CoA	3-oxo-cholesterol-4,22-diene-24-oyl-CoA	ITC	Isothermal titration calorimetry
3-OPC-CoA	3-oxo-pregn-4-ene-20-carboxyl-CoA	<i>M. smegmatis</i>	<i>Mycobacterium smegmatis</i>
4,9-DSHA	4,5,9,10-diseco-3-hydroxy-5,9,17-triox androsta-1(10),2-diene-4-carboxylic acid	<i>M. tuberculosis</i>	<i>Mycobacterium tuberculosis</i>
5-FU	5-fluorouracil	MALLS	Multiangle laser light scattering
<i>A. thaliana</i>	<i>Arabidopsis thaliana</i>	MDR	multidrug resistance
AB	human peroxisomal AB-thiolase	Mdr1	Multidrug resistance pump 1
ABC	ATP-binding cassette	MES	2-(N-morpholino)ethanesulfonic acid
ACC	hydrogen-bond acceptor probe	MFS	Major facilitator subfamily
AD	4-androsten-3,17-dione	MHR	middle homology region
ADA	N-(2-Acetamido)iminodiacetic acid	MR	molecular replacement
ADD	1,4-androstadiene-3,17-dione	Mrr1	Multidrug resistance regulator 1
AnD	hydrogen-bond acceptor/donor probe	Mrr1 '250'	Mrr1 '250' construct
APS	Ammonium peroxydisulfate	Mrr1 AR	Mrr1 Armadillo Repeat construct
ARO	aromatic probe	Mrr1 ARlong	Mrr1 elongated Armadillo Repeat construct
AtKAT	3-ketocyl-CoA thiolase in <i>A. thaliana</i>	Mrr1 FL	Mrr1 Full Length construct
ATP	Adenosine triphosphate	mtbTFE	<i>M. tuberculosis</i> trifunctional enzyme
BESSY	Berliner Elektronen-Speicherring Gesellschaft für Synchrotronstrahlung	MWCO	Molecular weight cut-off
BGMV	Buffered Glycerol-complex Medium for Yeast	NTA	nitrilotriacetic acid
BMMV	Buffered Methanol-complex Medium for Yeast	ORF	origin of replication
BRE	benzoyl response element	<i>P. pastoris</i>	<i>Pichia pastoris</i>
bZip	basic leucine zipper domain	PAD agar	<i>Pichia</i> -Adenine-Dropout Agar
<i>C. albicans</i>	<i>Candida albicans</i>	PCR	polymerase chain reaction
Cap1	<i>C. albicans</i> AP-1	PDB	Protein Data Bank
CCD	Charge-coupled device	PDIM	phthiocerol dimycocerosates
CDC	Center of Disease Control	PEG	polyethylene glycol
Cdr1/2	Candida drug resistance pump 1/2	pPinkHC	pichia pink high copy plasmid
CFU	colony forming units	pPinkLC	pichia pink low copy plasmid
CoA	Coenzyme A	pPinkαHC	secreting pichia pink high copy plasmid
CT	human cytosolic thiolase	RMSD	root mean square deviation
CV	column volume	<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
Da	Dalton	SCP2	peroxisomal SCP2-thiolase
DBD	DNA-binding domain	SD	Superdex
ddH2O	bidest water	SDS	Sodium dodecyl sulfate
DMSO	dimethyl sulfoxide	SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
DNA	Deoxyribonucleic acid	SE	size-exclusion
dNTP	desoxiribonucleotide triphosphate	SEC-MALLS	Size-exclusion chromatography Multiangle laser light scattering
DOHNAA	9,17-dioxo-1,2,3,4,10,19-hexanorandrost-5-oiic acid	SLIC	Sequence and ligation independent cloning
DON	hydrogen-bond donor probe	SOC medium	Super Optimal Broth medium with glucose
DOTS	directly observed treatment short-course	ssp.	subspecies
DRE	Drug response element	Strep-tag	streptavidin tag
DTT	dithiothreitol	<i>T. thermophilus</i>	<i>Thermus thermophilus</i>
<i>E. coli</i>	<i>Escherichia coli</i>	T1	mitochondrial T1-thiolase
ECL	enhanced chemiluminescent substrates	T2	mitochondrial T2-thiolase
EDTA	Ethylenediaminetetraacetic acid	Tac1	transcriptional activator of cdr genes 1
ESRF	European Synchrotron Radiation Facility	TB	tuberculosis
FLU	fluorine probe	TB medium	terrific broth medium
Flu1	Fluconazole resistance pump 1	TED	tris(carboxymethyl)ethylene diamine
FPLC	Fast Protein Liquid Chromatography	TEMED	Tetramethylethylenediamine
GST-tag	glutathione-S-transferase tag	TFE-thiolase	trifunctional enzyme thiolase
HAL	halogen probe	T _m	melting temperature
HA-tag	human influenza hemagglutinin - tag	WHO	World Health Organization
His ₆ -tag	hexahistidine tag	WT	wild type
HIV	Human immunodeficiency virus	XDR	extensively drug resistant
HRE	H ₂ O ₂ response element	YPD	Yeast peptone dextrose (glucose) medium

8.4 List of Figures, Tables and Equations

Figures

Figure 1 - Image of a SABHI agar plate culture of the fungus *Candida albicans* grown at 20 °C 8

Figure 2 - Histopathology of *Candida albicans* infection..... 9

Figure 3 - Selection of antifungal agents:..... 12

Figure 4 - Crystal structure of the DNA-binding domain (DBD) of Gal4: 18

Figure 5 - Location of functional regions and gain of function mutations in Mrr1: 18

Figure 6 - Illustration of the DNA-binding motif of Mrr1:..... 20

Figure 7 - *Mycobacterium tuberculosis*: 22

Figure 8 - *M. tuberculosis* colonies: 23

Figure 9 - Cholesterol degradation pathway of *Rhodococcus sp.* RHA1, *M. tuberculosis* H37Rv, and *M. bovis* bacillus Calmette Guerin. 29

Figure 10 - Proposed role of FadA5 in the side-chain degradation of cholesterol: 31

Figure 11 - Proposed thiolase reaction mechanisms of FadA5:..... 32

Figure 12 - *M. smegmatis* colonies..... 50

Figure 13 - Principle of cloning by restriction digestion and ligation: 54

Figure 14 - SLIC:..... 55

Figure 15 - Protein-binding to immobilized Co²⁺ ions (schematic illustration): 62

Figure 16 - Principle of size exclusion chromatography: 66

Figure 17 - Phase diagram of protein crystallization: 73

Figure 18 - Mrr1 constructs for expression in *E. coli*:..... 85

Figure 19 - Constructs of Mrr1 for expression in *P. pastoris*: 87

Figure 20 - 15% SDS-PAGE of the Mrr1 FL test expression with BL21 (DE3) cells: 89

Figure 21 - 15% SDS-PAGE of the affinity chromatography purification tests of Mrr1 FL: 90

Figure 22 - Thermofluor experiments with Mrr1 FL: 91

Figure 23 - Size-exclusion chromatography run with a HiPrep 26/60 Sephacryl HR S-300 column: 93

Figure 24 - Purification of Mrr1 FL: 94

Figure 25 - 15% SDS-PAGE of the Mrr1 AR test expression with BL21(DE3)-pLysS cells: 95

Figure 26 - 15% SDS-PAGE of the affinity chromatography purification tests of Mrr1 AR: 96

Figure 27 - Size-exclusion chromatography with a Superdex 200 10/300 GL column: 97

Figure 28 - 15% SDS-PAGE of the Mrr1 250 test expression with BL21(DE3)-pLysS cells: 98

Figure 29 - Thermofluor experiments with Mrr1 '250': 99

Figure 30 - 15% SDS-PAGE of Mrr1 '250' after affinity chromatography with Ni-NTA: 100

Figure 31 - Purification of Mrr1 '250': 101

Figure 32 - Size-exclusion elution profiles and SDS-PAGE analyses of Mrr1 '250' K335N and P683S: 103

Figure 33 - Size-exclusion elution profiles and SDS-PAGE analyses of the Mrr1 '250' G878E and G997V variants: 104

Figure 34 - 15% SDS-PAGE after Ni-NTA affinity chromatography of the Mrr1 '250'CT construct: 105

Figure 35 - Purification of the Mrr1 '250' CT construct: 105

Figure 36 - 15% SDS-PAGE analyses of the test expressions in *P. pastoris*: 107

Figure 37 - 15% SDS-PAGE analyses of the test expression control experiments in *P. pastoris*: 109

Figure 38 - Selected DLS measurements: 111

Figure 39 - Circular-dichroism experiments of Mrr1 '250' variants: 113

Figure 40 - 15% SDS-PAGE analysis of Mrr1 '250' G878E after size-exclusion chromatography: 115

Figure 41 - Crystallization conditions of Mrr1 '250' WT with cleaved His₆-tag 118

Figure 42 - Crystallization conditions of the Mrr1 '250' variants P683S and K335N with cleaved His₆-tag: 119

Figure 43 - 15% SDS-PAGE of the metal-affinity chromatography elution profile of FadA5 WT, and the C93S and C93A variants: 121

Figure 44 - Purification of FadA5 WT: 121

Figure 45 - SEC-MALLS graphs of FadA5 WT, C93S and C93A: 122

Figure 46 - Isothermal plots of the ITC experiments with FadA5 WT in the presence of different ligands: 123

Figure 47 - Crystallization conditions of FadA5 WT apo 126

Figure 48 - Overall structure of FadA5 WT apo I:..... 131

Figure 49 - Dimeric structure of thiolases 132

Figure 50 - Comparison of the two apo FadA5 structures: 133

Figure 51 - Details of apo structure II and its active site: 134

Figure 52 - Co-crystallization conditions of FadA5 WT and FadA5 C93S: 136

Figure 53 - Electron density and interaction network in the acetylated FadA5 WT(-acetyl)-CoA complex 139

Appendix

Figure 54 - OAS93 and CoA electron density in the FadA5 C93S-CoA-complex	141
Figure 55 - Ligand binding in the OPC-complex:	143
Figure 56 - Steroid-binding pocket of the OPC ligand:.....	145
Figure 57 - Structure and active site comparison of the FadA5 structures:	148
Figure 58 - Differences in ligand binding:.....	150
Figure 59 - Location of the thiolase specific water:	151
Figure 60 - Comparison of the FadA5 OPC-complex with (modeled) human thiolases	153
Figure 61 - Schematic illustration of secondary structure motifs, regulating domains and the degree of degradation of Mrr1:	158
Figure 62 - Models of the Mrr1 full length protein:.....	162
Figure 63 - Lead compound determination based on the FadA5 product 3-OPC-CoA:	170

Tables:

Table 1 - First- and second-line anti TB drugs	25
Table 2 - Selection of compounds in clinical trials or approved for MDR TB therapy.....	27
Table 3 - Technical Equipment	36
Table 4 - Consumables.....	37
Table 5 - List of enzymes and additives.....	38
Table 6 - List of commercially available Kits used during this study	39
Table 7 - List of bacterial strains and yeast strains	39
Table 8 - List of expression vectors	40
Table 9 - Components for DNA agarose gel electrophoresis	41
Table 10 - List of antibiotic solutions and concentrations used	41
Table 11 - List of the used media and additives.....	42
Table 12 - List of buffers for protein purification and analysis.....	42
Table 13 - Buffers and Solutions for SDS-PAGE.....	44
Table 14 - Solutions for Western Blotting.....	44
Table 15 - Crystallization Screens.....	44
Table 16 - Crystallization Compounds.....	45
Table 17 - Computer software and Databases.....	46
Table 18 - PCR cycle and pipetting scheme for cloning or site-directed mutagenesis.....	48
Table 19 - PCR cycle and pipetting scheme for the colony PCR.....	48
Table 20 - Established metal-affinity chromatography procedures	63
Table 21 - Anion-exchange chromatography protocols.....	65
Table 22 - DLS parameter setting	70
Table 23 - Experimental parameters for the ITC experiment	72
Table 24 - Mrr1 constructs for <i>E. coli</i> expression.....	86
Table 25 - General information about the Mrr1 <i>E. coli</i> constructs.....	87
Table 26 - Mrr1 construct list for <i>P. pastoris</i> expression	88
Table 27 - General information about the Mrr1 <i>P. pastoris</i> constructs	88
Table 28 - Summary of the performed protein expression tests for all three expression vectors and the four strains	110
Table 29 - Summary of all performed CD experiments with Mrr1 '250' variants	113
Table 30 - Summary of all performed SEC-MALLS experiments with Mrr1 '250' variants.....	116
Table 31 - Summary of the SEC-MALLS experiments:.....	123
Table 32 - Results of the ITC measurements.....	125
Table 33 - Data collection and refinement statistics FadA5 WT apo.....	129
Table 34 - RMSD value comparison:	132
Table 35 - Data collection and refinement statistics of the FadA5 WT and C93S complex structures	146

Equations:

Equation 1 - Bragg's law:	76
Equation 2 - Calculation of the structure factors:.....	76
Equation 3 - Calculation of the electron density:.....	77
Equation 4 - Correlation between the intensities I_{hkl} and the structure factor amplitudes F_{hkl} :	77
Equation 5 - Alternative calculation of the structure factors F_{hkl} :	77
Equation 6 - Definition of the R_{merge} value:.....	79

Appendix

Equation 7 - Pearson correlation coefficient CC:	79
Equation 8 - Calculation of the Matthew's coefficient:	80
Equation 9 - Patterson function P_{uvw} :.....	80
Equation 10 - Calculation of the R-values:	81

Supplemental Figures:

Figure S 1 - Purification of FadA5 C93S:	186
Figure S 2 - Purification of FadA5 C93A:.....	187

Supplemental Tables:

Table S 1 - Used primers	188
Table S 2 - Composition and list of all used buffers during this work	189
Table S 3 - Composition of the standard thermofluor screen	191
Table S 4 - Composition of the FadA5 crystallization screen	192
Table S 5 - Composition of the Thiolase crystallization screen.....	194

8.5 DNA and amino acid sequences of protein constructs

The vector part of the constructs is shown in capital letters and the legend for the variant color code is assigned beneath the sequences.

FadA5 in pSD31

DNA sequence:

atg ccc gag gta gtt ttt atc cat gga tct cac cac cac cac cac gga tcc atg ggt tac cgg gtc atc gtt gaa gcc acc cgc agc ccc atc ggc
aaa cgc aac gga tgg ctg tgc ggg ctg cat gcc acc gag ttg ttg ggc gcg gtc caa aag gcg gtc gtc gac aag gcc ggc atc cag tcc ggc ctt
cac gcc ggt gac gtc gaa cag gtc atc ggc ggt tgc gtc acc cag ttc ggg gag caa tcc aac aac atc agc cgg gtc gcc tgg ctg acg gcc ggt ttg
ccc gaa cac gtc ggc gcc acc acc gtc gac tgc cag tgc ggc agc ggc cag cag gcc aac cat ctg att gcc ggg ttg atc gcg gcc ggt gcc atc
gat gtc ggc atc gcc tgc ggc atc gag gcg atg agc cgg gtc ggg ctg ggc gcc aac gcc ggg cgg gac cgc tgc ctg atc cgc gcg cag tca tgg
gat atc gac ctg ccg aac cag ttc gag gcc gcc gag cgg atc gcc aag cgg cgc ggc atc acc cgc gag gac gtc gat gtc ttc ggg ctc gag tgc
cag cga cgc gcg cag cgg gcc tgg gcg gag ggc cgc ttt gac cgc gag atc tgc ccg atc cag gcg ccg gtc ctc gac gag cag aat cag ccc acc
ggc gag cgg cgc ctg gtc ttt cgc gac cag ggc ctg cgc gag acc acg atg gcg ggg cta ggc gag ctg aaa ccg gtc ctc gag ggc gcc atc cac
acc cgc ggc acg tgc tgc cag atc tcc gac ggc gcg gca gcc gtc ttg tgg atg gac gaa gcc gtc gca cgt gcg cac ggc ctg acc ccg cgg gcc
cgg atc gtc gcc cag gca ctc gtc ggc gcc gag ccc tac tac cac ctg gac ggc ccg gtc cag tcc acc gcg aag gtc ctg gag aag gcc ggc atg
aag atc ggc gac atc gac atc gtc gag atc aac gag gcg ttc gcg tcc gtc gtc ctg tcc tgg gcg cgg gtc cac gag ccc gac atg gac cgg gtc
aac gtc aac ggc ggg gcg atc gcg ctg ggg cat ccg gtc ggc tgc acc ggc agc cgg ctg atc acc acc gcc ctg cac gag ctc gag gcg acc gac
cag agc ctc gcg ctg atc acc atg tgc gcc ggc ggg gcc ctg tcc acc ggc acc atc atc gag cgg att taa

Variants: C93A and C93S

Amino acid sequence:

MPEVVFHIGSHHHHHHGS

MGYPV IVEAT RSPIG KRNGW LSLGH ATELL GAVQK AVVDK AGIQS GLHAG DVEQV IGGCV TQFGE QSNNI SRVAW LTAGL PEHVG
ATTVD CQCGS GQQAN HLIAG LIAAG AIDVG IACGI EAMSR VGLGA NAGPD RSLIR AQSWD IDLPN QFEAA ERIAK RRGIT REDVD
VFGLE SQRRA QRAWA EGRFD REISP IQAPV LDEQN QPTGE RRLVF RDQGL RETTM AGLGE LKPV L EGGIH TAGTS SQISD GAAAV
LWMDE AVARA HGLTP RARIV AQALV GAEPY YHLDG PVQST AKVLE KAGMK IGDID IVEIN EAFAS VVLSW ARVHE PDMDR
VNVNG GAIAL GHPVG CTGSR LITTA LHELE RTDQS LALIT MCAGG ALSTG TIIER I*

C93A and C93S

Mrr1 Full length in pBadm11 / pPink α HC / pPinkHC / pPinkLC

DNA sequence:

ATG AAA CAT CAC CAT CAC CAT CAC CCC ATG AGC GAT TAC GAC ATC CCC ACT ACT GAG AAT CTT TAT TTT CAG GGC GCC
ATG AGT ATT GCA ACC ACC CCG ATT GAA ACC CCG AAA AGC CCG AAA TCA ACC GAA CCG CAG GTT CGT AAA CGT AAA AAA
GTT AGC ACC GTT TGT ACC AAT TGC CGC AAA CGC AAA ATT CGT TGT GAT CGT CAG CAT CCG TGC AAT AAT TGC ATC AAA
AGC AAA AAA CAC AAC GCC TGC GTG TAT GAT GAT GGT CAG GTT AGT CCG GCA AAT TTT AGC ACC AAT GGT AGC TCA CAT
GGT AAT ACC GTT CCG GAA AGC CGT CCG TAT GAA GAA AGC GCA CGT ATT CCG ATT CGT TTT GAT GCC GAA GCA CCG CGT
AAA AAA AGC AAA CCG AAT ACC CCG AAT AAC GAA CGC AAA AAC AGC AAA AAA AGT CCG GAT AAT ACC GTG GCA AAT AAT
CAG CAG ACC GCA AGC GAA AAT GAA GTT ACC ATT ACC CTG AGC GAA CTG AAT ATG CTG AAA CAG CGT CTG CAA AAT ATC
GAA GCA AAC ATT AAT GCA CAG AGC AAT CCG CAG AGT AAT CCG AGC TAT GTT CCG CAG ACA CCG GCA TAT CCG ACC CAG
CCG AAT ATT CTG CCT CCG CCT GTT AGC TTT AAT AGC TGG TCA CCG AAA CAG AGT AAT GAA CGT GTT ATG TTT AGT CCG
CAG CAG CGT CTG ACC ACC AAT TAT AAC GTT AGC CAT ACC CGT GGC CAG AGC CCG AGC ATT CAG CTG CCA CCG CTG AGC
TTT AAA GAT ACA CCG CGT GCA AGC ATT GAT AGC GCA CCG CTG TAT AGC GAA ATG AGC CCT CCG CGT AGC GAT CTG ATT
GCA AGC AGC CTG ACC TCA CCG GAA AGC ATT CAG ATG AGC GTT AGC GGT GAT GTT GTT GGT GTT AAT CCG TAT CTG AAT
GAA ACC GAA ACC ATC AAC TTC TAT GAT GGC TAT ACC AGC ATT TGC GTG CGT GAT TTT CGT CGT GTT AAT CAT GGT CCG TTT
GCA TGG TCA AGC CTG ATG CGT AAA GAT AAA GCA CTG AGC AGC CTG TGG AAT CAT ATC CTG AAA AAG AAA GAG AAG AAG
AAC GTT GCA AGC CAG ACC TTT GTT TTT GGT CAG GAT GTT CAC GAA ATC AGC CAA GAA AAT ACC CAG CTG GTT GCC AGC
GAA AGC AAT GAA AGT GAA ACC AAG TTC AAA AAA AAG ACC CTG GAA ACC TTC GGC TTT AAT GAT GTT GTG CCG TAT GAC
ATC CTG AAG AAA AAA CTG CAA ACC CAG ATC AAC AAA ACC ACC AGT CCG CTG GGT CTG ACC CTG TAT GAG GAA CAG GTG
AAT ATG GAA CTG CAA CTG GTT GAT CGT ATT CAT CAG CAG CTG CCG AAA AAA AAG GTT CTG TGG AAA CTG ATC GAT CGC

Appendix

TTT TTT AGC CTG CTG TAT CCG TTT ATG CCG TTT CTG GAT GAA ATT GAT TTT CGC GAA TCC GTC ACC AAA ATT ATC GGT GAA ACC GAG TAT AAG GAC GAG AAA ATC AAA GAA CTG AAG GTG GAA AAA CGT CTG GAT CTG GCA GTT ATT GGT GTT CTG CTG ATC ATT CTG CGT ATG AGC TAT CTG AGC CTG TTT TGT AAT AAA GAA AGC GTG AAT GAA ATG CGC CTG AAA ACC ACC GAT CCG AGT CCG GAA GCA CAG GAC ATG AAA TAT CTG CTG CAA AAT CCG ATT GGC ATT AGC CTG ATT GAT AGT GCA CAG AAT TGC CTG CAA TAC TTT GAC ATT TTT CGC AAA ACC TCC ATG CCG GTT TTA CAG TGT GCA TAT TTT CTG CAA CTG TAC CAT ATC TTT GCA CCG GAA GAT GGT GAT GAC GGT GAT GGT GCA GAT ACC TAT GCA CTG AAT AGC ATG GTT GTT CGT ATG GCA TAT AGC ATG GGT CTG AAT CGT GAA CCG GAT AAT TTC AAA GAT GTG CTG AAT GAT AAG CGC CAG AAT CAT CTG GGT CGT AAA ATT TGG CAT TTT CTG GTG ATT GGC GAC GTG CAT AAT AGC TAT GCA TTT GGC ACC CCG AAA CTG ATT GGT GAT GAT TTT TAT GAT ACC AAG GTG CCG TTT ATC GAA GAA GGT AAC GAA AAC CTG ATC GAT AAA AGC CTG GAT CAG TAT GTT ACC AAA AGC GTT TTT CCG GGT TAC TTC AGC ATT TAT AAC AGC GTT GAC CAG ATT CTG AAG CTG ATT CTG AGC GTG AGC CGT CGT AGC AAA GTT AGC GAA ATT TGC AAA ATT CTG AAC CAG TTC GAA ATC GGT ATC GCA GAA CAG TAT GGT ACA CTG AGC GAT TGT CTG AAA CCT AAA GAA AAT CTG ATC CAC ATT TTT GCC CGT AAT ATG CCC GTG AAA ATG TAT ATT TCC CTG AAA AGC TTT CTG GTG TCC GTT TAT TTC CAC CTG TTT CTG TAT TAC GAG CAC AAA AAT GAT AGC CTG AGC TTT TTC TAC CTG CCG AAA ATC TTA AAA ACC GGT GCC GGT GAT ATT ATG CCG CAT TAT TTT GAA CTG CTG GGT AAT AGC GAA GTT GTT TGC GAT ATG GTG ATC AAC CCT AAA CTG ATT CAG ATC ATC CAT AAA GCC AAC CAG ATT AAC ATT GCC CTG ATT ATT CGC GTG AAC ATG AGC ATT TAC CGC ATG AAA AAT AGC CAG CAT CAT GCC GAG AAC TGC AAA AAA GAT GAT TTC TAC TAC AGC TAC TAT AAA GAA CTG TGC AAG TTT AGC AGT TGT CTG ACC CGT TGT GCA GAA GTT GGT ATT GCA GCA GTT AGC AAA CTG AGC ACC CGT TAT TAC TAT GCA TGG AAA ATT ACC AAA GGC CAC AAC TTT CTG CTG AAA ACA ATT ACC AGC ATG GAA TTT TAT GAG AAA GAA AGC ACC AAC GCC CAA GAA ATT ACA CTG CCG AAA TAC AAA CTG GAA CAA ATT GCC GAT CTG GAA AAC ATT TGT GAA GTG GCC CTG AAT AAA CTG GGT AAA ACC AGC GTT ATG GGT GAT GAA TTT TGT AGC AAC GTG AAC TAC AAG AAA TAT AAG GGC GAT CAG ACC TAT AGC ACC AGC AGC GAA AGT AGC AGC ACA CCG AAT AAA GAT TCT CCG CTG GAT AGC CGT AAA TAT ACC AAT GAT TTT GGT CTG GAC CTG GTG AAT AAC CAA GAA ATC GAC AAA ATC TGG TTA CAG ATG CTG AGC CTG AAA AGT GAA GAA GCA CAG CAG CAG CGC CAG CAA GAA AGC CAG CCG TTT ACC AGC AGC CAG AGC CAG AGT CAG TCA CCG CTG ACC AGC GCA AAT CAG GGT TAT ATG CCT CGT CCG GAA TCA CGT CGT GGT AGT TAT TAT GGC AAT ACC CCG TTT GCA CTG GAA AAC TTA AAC TTT GAT GGC TTT GGT GGC CAG TCA AAA AGC AGC AAT AAT GGT GAA GCC GAT CTG AGC AGC TTT GAT TTC TTT GTG GAT CTG CCG TTT GAT CAG CTG TTT ACC AAC TAA

Amino acid sequence:

[MKHHHHHPMSDYDIPTTENLYFQGA](#)

MSIAT TPIET PKSPK STEPQ VRKRK KVSTV CTNCR KRKIR CDRQH PCNNC IKSCK HNACV YDDGQ VSPAN FSTNG SSHGN TVPES RPYEE SARIP IRFDA EAPRK KSKPN TPNNE RKNSK KSPDN TVANN QQTAS ENEVT ITLSE LNMLK QRLQN IEANI NAQSN PQSNP SYVPQ TPAYP TQPNL LPPPV SFNSW SPKQS NERVM FSPQQ RLTTN YNVSH TRGQS PSIDL PPLSF KDTPR ASIDS APLYS EMSPP RSDLI ASSLT SPESI QMSVS GDVVG VNPYL NETET INFYD GYTSI CVRDF RRVNH GPFOW SSLMR KDKAL SSLWN HILKK KEKKN VASQT VFVGG DVHEI SQENT QLVAS ESNES ETKFK KKTLE TFGFN DVVPY DILKK KLQIQ INKTT SPLGL TLYEE QVNME LQLVD RIHQQ LPKKK VLWKL IDRFF SLLYP FMPFL DEIDF RESVT KIIGE TEYKD EKIKE LKVEK RLDLA VIGVL LIILR MSYLS LFCNK ESNVE MRLKT TDPSA EAQDM KYLLQ NPIGI SLIDS AQNCL QYFDI FRKTS MPVLQ CAYFL QLYHI FAPED GDDGD GADTY ALNSM VVRMA YSMGL NREPD NFKDV LNDKR QNHLG RKIWH FLVIG DVHNS YAFGT PKLIG DDFYD TKVPF IEEGN ENLID KSLDQ YVTKS VFPGY FSIYN SVDQI LKLIL SVSRR SKVSE ICKIL NQFEI GIAEQ YGTLS DCLKP KENLI HIFAR NMPVK MYISL KSFLV SVYFH LFLYY EHKND SLSFF YLRKI LKTGA GDIMP HYFEL LGNSE VVCDM VINPK LIQII HKANQ INIAL IIRVN MSIYR MKNSQ HHAEN CKKDD FYSY YKELC KFSSC LTRCA EVGIA AVSKL STRYY YAWKI TKGHN FLLKT ITSME FYEKE STNAQ EITLP KYKLE QIADL ENICE VALNK LGKTS VMGDE FCSNV NYKKY KGDQT YSTSS ESSST PNKDS PLDSR KYTND FGLDL VNNQE IDKIW LQMLS LKSEE AQQQR QQESQ PFTSS QSQSQ SPLTS ANQGY MPRPE SRRGS YYGNT PFALE NLNFD GFGGQ SKSSN NGEAD LSSFD FFDVL PFDQL FTN*

Mrr1 Armadillo Repeat in pBadm11

DNA sequence:

[ATG AAA CAT CAC CAT CAC CAT CAC CCC ATG AGC GAT TAC GAC ATC CCC ACT ACT GAG AAT CTT TAT TTT CAG GGC GCC ACC AGC ATT TGC GTG CGT GAT TTT CGT CGT GTT AAT CAT GGT CCG TTT GCA TGG TCA AGC CTG ATG CGT AAA GAT AAA GCA CTG AGC AGC CTG TGG AAT CAT ATC CTG AAA AAG AAA GAG AAG AAG AAC GTT GCA AGC CAG ACC TTT GTT TTT GGT CAG GAT GTT CAC GAA ATC AGC CAA GAA AAT ACC CAG CTG GTT GCC AGC GAA AGC AAT GAA AGT GAA ACC AAG TTC AAA AAA AAG ACC CTG GAA ACC TTC GGC TTT AAT GAT GTT GTG CCG TAT GAC ATC CTG AAG AAA AAA CTG CAA ACC CAG ATC AAC AAA ACC ACC AGT CCG CTG GGT CTG ACC CTG TAT GAG GAA CAG GTG AAT ATG GAA CTG CAA CTG GTT GAT CGT ATT CAT CAG CAG CTG CCG AAA AAA AAG GTT CTG TGG AAA CTG ATC GAT CGC TTT TTT AGC CTG CTG TAT CCG TTT ATG CCG TTT CTG GAT GAA ATT GAT TTT CGC GAA TCC GTC ACC AAA ATT ATC GGT GAA ACC GAG TAT AAG GAC GAG AAA ATC AAA GAA CTG AAG GTG GAA AAA CGT CTG GAT CTG GCA GTT ATT GGT GTT CTG CTG ATC ATT CTG CGT ATG AGC TAT CTG AGC CTG TTT TGT AAT AAA GAA AGC GTG AAT GAA ATG CGC CTG AAA ACC ACC GAT CCG AGT CCG GAA GCA CAG GAC ATG AAA TAT](#)

Appendix

CTG CTG CAA AAT CCG ATT GGC ATT AGC CTG ATT GAT AGT GCA CAG AAT TGC CTG CAA TAC TTT GAC ATT TTT CGC AAA
ACC TCC ATG CCG GTT TTA CAG TGT GCA TAT TTT CTG CAA CTG TAC CAT ATC TTT GCA CCG GAA GAT GGT GAT GAC GGT
GAT GGT GCA GAT ACC TAT GCA CTG AAT AGC ATG GTT GTT CGT ATG GCA TAT AGC ATG GGT CTG AAT CGT GAA CCG GAT
AAT TTC AAA GAT GTG CTG AAT GAT AAG CGC CAG AAT CAT CTG GGT CGT AAA ATT TGG CAT TTT CTG GTG ATT GGC GAC
GTG CAT AAT AGC TAT GCA TTT GGC ACC CCG AAA CTG ATT GGT GAT GAT TTT TAT GAT ACC AAG GTG CCG TTT ATC GAA
GAA GGT AAC GAA AAC CTG ATC GAT AAA AGC CTG GAT CAG TAT GTT ACC AAA AGC GTT TTT CCG GGT TAC TTC AGC ATT
TAT AAC AGC GTT GAC CAG ATT CTG AAG CTG ATT CTG AGC GTG AGC CGT CGT AGC AAA GTT AGC GAA ATT TGC AAA ATT
CTG AAC CAG TTC GAA ATC GGT ATC GCA GAA CAG TAT GGT ACA CTG AGC GAT TGT CTG AAA CCT AAA GAA AAT CTG ATC
CAC ATT TTT GCC CGT AAT ATG CCC GTG AAA ATG TAT ATT TCC CTG AAA AGC TTT CTG GTG TCC GTT TAT TTC CAC CTG TTT
CTG TAT TAC GAG CAC AAA AAT GAT AGC CTG AGC TTT TTC TAC CTG CGG AAA ATC TTA AAA ACC GGT GCC GGT GAT ATT
ATG CCG CAT TAT TTT GAA CTG CTG GGT AAT AGC GAA GTT GTT TGC GAT ATG GTG ATC AAC CCT AAA CTG ATT CAG ATC
ATC CAT AAA GCC AAC CAG ATT AAC ATT GCC CTG ATT ATT CGC GTG AAC ATG AGC ATT TAC CGC ATG AAA AAT AGC CAG
CAT CAT GCC GAG AAC TGC AAA AAA GAT GAT TTC TAC TAC AGC TAC TAT AAA GAA CTG TGC AAG TTT AGC AGT TGT CTG
ACC CGT TGT GCA GAA GTT GGT ATT GCA GCA GTT AGC AAA CTG AGC ACC CGT TAT TAC TAT GCA TGG AAA ATT ACC AAA
GGC CAC AAC TTT CTG CTG AAA ACA ATT ACC AGC ATG GAA TTT TAT GAG AAA GAA AGC ACC AAC GCC CAA GAA ATT ACA
CTG CCG AAA TAC AAA CTG GAA CAA ATT GCC GAT CTG GAA AAC ATT TGT GAA GTG GCC CTG AAT AAA CTG GGT AAA ACC
AGC GTT ATG GGT GAT GAA TTT TGT AGC AAC GTG AAC TAC AAG AAA TAT AAG GGC GAT CAG ACC TAT AGC ACC AGC AGC
GAA AGT AGC AGC ACA CCG AAT AAA GAT TCT CCG CTG GAT AGC CGT AAA TAT ACC AAT GAT TTT GGT CTG GAC CTG GTG
AAT AAC CAA GAA ATC GAC AAA ATC TGG TTA CAG ATG CTG AGC CTG AAA AGT GAA GAA GCA CAG CAG CAG CGC CAG CAA
GAA AGC CAG CCG TTT ACC TAA

Amino acid sequence:

[MKHHHHHPMSDYDIPTTENLYFQGA](#)

TSICV RDFRR VNHGP FAWSS LMRKD KALSS LWNHI LKKKE KKNVA SQTFV FGQDV HEISQ ENTQL VASES NESET KFKKK TLETF
GFNDV VPYDI LKKKL QTQIN KTTSP LGLTL YEEQV NMELO LVDRI HQQLP KKKVL WKLID RFFSL LYPFM PFLDE IDFRE SVTKI IGETE
YKDEK IKELK VEKRL DLAVI GVLLI ILRMS YLSLF CNKES VNEMR LKTTD PSPEA QDMKY LLQNP IGISL IDSAQ NCLQY FDI FR KTSMP
VLQCA YFLQL YHIFA PEDGD DGDGA DTYAL NSMNV RMAYS MGLNR EPDNF KDVLN DKRQN HLGK IWHFL VIGDV HNSYA
FGTPK LIGDD FYDTK VPFIE EGNEN LIDKS LDQYV TKS VF PGYFS IYNSV DQILK LILSV SRRSK VSEIC KILNQ FEIGI AEQYG TSDC
LKPKE NLIHI FARNM PVKMY ISLKS FLVSV YHFLF LYYEH KNDSL SFFYL RKILK TGAGD IMPHY FELLG NSEVV CDMVI NPKLI QIIHK
ANQIN IALII RVNMS IYRMK NSQHH AENCK KDDFY YSYK ELCKF SSCLT RCAEV GIAAV SKLST RYYYY WKITK GHNFL LKTIT
SMEFY EKEST NAQEI TLPKY KLEQI ADLEN ICEVA LNKLG KTSVM GDEFK SNVNY KKYKG DQTYS TSSES SSTPN KDSPL DSRKY
TNDFG LDLVN NQEIDK IWLQM LSLKS EEAQQ QRQQE SQPFT*

Mrr1 Armadillo Repeat long in pPink α HC / pPinkHC / pPinkLC

DNA sequence:

[ATG AAA CAT CAC CAT CAC CAT CAC CCC ATG AGC GAT TAC GAC ATC CCC ACT ACT GAG AAT CTT TAT TTT CAG GGC GCC ACC](#)
AGC ATT TGC GTG CGT GAT TTT CGT CGT GTT AAT CAT GGT CCG TTT GCA TGG TCA AGC CTG ATG CGT AAA GAT AAA GCA
CTG AGC AGC CTG TGG AAT CAT ATC CTG AAA AAG AAA GAG AAG AAG AAC GTT GCA AGC CAG ACC TTT GTT TTT GGT CAG
GAT GTT CAC GAA ATC AGC CAA GAA AAT ACC CAG CTG GTT GCC AGC GAA AGC AAT GAA AGT GAA ACC AAG TTC AAA AAA
AAG ACC CTG GAA ACC TTC GGC TTT AAT GAT GTT GTG CCG TAT GAC ATC CTG AAG AAA AAA CTG CAA ACC CAG ATC AAC
AAA ACC ACC AGT CCG CTG GGT CTG ACC CTG TAT GAG GAA CAG GTG AAT ATG GAA CTG CAA CTG GTT GAT CGT ATT CAT
CAG CAG CTG CCG AAA AAA AAG GTT CTG TGG AAA CTG ATC GAT CGC TTT TTT AGC CTG CTG TAT CCG TTT ATG CCG TTT
CTG GAT GAA ATT GAT TTT CGC GAA TCC GTC ACC AAA ATT ATC GGT GAA ACC GAG TAT AAG GAC GAG AAA ATC AAA GAA
CTG AAG GTG GAA AAA CGT CTG GAT CTG GCA GTT ATT GGT GTT CTG CTG ATC ATT CTG CGT ATG AGC TAT CTG AGC CTG
TTT TGT AAT AAA GAA AGC GTG AAT GAA ATG CGC CTG AAA ACC ACC GAT CCG AGT CCG GAA GCA CAG GAC ATG AAA TAT
CTG CTG CAA AAT CCG ATT GGC ATT AGC CTG ATT GAT AGT GCA CAG AAT TGC CTG CAA TAC TTT GAC ATT TTT CGC AAA
ACC TCC ATG CCG GTT TTA CAG TGT GCA TAT TTT CTG CAA CTG TAC CAT ATC TTT GCA CCG GAA GAT GGT GAT GAC GGT
GAT GGT GCA GAT ACC TAT GCA CTG AAT AGC ATG GTT GTT CGT ATG GCA TAT AGC ATG GGT CTG AAT CGT GAA CCG GAT
AAT TTC AAA GAT GTG CTG AAT GAT AAG CGC CAG AAT CAT CTG GGT CGT AAA ATT TGG CAT TTT CTG GTG ATT GGC GAC
GTG CAT AAT AGC TAT GCA TTT GGC ACC CCG AAA CTG ATT GGT GAT GAT TTT TAT GAT ACC AAG GTG CCG TTT ATC GAA
GAA GGT AAC GAA AAC CTG ATC GAT AAA AGC CTG GAT CAG TAT GTT ACC AAA AGC GTT TTT CCG GGT TAC TTC AGC ATT
TAT AAC AGC GTT GAC CAG ATT CTG AAG CTG ATT CTG AGC GTG AGC CGT CGT AGC AAA GTT AGC GAA ATT TGC AAA ATT
CTG AAC CAG TTC GAA ATC GGT ATC GCA GAA CAG TAT GGT ACA CTG AGC GAT TGT CTG AAA CCT AAA GAA AAT CTG ATC
CAC ATT TTT GCC CGT AAT ATG CCC GTG AAA ATG TAT ATT TCC CTG AAA AGC TTT CTG GTG TCC GTT TAT TTC CAC CTG TTT
CTG TAT TAC GAG CAC AAA AAT GAT AGC CTG AGC TTT TTC TAC CTG CGG AAA ATC TTA AAA ACC GGT GCC GGT GAT ATT

Appendix

ATG CCG CAT TAT TTT GAA CTG CTG GGT AAT AGC GAA GTT GTT TGC GAT ATG GTG ATC AAC CCT AAA CTG ATT CAG ATC
ATC CAT AAA GCC AAC CAG ATT AAC ATT GCC CTG ATT ATT CGC GTG AAC ATG AGC ATT TAC CGC ATG AAA AAT AGC CAG
CAT CAT GCC GAG AAC TGC AAA AAA GAT GAT TTC TAC TAC AGC TAC TAT AAA GAA CTG TGC AAG TTT AGC AGT TGT CTG
ACC CGT TGT GCA GAA GTT GGT ATT GCA GCA GTT AGC AAA CTG AGC ACC CGT TAT TAC TAT GCA TGG AAA ATT ACC AAA
GGC CAC AAC TTT CTG CTG AAA ACA ATT ACC AGC ATG GAA TTT TAT GAG AAA GAA AGC ACC AAC GCC CAA GAA ATT ACA
CTG CCG AAA TAC AAA CTG GAA CAA ATT GCC GAT CTG GAA AAC ATT TGT GAA GTG GCC CTG AAT AAA CTG GGT AAA ACC
AGC GTT ATG GGT GAT GAA TTT TGT AGC AAC GTG AAC TAC AAG AAA TAT AAG GGC GAT CAG ACC TAT AGC ACC AGC AGC
GAA AGT AGC AGC ACA CCG AAT AAA GAT TCT CCG CTG GAT AGC CGT AAA TAT ACC AAT GAT TTT GGT CTG GAC CTG GTG
AAT AAC CAA GAA ATC GAC AAA ATC TGG TTA CAG ATG CTG AGC CTG AAA AGT GAA GAA GCA CAG CAG CAG CGC CAG CAA
GAA AGC CAG CCG TTT ACC AGC AGC CAG AGC CAG AGT CAG TCA CCG CTG ACC AGC GCA AAT CAG GGT TAT ATG CCT CGT
CCG GAA TCA CGT CGT GGT AGT TAT TAT GGC AAT ACC CCG TTT GCA CTG GAA AAC TTA AAC TTT GAT GGC TTT GGT GGC
CAG TCA AAA AGC AGC AAT AAT GGT GAA GCC GAT CTG AGC AGC TTT GAT TTC TTT GTG GAT CTG CCG TTT GAT CAG CTG
TTT ACC AAC TAA

Amino acid sequence:

[MKHHHHHPMSDYDIPTTENLYFQGA](#)

TSICV RDRFR VNHP FAWSS LMRKD KALSS LWNHI LKKKE KKNVA SQTFV FGQDV HEISQ ENTQL VASES NESET KFKKK TLETF
GFNDV VPYDI LKKKL QTQIN KTTSP LGLTL YEEQV NMELQ LVDRI HQQLP KKKVL WKLID RFFSL LYPFM PFLDE IDFRE SVTKI IGETE
YKDEK IKELK VEKRL DLAVI GVLLI ILRMS YLSLF CNKES VNEMR LKTTD PSPEA QDMKY LLQNP IGISL IDSAQ NCLQY FDIFR KTSMP
VLQCA YFLQL YHIFA PEDGD DGDGA DTYAL NSMNV RMAYS MGLNR EPDNF KDVLN DKRQN HLGRR IWHFL VIGDV HNSYA
FGTPK LIGDD FYDTK VPFIE EGNEN LIDKS LDQYV TKSVF PGYFS IYNSV DQILK LILSV SRRSK VSEIC KILNQ FEIGI AEQYG TSDC
LKPKE NLIHI FARNM PVKMY ISLKS FLVSV YFHLF LYYEH KNDSL SFFYL RKILK TGAGD IMPHY FELLG NSEVV CDMVI NPKLI QIIHK
ANQIN IALII RVNMS IYRMK NSQHH AENCK KDDFY YSYYK ELCKF SSCLT RCAEV GIAAV SKLST RYYA WKITK GHNFL LKTIT
SMEFY EKEST NAQEI TLPKY KLEQI ADLEN ICEVA LNKLG KTSVM GDEFK SNVNY KKYKG DQYTS TSSES SSTPN KDSPL DSRKY
TNDFG LDLVN NQEIQ IWLQM LSLKS EEAQQ QRQQE SQPFT SSQSQ SQSPL TSANQ GYMPR PESRR GSYYG NTPFA LENLN
FDGFG GQSKS SNNGE ADLSS FDFV DLPFD QLFTN*

Mrr1 '250' in pBadm11 / pPink α HC / pPinkHC / pPinkLC

DNA sequence:

ATG AAA CAT CAC CAT CAC CAT CAC CCC ATG AGC GAT TAC GAC ATC CCC ACT ACT GAG AAT CTT TAT TTT CAG GGC GCC
ATG GCA AGC GAA ATG AGC CCT CCG CGT AGC GAT CTG ATT GCA AGC AGC CTG ACC TCA CCG GAA AGC ATT CAG ATG AGC
GTT AGC GGT GAT GTT GTT GGT GTT AAT CCG TAT CTG AAT GAA ACC GAA ACC ATC AAC TTC TAT GAT GGC TAT ACC AGC
ATT TGC GTG CGT GAT TTT CGT CGT GTT AAT CAT GGT CCG TTT GCA TGG TCA AGC CTG ATG CGT AAA GAT AAA GCA CTG
AGC AGC CTG TGG AAT CAT ATC CTG AAA AAG AA GAG AAG AAG AAC GTT GCA AGC CAG ACC TTT GTT TTT GGT CAG GAT
GTT CAC GAA ATC AGC CAA GAA AAT ACC CAG CTG GTT GCC AGC GAA AGC AAT GAA AGT GAA ACC AAG TTC AAA AAA AAG
ACC CTG GAA ACC TTC GGC TTT AAT GAT GTT GTG CCG TAT GAC ATC CTG AAG AAA AAA CTG CAA ACC CAG ATC AAC AAA
ACC ACC AGT CCG CTG GGT CTG ACC CTG TAT GAG GAA CAG GTG AAT ATG GAA CTG CAA CTG GTT GAT CGT ATT CAT CAG
CAG CTG CCG AAA AAA AAG GTT CTG TGG AAA CTG ATC GAT CGC TTT TTT AGC CTG CTG TAT CCG TTT ATG CCG TTT CTG
GAT GAA ATT GAT TTT CGC GAA TCC GTC ACC AAA ATT ATC GGT GAA ACC GAG TAT AAG GAC GAG AAA ATC AAA GAA CTG
AAG GTG GAA AAA CGT CTG GAT CTG GCA GTT ATT GGT GTT CTG CTG ATC ATT CTG CGT ATG AGC TAT CTG AGC CTG TTT
TGT AAT AAA GAA AGC GTG AAT GAA ATG CGC CTG AAA ACC ACC GAT CCG AGT CCG GAA GCA CAG GAC ATG AAA TAT CTG
CTG CAA AAT CCG ATT GGC ATT AGC CTG ATT GAT AGT GCA CAG AAT TGC CTG CAA TAC TTT GAC ATT TTT CGC AAA ACC TCC
ATG CCG GTT TTA CAG TGT GCA TAT TTT CTG CAA CTG TAC CAT ATC TTT GCA CCG GAA GAT GGT GAT GAC GGT GAT GGT
GCA GAT ACC TAT GCA CTG AAT AGC ATG GTT GTT CGT ATG GCA TAT AGC ATG GGT CTG AAT CGT GAA CCG GAT AAT TTC
AAA GAT GTG CTG AAT GAT AAG CGC CAG AAT CAT CTG GGT CGT AAA ATT TGG CAT TTT CTG GTG ATT GGC GAC GTG CAT
AAT AGC TAT GCA TTT GGC ACC CCG AAA CTG ATT GGT GAT GAT TTT TAT GAT ACC AAG GTG CCG TTT ATC GAA GAA GGT
AAC GAA AAC CTG ATC GAT AAA AGC CTG GAT CAG TAT GTT ACC AAA AGC GTT TTT CCG GGT TAC TTC AGC ATT TAT AAC
AGC GTT GAC CAG ATT CTG AAG CTG ATT CTG AGC GTG AGC CGT CGT AGC AAA GTT AGC GAA ATT TGC AAA ATT CTG AAC
CAG TTC GAA ATC GGT ATC GCA GAA CAG TAT GGT ACA CTG AGC GAT TGT CTG AAA CCT AAA GAA AAT CTG ATC CAC ATT
TTT GCC CGT AAT ATG CCC GTG AAA ATG TAT ATT TCC CTG AAA AGC TTT CTG GTG TCC GTT TAT TTC CAC CTG TTT CTG TAT
TAC GAG CAC AAA AAT GAT AGC CTG AGC TTT TTC TAC CTG CCG AAA ATC TTA AAA ACC GGT GCC GGT GAT ATT ATG CCG
CAT TAT TTT GAA CTG CTG GGT AAT AGC GAA GTT GTT TGC GAT ATG GTG ATC AAC CCT AAA CTG ATT CAG ATC ATC CAT
AAA GCC AAC CAG ATT AAC ATT GCC CTG ATT ATT CGC GTG AAC ATG AGC ATT TAC CGC ATG AAA AAT AGC CAG CAT CAT
GCC GAG AAC TGC AAA AAA GAT GAT TTC TAC TAC AGC TAC TAT AAA GAA CTG TGC AAG TTT AGC AGT TGT CTG ACC CGT
TGT GCA GAA GTT GGT ATT GCA GCA GTT AGC AAA CTG AGC ACC CGT TAT TAC TAT GCA TGG AAA ATT ACC AAA GGC CAC

Appendix

AAC TTT CTG CTG AAA ACA ATT ACC AGC ATG GAA TTT TAT GAG AAA GAA AGC ACC AAC GCC CAA GAA ATT ACA CTG CCG
AAA TAC AAA CTG GAA CAA ATT GCC GAT CTG GAA AAC ATT TGT GAA GTG GCC CTG AAT AAA CTG GGT AAA ACC AGC GTT
ATG GGT GAT GAA TTT TGT AGC AAC GTG AAC TAC AAG AAA TAT AAG GGC GAT CAG ACC TAT AGC ACC AGC AGC GAA AGT
AGC AGC ACA CCG AAT AAA GAT TCT CCG CTG GAT AGC CGT AAA TAT ACC AAT GAT TTT **G**GT CTG GAC CTG GTG AAT AAC
CAA GAA ATC GAC AAA ATC TGG TTA CAG ATG CTG AGC CTG AAA AGT GAA GAA GCA CAG CAG CAG CGC CAG CAA GAA AGC
CAG CCG TTT ACC AGC AGC CAG AGC CAG AGT CAG TCA CCG CTG ACC AGC GCA AAT CAG GGT TAT ATG CCT CGT CCG GAA
TCA CGT CGT GGT AGT TAT TAT GGC AAT ACC CCG TTT GCA CTG GAA AAC TTA AAC TTT GAT GGC TTT GGT GGC CAG TCA
AAA AGC AGC AAT AAT GGT GAA GCC GAT CTG AGC AGC TTT GAT TTC TTT GTG GAT CTG CCG TTT GAT CAG CTG TTT ACC
AAC TAA

The **orange** nucleotides indicate the residues that were inserted due to the cloning procedure ahead of the Mrr1 '250' insert DNA.

K335N: A → C; P683S: C → T; G878E: GT → AG; G997V: G → T

Amino acid sequence:

MKHHHHHPMSDYDIPTTENLYFQGA

MASEM SPPRS DLIAS SLTSP ESIQM SVSGD VVGVN PYLNE TETIN FYDGY TSICV RDFRR VNHGP FAWSS LMRKD KALSS LWNHI
LKKKE KKNVA SQTFV FGQDV HEISQ ENTQL VASES NESET KFKKK TLETF GFNDV VPYDI LKK**KL** QTQIN KTTSP LGLTL YEEQV
NMELQ LVDRI HQQLP KKKVL WKLID RFFSL LYPFM PFLDE IDFRE SVTKI IGETE YKDEK IKELK VEKRL DLAVI GVLLI ILRMS YLSLF
CNKES VNEMR LKTTD PSPEA QDMKY LLQNP IGISL IDSAQ NCLQY FDIFR KTSMP VLQCA YFLQL YHIFA PEDGD DGDGA DTYAL
NSMVV RMAYS MGLNR EPDNF KDVLN DKRQN HLGK IWHFL VIGDV HNSYA FGTPK LIGDD FYDTK VPFI EGNEN LIDKS LDQYV
TKSVF **P**GYFS IYNSV DQILK LILSV SRRSK VSEIC KILNQ FEIGI AEQYG TSDC LKPKE NLIHI FARNM PVKMY ISLKS FLVSV YFHLF
LYYEH KNDSL SFFYL RKILK TGAGD IMPHY FELLG NSEVV CDMVI NPKLI QIIHK ANQIN IALII RVNMS IYRMK NSQHH AENCK
KDDFY YSYK ELCKF SSCLT RAEV **G**IAAV SKLST RYYA WKITK GHNFL LKTIT SMEFY EKEST NAQEI TLPKY KLEQI ADLEN ICEVA
LNKLG KTSVM GDEFC SNVNY KKYKG DQTYS TSSES SSTPN KDSPL DSRKY TNDF**G** LDLVN NQEIDK IWLQM LSLKS EEAQQ QRQEQ
SQPFT SSQSQ SOSPL TSANQ GYMPR PESRR GSYYG NTPFA LENLN FDGFG GQSKS SNNGE ADLSS FDFV DLPFD QLFTN*

The **orange** amino acids indicate the residues that were inserted due to the cloning procedure ahead of the Mrr1 '250' insert DNA.

K335N; P683S; G878E; G997V

Acknowledgements

Acknowledgements

First of all I would like to thank Professor Caroline Kisker for her trustful supervision and the guidance during my work on the PhD thesis. In addition, I would like to thank her for her patience and help in the processes of manuscript and thesis preparation.

Special thanks also to Professor Joachim Morschhäuser and Professor Nicole Sampson for being members of my thesis committee and for their collaboration on our projects, the helpful discussions and their good advices during the committee meetings.

This thesis would not have been possible without the contribution of all members of the Kisker and Schindelin laboratories. Thank you all for the almost always nice working atmosphere and the activities in and out of the lab.

I would also like to thank Professor Hermann Schindelin, Dr. Jochen Kuper and Dr. Florian Sauer for their contribution to the structure solutions and their always open ears for any questions concerning X-ray crystallography, data reduction and refinement processes.

In particular, I would also like to thank our technical assistants for keeping the lab running. Thank you Gudrun, Monika and Nicole. In a similar way Dr. Bernhard Fröhlich and Christian Weinberger shall be acknowledged here, as without them my data would have been lost several times in the last years.

Our small SBDD subgroup always provided an important and supporting working environment, therefore I would like to thank the current and former colleagues of the 'Seuchengruppe', namely Dr. Maria Neumann, Dr. Johannes Schiebel, Sandra Eltschkner, Uwe Dietzel and Yesid Ramirez.

I am particularly grateful for the fruitful collaboration with our collaborators in Professor Nicole Sampson's laboratory in Stony Brook, namely Rui Lu.

In the end I would like to thank my family and friends at home, in Würzburg and all over the world for distractions but also support during the last years.

In particular I want to thank my father Wolfgang Schäfer and my brother Kai for their support and understanding throughout the last years.

This page was intentionally left empty.

Affidavit

(eidesstattliche Erklärung)

I hereby declare that my thesis entitled

Approaching antimicrobial resistance – Structural and functional characterization of the fungal transcription factor Mrr1 from *Candida albicans* and the bacterial β -ketoacyl-CoA thiolase FadA5 from *Mycobacterium tuberculosis*.

is the result of my own work. I did not receive any help or support from commercial consultants. All sources and/or materials applied are listed and specified in the thesis.

Furthermore, I verify that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Würzburg

List of publications

C.M. Schaefer, R. Lu, N.M. Nesbitt, J. Schiebel, N.S. Sampson, C. Kisker; FadA5 a thiolase from *Mycobacterium tuberculosis* – a unique steroid-binding pocket reveals the potential for drug development against tuberculosis; *Structure*; **in press**

J. Schiebel, K. Kapilashrami, A. Fekete, G.R. Bommineni, **C.M. Schaefer**, M.J. Mueller, P.J. Tonge, C. Kisker; Structural basis for the recognition of mycolic acid precursors by KasA, a condensing enzyme and drug target from *Mycobacterium tuberculosis*; *JBC*; **2013**