

**Biochemical and Structural Characterization of
 β -Hydroxylases
in Non-Ribosomal Peptide Biosynthesis**

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Daniel A. Wirtz

aus

Köln

Bonn 2022

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachterin: Prof. Dr. Gabriele M. König

2. Gutachterin: Prof. Dr. Tanja Schneider

Tag der Promotion: 26.01.2022

Erscheinungsjahr: 2022

Table of Contents

Abstract.....	1
1 Introduction	3
1.1 Non-ribosomal peptides from bacteria	3
1.2 Non-ribosomal peptide synthetases	4
1.2.1 Basic architecture.....	4
1.2.2 In <i>trans</i> acting hydroxylases	8
1.3 Lipid II-binding antibiotics.....	16
1.3.1 Hypeptin	19
1.3.2 Lysobactin.....	21
1.4 FR900359	24
1.4.1 Structure and activity	24
1.4.2 Utilization as a pharmacological tool, clinical and ecological function	25
1.4.3 Biosynthesis.....	26
2 Aims of the study.....	29
2.1 Hypeptin	29
2.2 Lysobactin	29
2.3 FR900359	30
3 Results	31
3.1 Hypeptin	31
3.1.1 Phylogenetic analysis of 16S-rRNA DNA	31
3.1.2 Description of the biosynthetic potential of <i>Lysobacter</i> sp. K5869.....	33
3.1.3 Description of the <i>hyn</i> biosynthetic gene cluster	33
3.1.4 Cloning and heterologous expression of <i>hyn</i> genes.....	36
3.1.5 <i>In vitro</i> activity of A domains.....	38
3.1.6 Cloning and heterologous expression of <i>hynC</i> and <i>hynE</i>	41
3.1.7 <i>In vitro</i> activity of HynE.....	43
3.1.8 <i>In vitro</i> activity of HynC.....	45
3.1.9 Stereochemistry of hydroxylated amino acids	46
3.1.10 Bioinformatic analysis of HynA ₅ C ₂	47
3.1.11 <i>In vitro</i> activity of HynA ₅ C ₂	49
3.2 Lysobactin	51
3.2.1 Cloning and heterologous expression of <i>lyb</i> genes.....	51
3.2.2 <i>In vitro</i> activity of A domains.....	52
3.2.3 UV/Vis spectrum of LybC.....	54
3.2.4 <i>In vitro</i> activity of LybC.....	54
3.3 FR900359	55

3.3.1	Cloning and heterologous expression of <i>frs</i> genes	55
3.3.2	<i>In vitro</i> activity of A domains.....	57
3.3.3	UV/Vis spectrum of FrsH.....	59
3.3.4	<i>In vitro</i> activity of FrsH.....	59
3.4	Investigations of the substrate recognition of FrsH.....	61
3.4.1	Bioinformatic analysis on the recognition mechanism of <i>trans</i> acting NHDM	61
3.4.2	Size-exclusion chromatography.....	64
3.4.3	Isothermal titration calorimetry	67
3.4.4	Substitution of <i>frsH</i> <i>in vivo</i>	70
3.4.5	Structure elucidation of FrsH.....	72
3.4.6	Structure of FrsH	75
3.4.7	Dimeric character of FrsH	79
3.4.8	Putative docking interface of FrsH	83
3.4.9	Putative docking interface of FrsA ₁ A _{main}	86
3.4.10	<i>In vitro</i> assay to verify the interaction of FrsH and FrsA ₁ A _{main}	88
4	Summary and Conclusion.....	93
4.1	Hypeptin	93
4.2	Lysobactin	95
4.3	FR900359	96
5	Material and Methods.....	99
5.1	Chemicals and reagents	99
5.2	Bacterial strains	99
5.3	Plasmids.....	100
5.4	Media and buffer.....	101
5.4.1	Media	101
5.4.2	Buffers	102
5.5	Microbiological methods	103
5.5.1	Cultivation of bacterial strains.....	103
5.5.2	Cryopreservation of bacterial strains	103
5.5.3	Chemical transformation	103
5.5.4	Electroporation	104
5.6	Molecular biological methods	104
5.6.1	Primer design.....	104
5.6.2	Polymerase chain reaction (PCR).....	105
5.6.3	DNA isolation.....	106
5.6.4	Agarose gel electrophoresis.....	107
5.6.5	DNA isolation from agarose gels.....	107
5.6.6	Restriction hydrolysis	107
5.6.7	Dephosphorylation of DNA fragments.....	108

5.6.8	Ligation.....	108
5.6.9	Gibson assembly	108
5.6.10	Sanger sequencing	109
5.6.11	Generation of specific <i>E. coli</i> expression strains	109
5.7	Protein purification and analysis	110
5.7.1	Protein expression.....	110
5.7.2	Protein purification from expression culture	110
5.7.3	Desalting and concentration	111
5.7.4	Concentration measurement	111
5.7.5	Storage of proteins.....	111
5.7.6	Fast protein liquid chromatography (FPLC).....	111
5.7.7	SDS-Polyacrylamide gel electrophoresis (SDS-PAGE).....	112
5.7.8	Thrombolytic removal of N-terminal His-tag.....	113
5.7.9	Peptide cleavage utilizing TEV-protease.....	113
5.7.10	Analytical size-exclusion chromatography.....	114
5.8	<i>In vitro</i> assays	115
5.8.1	$\gamma^{18}\text{O}_4$ -ATP exchange assay	115
5.8.2	Hydroxylation assay	116
5.8.3	Fe/ α KG-dependent hydroxylase functionality assay with HynE.....	116
5.8.4	Side chain assembly assay	117
5.8.5	Epimerisation activity assay	117
5.9	Characterization of protein-protein interactions	118
5.9.1	Size exclusion chromatography	118
5.9.2	Isothermal titration calorimetry (ITC)	119
5.10	<i>In vivo</i> complementation of FrsH	119
5.10.1	Construction of complemented strains	119
5.10.2	Fermentation of <i>C. vaccinii</i> and extraction of FR.....	120
5.11	Bioinformatic methods	121
5.11.1	Determination of the reading frame.....	121
5.11.2	AntiSmash.....	121
5.11.3	Construction of phylogenetic trees with MEGA X.....	121
5.11.4	Visualization of proteins with PyMOL.....	121
5.11.5	SWISS-MODEL	122
5.11.6	I-TASSER.....	123
5.11.7	SwarmDock	123
5.11.8	ClusPro	124
6	References	127
7	List of abbreviations	135
8	Appendix	137

8.1	Supplementary data	137
8.1.1	Sequence alignments	137
8.1.2	Sequence data for phylogenetic analysis of C domains.....	139
8.1.3	<i>In silico</i> construction of FrsA _{1A} _{main}	140
8.1.4	Side chain assembly with mutated enzymes	141
8.1.5	NRM data of hypeptin	146
8.2	Primer	152
8.2.1	Primers for the hypeptin project	152
8.2.2	Primers for the lysobactin project.....	153
8.2.3	Primers for the FR900359 project	154
8.3	Plasmid maps	155
9	Publications	159
10	Acknowledgements.....	161

Abstract

Hydroxylases are essential in the biosynthesis of natural products by functionalization of building blocks, which derive from primary metabolism. They provide hydrophily for specific interactions or enable further derivatisation, which ultimately results in a natural product with outstanding activity. Hydroxylases, which are associated with non-ribosomal peptide synthetases (NRPS) contribute to the generation of diverse compounds with great significance for human society such as antibiotics or tumour therapeutics. Their wide-ranging importance is in stark contrast to the scarce knowledge about their structure, functionality, and substrate recognition. Hence, the *trans* acting hydroxylases in the biosynthetases of the non-ribosomal peptides hypeptin, lysobactin and FR900359 were investigated in this work.

Hypeptin is a cyclic octadepsipeptide from *Lysobacter* sp. with strong antibacterial activity against Gram-positive strains. The responsible biosynthetic gene cluster, (BGC) *hyn* encodes for two *trans* acting hydroxylases that are responsible for the β -hydroxylation of four amino acid residues. HynC belongs to the family of non-heme diiron monooxygenases (NHDM) and HynE is an iron- and α -ketoglutarate dependent hydroxylase. The characterization of HynC and HynE *in vitro* required the reconstitution of cognate NRPS modules, which were partly dependent on the MbtH-like protein, HynMLP that was not encoded within the *hyn* BGC. The successful reconstitution of HynC and HynE *in vitro* enabled determination of their substrate specificities, which were contradicting with the published stereoconfiguration of hypeptin. Finally, the reassigned configuration could be verified by in-depth NMR analyses.

The cyclic depsipeptide lysobactin shows a similar activity compared to hypeptin. The respective *lyb* BGC encodes for the interesting NHDM LybC, which is predicted to hydroxylate three structurally diverse amino acids. The cognate NRPS modules were successfully reconstituted *in vitro*, but no activity of the hydroxylase could be detected.

The selective and potent G α q protein inhibitor FR9003559 (FR) is a cyclic octadepsipeptide with three β -hydroxylated leucine residues. During its biosynthesis, the monomodular NRPS FrsA, together with the NHDM synthesise the FR side chain *N*-propionylhydroxyleucine. Here, the biosynthesis of the side chain was reconstituted *in vitro*. To get a deeper insight into the mechanism of substrate recognition, the structure of FrsH was elucidated by crystallization trails and subsequent X-ray absorption spectroscopy. Additionally, the gene *frsH* was functionally substituted by *hynC* and *lybC* in the native producer *Chromobacterium vaccinii*, indicating a conserved structure of the NHDMs for substrate recognition. The gathered information were used for *in silico* docking studies to elucidate crucial residues for the interaction of FrsH and FrsA. Lastly, an assay was established to verify the outcome of the docking studies *in vitro*.

1 Introduction

This thesis focusses on a special class of hydroxylases, which is utilized by bacteria to introduce hydroxyl groups into non-ribosomal peptides (NRPs) at the β -carbon of single amino acid residues. Given the obligatory dependency of the hydroxylases on the respective non-ribosomal peptide synthetase, these are introduced first before the three relevant classes of related β -hydroxylases are described. Lastly, the compounds, which are relevant for this work, as well as their biosynthesis, will be introduced. Hypeptin and lysobactin both belong to the class of lipid II binding antibiotics and FR900359 is a Gαq protein inhibitor.

1.1 Non-ribosomal peptides from bacteria

Secondary metabolites (also referred to as natural products, NP) play versatile roles in the life cycle of bacteria. In contrast to primary metabolites, they are not essential for the life cycle of the organism *per se*, but enable adaption to harmful environments, communication, and predatory or protective behaviour. The multifaceted functionality of natural products is mirrored by their vast structural diversity, which is achieved by functionalization of building blocks from the primary metabolism by fundamentally different biosynthetic pathways. In bacterial genomes, the enzymes that are necessary for the production of a specific NP are located in direct contact with each other, forming so called biosynthetic gene clusters (BGCs). For the bacterium, this organisation enables the transfer of the ability to produce the respective NP in a single occasion of horizontal gene transfer. Nowadays, for researchers, bacterial BGCs are easy to detect and to analyse with relatively simple bioinformatic tools, once the genomic DNA sequence is obtained.^[1]

A major group of bacterial NPs are NRPs. In contrast to the fundamental process of ribosomal peptide and protein synthesis, NRPs are almost exclusively found in bacteria and fungi. So far, the function of most bioinformatically detected and chemically characterized NRPs remains elusive. Nevertheless, many NRPs are regarded as virulence factors and toxins. The compound enterobactin from *E. coli*, for example, is produced under Fe(III)-deficient conditions. Outside the cell, it coordinates the metal ion with extremely high affinity and is then re-imported to secure iron supply for cellular processes. This facilitates the invasion of human tissue, where enterobactin is able to capture Fe(III) from human transferrin and haemoglobin.^[2] As a second example, the plant pathogen *Ralstonia solanacearum* produces ralsolamycin to induce the formation and invasion of chlamydospores of soil dwelling fungi. This fungal phenotype enhances the environmental persistence and probably enables the long-term survival of the invading bacteria in absence of the actual host plant.^[3] Furthermore, several NRPs and derivatives thereof are used to treat bacterial or fungal infections. Vancomycin and teicoplanin from *Amycolatopsis orientalis* and *Actinoplanes*

teichomyceticus are antibiotics of last resort to treat infections of methicillin resistant *Staphylococcus aureus* (MRSA).^[4,5] Additionally, bacterial NRPs contribute to the repertoire of anticancer drugs with bleomycin from *Streptomyces verticillus* and romidepsin from *Chromobacterium violaceum*.^[6]

Taken together, the class of bacterial NRPs has a significant influence on human society. New interesting compounds are reported regularly, but detailed biochemical investigation is hampered by low production yields from the native producer and complex chemical synthesis. In this context, a deeper comprehension of the biosynthesis might facilitate efficient heterologous expression or open doors for chemoenzymatic syntheses. Furthermore, detailed knowledge of the native BGC enables precise modification of the production system for the formation of new NRP derivatives with altered pharmacological and pharmacokinetic properties.^[7]

1.2 Non-ribosomal peptide synthetases

1.2.1 Basic architecture

Non-ribosomal peptide synthetases (NRPS) are biocatalytic machineries in bacteria and fungi. Depending on the product peptide, their size ranges between 100 kDa and 150 MDa. NRPS assemble peptides in a stepwise process, in which each amino acid is recruited and modified separately, resulting in products of outstanding structural diversity. In general, NRPS follow the co-linearity rule: the number and order of the elongation modules corresponds to the arrangement of the amino acid residues in the final peptide. However, there are a growing number of reports about NRPS systems that violate this rule.^[8] The most important exceptions are iterative systems, where one or more modules perform several elongation steps before the peptide is passed downstream along the assembly line, *e.g.* the enterobactin biosynthesis in *E. coli*.^[9] Due to the higher complexity and low relevance for this work of iterative systems, this introduction will solely consider “classical” NRPS that work in a linear manner.

The minimal NRPS elongation module consists of a condensation (C), adenylation (A), and thiolation (T) domain, which work together to assemble a specific substrate and integrate it into the growing peptide chain in a stepwise process (Figure 1.1). The first two steps of adenylation and thiolation are catalysed by the A domain. The binding pockets of A domains display a specificity conferring pattern of amino acids residues, also known as the Stachelhaus code, to recognize a single, most often proteinogenic, amino acid as substrate. This pattern is used to reliably predict the amino acid substrate from the primary structure of the enzyme.^[10] Once coordinated in the active site, the substrate is activated under the expense of ATP to an aminoacyl-adenylate. In the second step, this intermediate is covalently bound to the adjacent T domain via a thioester.

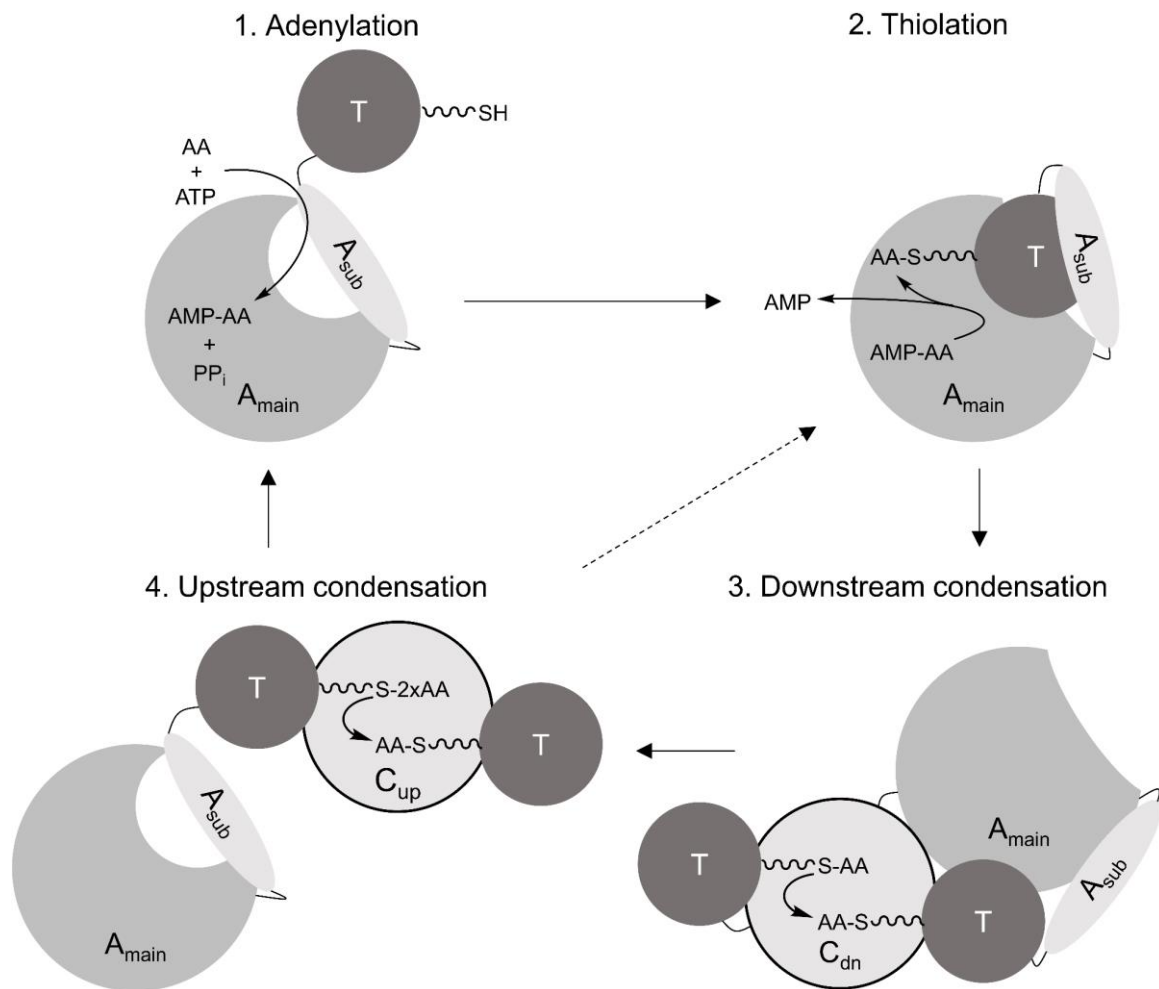


Figure 1.1: Simplified elongation cycle of a NRPS module. In adenylation state, the adenylation (A) domain activates a specific amino acid (AA) in the active site, which is shaped by the larger (A_{main}) and the smaller (A_{sub}) subdomains. The activated aminoacid-adenylate (AMP-AA) remains coordinated by A_{main} , whereas pyrophosphate (PP_i) is released. Afterwards, A_{sub} swings open to pull the thiolation (T) domain towards the active site. In the thiolation step, the amino acid is thiolated onto the phosphopantetheine moiety of the T domain. If present, the loaded T domain is then coordinated in the acceptor side of the downstream condensation (C_{dn}) domain, which catalyses the transfer of the amino acid from the preceding module. As a final step, the T domain converges to the donor binding site of the upstream C domain (C_{up}) to pass on the peptide to the subsequent module. Simultaneously to downstream condensation, the A domain is in adenylation state to initiate the next elongation cycle. Thus, after condensation, the T domain directly adopts thiolation state. Possible modifications by *in cis* acting domains or *in trans* acting enzymes usually occur after thiolation and prior to condensation.

The T domain is also referred to as a peptidyl carrier domain to highlight its noncatalytic function. After translation, T domains need to be activated by the installation of a 4'-phosphopantetheinyl (PPant) moiety as a prosthetic group derived from coenzyme A to the hydroxyl side chain of a highly conserved serine residue.^[11] The terminal thiol group serves as nucleophile in the thiolation reaction of the A domain. Once bound to the amino acid, the T domain translocates to the downstream C domain, which couples the bound peptide onto the amino acid of the next T domain in a condensation reaction, thereby releasing the upstream T domain for the next elongation cycle.

Thus, initiating modules of NRPS most often only consist of A and T domains.^[8] The terminus of NRPS might display various domains. Frequently, thioesterase domains (TE) catalyse hydrolytic release of linear peptides, or perform inter- or intramolecular esterifications. Further terminal domains perform thio reduction (TD domains) or cyclisation (Cy domains) of the terminal amino acid to release the peptide from the assembly line.^[12-14]

Structural insight into single NRPS domains was generally easily achieved since the early 2000s.^[15] However, the true power of NRPS originates from the manifold interplay of the single domains throughout the catalytic cycle of adenylation, thiolation and condensation reaction. Within the last five years, crystal structures of multidomain or even multimodular NRPS machineries were published, revealing a staggering conformational flexibility that allows major rearrangements of the domains between the single catalytic steps.^[16-19]

A domains commonly have a molecular weight of around 60 kDa and separate into two subdomains. The larger subdomain A_{main} comprises major parts of the active site and is linked to the small, 10 kDa A_{sub} by a highly flexible linker region. A_{sub} contains a conserved lysine residue that stabilizes the substrate amino acid during adenylation by interacting with its carboxylate moiety.^[17,20] After the adenylation reaction occurred and pyrophosphate is released, A_{sub} swings open to impede further adenylation. Additionally, the T domain, likewise connected by a flexible linker region, is coordinated towards A_{main} , such that its PPant moiety approaches the adenylate for the subsequent thiolation reaction. This rearrangement involves a 140° rigid body torsion of A_{sub} . Subsequent to thiolation, the T domain performs a 75° rotation and 61 Å translocation, combined with a 180° rotation and 21 Å translocation of A_{sub} to carry the bound substrate the distance of 50 Å to the upstream C domain.^[16,18]

Many A domains heavily depend on the non-covalent interaction with small (~8 kDa) MbtH-like proteins (MLP), to maintain their activity and structure. Generally, the occurrence of an MLP within a NRPS BGC indicates the dependence of at least some encoded A domains. Interestingly, the interaction is not specific, so MLPs are exchangeable across different NRPS systems *in vivo* and *in vitro*.^[21] Crystal structures of whole NRPS, co-crystallized with a cognate MLP, show a dense, non-covalent interaction with a specific binding site of A_{main} , but no major structural changes compared to A domains that do not depend on the interaction.^[19] Thus, MLPs rather seem to function as allosteric modulators than as chaperones, however the detailed function could not yet be determined.

T domains are 10 kDa small domains that adopt a four-helix bundle fold. The N-terminus of helix two harbours the highly conserved serine residue, which gets phosphopantetheinylated by respective promiscuous phosphopantetheinyl transferases to convert the domain from *apo* to *holo* state, thereby functionalizing the whole NRPS module.^[11] The substrate carrying T domain directly

interacts with the A and C domain, as well as all further modifying domains and in *trans* acting enzymes (see below). Unsurprisingly, analyses of complex crystal structures showed that the residues surrounding the conserved serine predominantly participate in complex coordination, but no specificity conferring residues could be identified.^[22] This raises the question on how the *holo* T domain selects the binding partner to mediate effective substrate trafficking and correct directionality. Current evidence suggests that the selectivity does not emerge from the T domain, but from its binding partners that display a “gatekeeping” selectivity towards the bound substrate.^[8] An analogous mechanism is also relevant for the distantly related fatty acid synthetases.^[23]

C domains are 50 kDa domains with a pseudosymmetric V-shape that are rigidly packed against the A domain of the respective module. The N-terminal C_{don} subdomain binds the substrate loaded T domain of the upstream module and C_{acc} binds the respective counterpart from downstream. Consequently, the PPant moieties of the T domains protrude from opposite sides into the C domain to arrange their bound amino acids near the catalytic “consensus” motive HHxxxDG, which is located at the connecting loop between C_{acc} and C_{don}.^[24] The second histidine of the consensus motive was proposed to act as a Lewis base to abstract the proton from the α -amino group of the thioester and initiate the condensation reaction,^[25] or to stabilize a possible anionic transition state post C-N bond formation.^[26] Apart from the central condensation reaction, C domains exhibit specificity towards their substrate, especially the acceptor subdomain. Several interplaying mechanisms ensure that the non-loaded PPant may not enter the substrate channel, but predominantly binds there once the amino acid is loaded.^[27] Additionally, erroneously loaded amino acids are repelled as well as such that ought to be modified prior to condensation.^[5]

The vast structural diversity of NRPs most often emerges from the individual modification of the amino acid after thiolation by additional in *cis* acting domains within the NRPS module. These modifying domains perform reactions such as *N*- or *O*-methylation,^[28] formylation,^[29] hydroxylation and many more.^[30] Most modifications are introduced by specialized C domains that contain the overall V-shaped fold, but have lost the original condensation activity. Common examples are epimerases (E), dual condensation/epimerization domains (C/E), cyclases, and starter C domains (C_{starter}).^[24,31] The latter are solely located in initiation modules of NRPS and catalyse the condensation of a fatty acid, from an acyl carrier protein or coenzyme A (CoA) to the acceptor amino acid in an *N*-acylation reaction.

1.2.2 In trans acting hydroxylases

In contrast to *cis* acting domains that are an integrative part of NRPS, *trans* acting enzymes are encoded on additional genes of the BGC. The resulting enzymes recognize the specific cognate module to perform the reactions such as halogenation or thiolation.^[32] The most interesting group here are β -hydroxylases, which functionalize C-H bonds and thereby expand the chemical space of the amino acid substrate. To overcome the very high activation energy, which is necessary to reach the C-H bond dissociation energy (420 kJ/mol),^[33] the hydroxylases utilize high-valent metal-ion intermediates. These in turn require tight control of the catalytic cycle to prevent cellular damage. The complex and energy consuming reaction gives good return as it introduces a functional group at a natively inert amino acid residue, making it susceptible for further modifications like macrocyclization,^[13,34,35] glycosylation,^[5] or transesterification.^[13] Additionally, free hydroxyl moieties often contribute to the biological activity of the final compound.^[13,36] So far, three enzyme families of *trans* acting β -hydroxylases in NRPS are described.

1.2.2.1 *Cytochrome P450 hydroxylases*

The cytochrome P450 superfamily is common to the metabolism of almost all eukaryotic and procaryotic organisms, and plays versatile roles in both anabolic and catabolic processes. Despite the diversity in terms of selected substrates and catalysed reactions, all members of this superfamily exhibit a well conserved structure and catalytic mechanism. The enzyme's activity depends on a prosthetic heme group, coordinated and encapsuled by a group of α -helices.^[37] In resting state, the heme-coordinated central Fe(III) is coordinated to a water molecule. The catalytic cycle is initiated by binding of the substrate in the adjacent binding pocket, which leads to spatial displacement of the water (Figure 1.2). The transfer of an electron from a redox partner (typically a ferredoxin) leads to reduction of Fe to ferric state (II), which reacts with environmental dioxygen and instantly triggers the delivery of a second electron, generating a ferric peroxyanion. In some cases, this transition state is used to generate peroxy-bridged products. Otherwise, two protons are transferred to the distal oxygen. Thus, the distal oxygen is released as water, leaving the proximal oxygen double-bound to the iron core, in a highly reactive iron-(IV)-oxo intermediate. The catalytic cycle proceeds with the abstraction of a proton from the bound substrate to a transient hydroxyl Fe(IV) intermediate. Rebound of the hydroxyl group to the radical substrate leads to dissociation of the remaining Fe(III), which subsequently coordinates water to regain resting state.^[38]

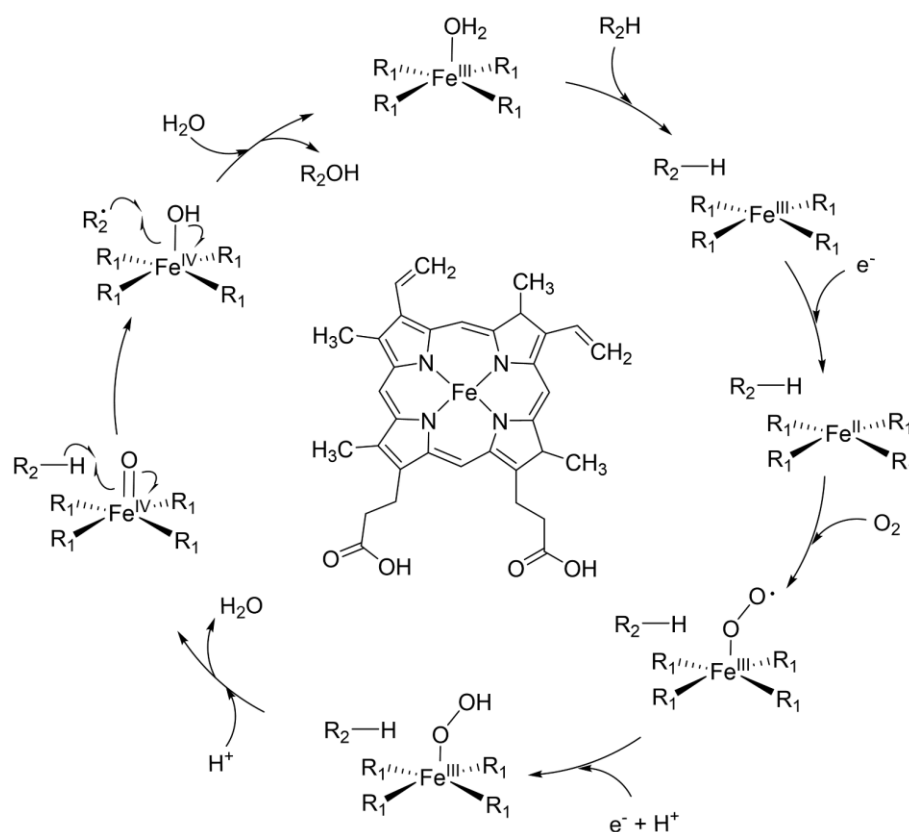


Figure 1.2: Simplified catalytic cycle of cytochrome P450 hydroxylases. The highly reactive Fe(IV)-oxo species is generated by controlled successive delivery of electrons and protons. Fe-coordinating type-b heme is depicted in the middle. The central iron is coordinated by a protoporphyrin ring. Figure adapted from Cook *et al.*^[39]

Members of the cytochrome P450 superfamily in NRP biosynthesis often target monomodular NRPS systems consisting of A, T, and TE domains. The loaded substrate amino acid is β -hydroxylated by a *trans* acting cytochrome P450 hydroxylase (CYP450) and is subsequently released by the TE domain. This mechanism provides the respective non-proteinogenic amino acid for the final compound, which is synthesized in a NRPS-independent manner. Common examples are the biosyntheses of nikkomycin^[40] and novobiocin.^[41] Multimodular NRPS systems, targeted by CYP450s were investigated in the case of skyllamycin. Here, the CYP450 P450_{sky} is responsible for β -hydroxylation of L-Phe₅, *O*-Me-L-Tyr₇, and L-Leu₁₁. In this thesis, the initial, as well as the following publication towards P450_{sky} focussed on its substrate recognition and are summarized in chapter 1.2.2.4.^[42,43]

1.2.2.2 *α -Ketoglutarate dependent hydroxylases*

Members of the iron- and α -ketoglutarate (also known as α -oxoglutarate) dependent dioxygenases are found in mammals and take part in important biological processes, such as collagen biosynthesis, L-carnitin biosynthesis, post-translational protein modification, epigenetic regulation and energy metabolism. In plants, bacteria and fungi, they are utilized for secondary metabolism to biosynthesize new natural products or to decompose xenobiotics.^[14,44] The most frequently occurring reactions are hydroxylations, but sequential oxidation, halogenation, demethylation, desaturation, epoxidation and ring formations, as well as ring expansions have also been reported.^[45] Structurally, the enzymes coordinate an iron atom in a double stranded β -helix core fold by a conserved H-X-D/E-X_n-H motive. The obligatory cofactor α -ketoglutarate bidentally binds to Fe(II), together with an environmental water molecule, to stabilize the activated iron core. Similar to cytochrome 450 hydroxylases, binding of the substrate within its binding pocket displaces the water molecule and leaves a vacant coordination site for dioxygen (Figure 1.3). The resulting Fe(III)-superoxo complex is short lived as the distal oxygen directly attacks the β -carbon of α -ketoglutarate, leading to oxidative decarboxylation of the cofactor to succinate. This process abstracts another electron off the system, leaving a high-valent Fe(IV)-oxo species. Comparable to the mechanism of cytochrome P450-containing hydroxylases, the alcohol product is formed by hydrogen abstraction of the substrate and subsequent OH rebound to the radical intermediate. The catalytic cycle is finalized by dissociation of the hydroxylated product and succinate, followed by recruitment of α -ketoglutarate and water.^[46]

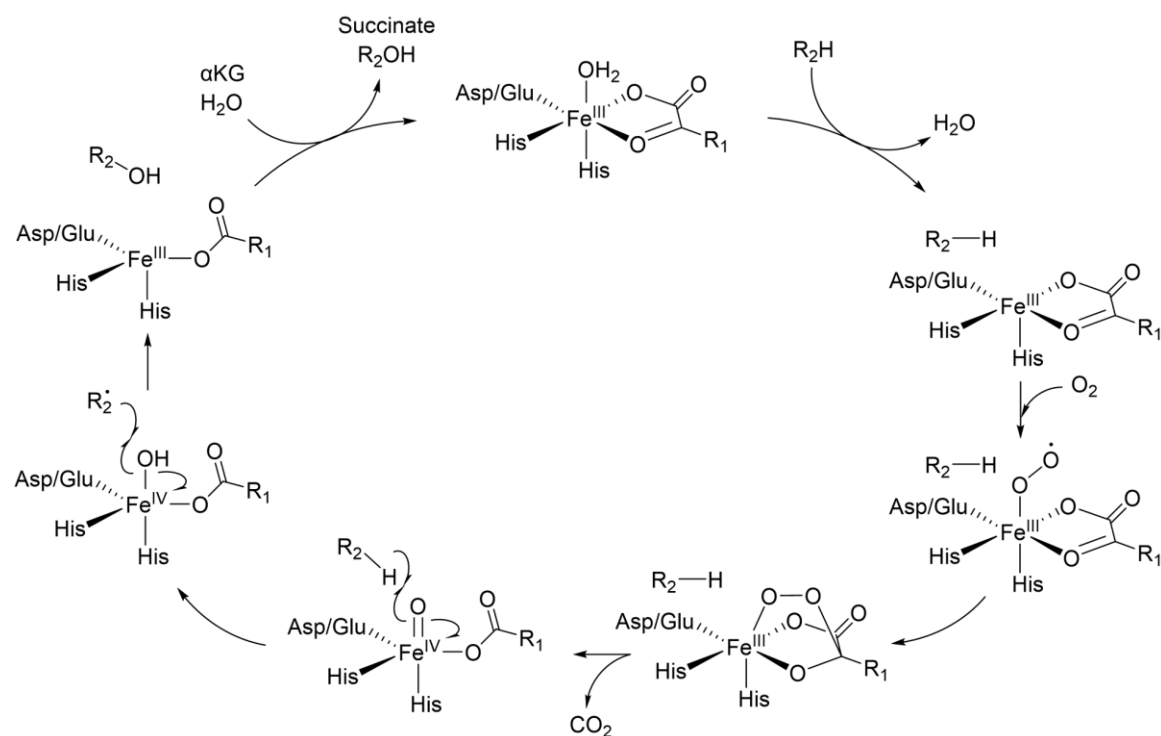


Figure 1.3: Generalized catalytic cycle of iron- and α -ketoglutarate dependent hydroxylases. The highly reactive Fe(IV)-oxo species is generated by oxidative decarboxylation of the cofactor α -ketoglutarate (α KG). (R_1 : CH_2-CH_2-COOH ; R_2 : substrate). Figure adapted from Zwick *et al.*^[47]

In NRPS-related systems, members of the iron- and α -ketoglutarate dependent dioxygenases superfamily either hydroxylate free amino acids as substrate, which is then recognized by the respective A domain,^[48] or they act on T domain-bound substrates. These in *trans* acting iron- and α -ketoglutarate dependent hydroxylases (Fe/ α KG) are frequently reported, for example in the BGSs of kutznerides,^[49] cystobactamide^[50] or potashchelins.^[51] Most of these reports refer to the prototype Fe/ α KG, SyrP from syringomycin biosynthesis^[52] and biochemically verify the hydroxylation reaction without further investigations regarding structure, substrate selectivity or kinetics of the enzyme. Recently, two research groups reported about Fe/ α KGs that are recruited to the cognate NRPS module by a special, noncatalytic C domain, which was named I domain (I for interaction). Additionally, phylogenetic analyses revealed a strict clustering dependent on the stereocontrol and substrate amino acid of the hydroxylases.^[51,53] Interestingly, these findings seem to be limited to siderophore BGCs, as such kind of Fe/ α KGs and I domains were not reported in other NRPS so far.

1.2.2.3 Non-heme diiron monooxygenases

NRPS-related *trans* acting non-heme diiron monooxygenases (NHDMs) belong to the superfamily of metallo-β-lactamase (MBL)-fold hydrolases.^[54] Some members of this superfamily have sequence similarities of down to 5 %, but all show a conserved αββα-fold and a H-x-H-x-D-H motif, which most often coordinates binuclear metal complex in the active site.^[55] The interplay of two metal ions enables the enzymes to perform complex reactions, the most infamous subfamily catalyses the oxidative ring-opening of β-lactam antibiotics. Further listed enzyme types are sulfatases, sulfur dioxygenases, glyoxylases, ribonucleases, and many more.^[55,56] Depending on the coordinating ligands of the dinuclear cluster and the conducted reactions, the generation of high-valent Fe(IV)-species is achieved by different catalytic cycles. Some MBLs even circumvent the formation of Fe(IV)-intermediates by utilizing μ-1,2-peroxo bridged Fe(III)-species for the hydroxylation event. Thus, a generalized catalytic cycle cannot be depicted.^[57]

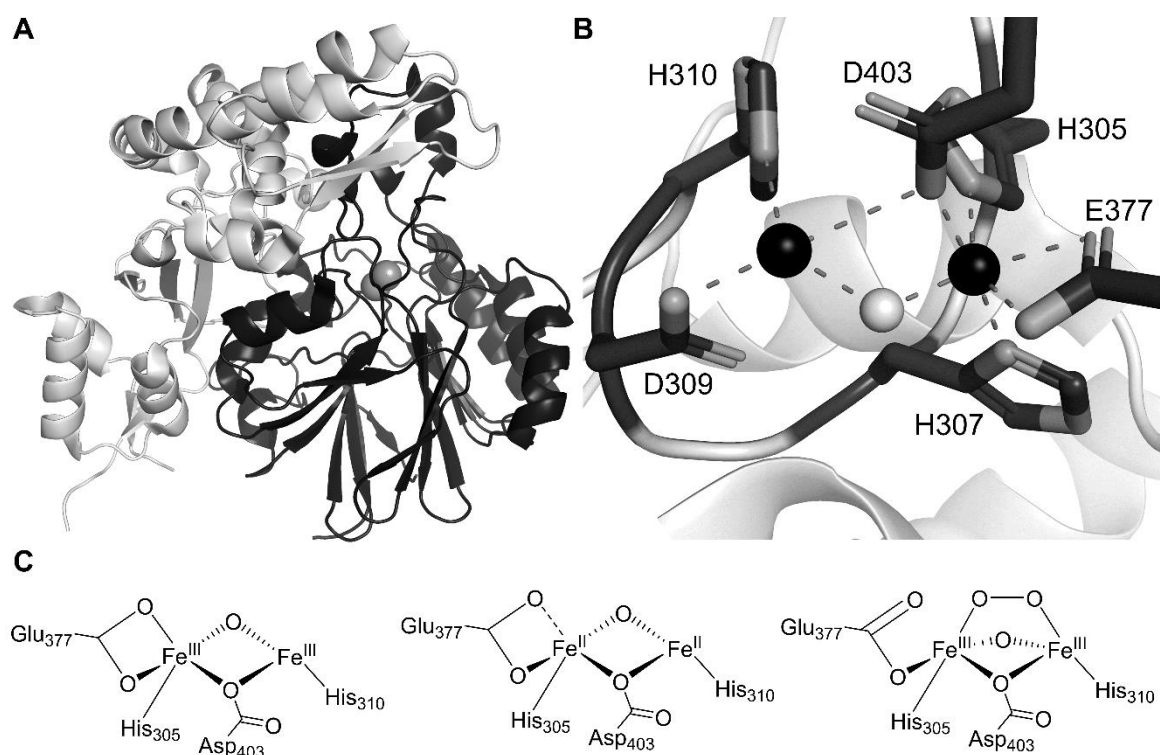


Figure 1.4: Structure of CmlA. **A** Ribbon model of a monomeric unit. The C terminal domain (dark grey) displays the metallo β-lactamase fold consisting of two central β-sheets, each flanked by three α-helices and coordinates the dinuclear iron cluster (spheres). The N-terminal domain (light grey) is rather distorted. **B** Coordination of the diiron cluster (dark spheres) in the active site of CmlA by the given amino acids and a bridging oxoanion (light grey sphere). The additional acetate was omitted for the sake of clarity. **C** Proposed structures of the dinuclear cluster in diferrous resting state, activated diferic state and upon O₂ binding in peroxo state. Adapted from Jasniewsky *et al.*^[58]

The prototype NHDM CmlA catalyses the β-hydroxylation of T domain bound *L-p*-aminophenylalanine (*L*-PAPA) in chloramphenicol biosynthesis.^[54] The successful structure elucidation, of CmlA by X-ray absorption spectroscopy, verified the initially hypothesized MBL-fold of the C-terminal subdomain (Figure 1.4). The N-terminal part however exhibits a rather

distorted structure of α -helices and small antiparallel β -sheets. The obvious role of the N-terminal subdomain is the formation of a large protruding arm, which enables the dimerization of two CmlA molecules. Further spectroscopic analyses, including UV/Vis, Mössbauer, EPR (electron paramagnetic resonance), XANES (X-ray absorption near-edge structure spectroscopy), and EXAFS (extended X-ray absorption fine structure), unambiguously verified the presence of an oxo bridged diiron cluster in the active site, coordinated by ligands of the MBL-fold.^[59] Fe1 is monodentally coordinated by H305, H307, and bidentally coordinated to E377. Fe2 binds monodentally to H310 and D309. In the crystal structure, Fe2 was additionally coordinated by an acetate, which did not affect the catalytic or spectroscopic properties of the enzyme in solution. Both iron atoms were additionally bridged by a μ -1,1 carboxylate from D403 and an oxoanion. Mutational analyses validated that binding of the substrate induces a shift of E377 to adopt a monodentally bound state. The now vacant coordination site of Fe1 is then directly occupied by dioxygen, generating a μ -1,2-peroxo species, bridging the two Fe(III) cores.^[58,59] From this initial state, several possible reaction paths consequently lead to the functionalization of the adjacent substrate by one oxygen, while the second one is reduced to water that dissociates. The iron cluster subsequently gets reconditioned by the transfer of two electrons. Identification of the responsible electron transport system still remains elusive. The chloramphenicol BGC does not encode for a respective enzyme, so candidates from the primary metabolism such as ferredoxin or flavo-iron-sulfur reductases have been suggested as mediators of the electron transfer. For *in vitro* experiments, CmlA was initially activated via chemical reduction with sodium dithionite in the presence of methyl viologen as the chemical electron transmitter. In subsequent studies, the enzyme was more reliably activated when the dithionite was substituted by NADH.^[60]

1.2.2.4 Substrate recognition of *trans* acting hydroxylases in NRPS

In general, *trans* acting hydroxylases in NRPS that modify T domain-bound substrates seem to distinguish the substrate rather by the structure or nature of the cognate NRPS module, than by the structure of the substrate itself. The first indication for this mechanism was provided by investigations into the biosynthesis of skyllamycin. Here, the *trans* acting CYP450, P450_{sky} is responsible for β -hydroxylation of the three T domain-bound amino acids, L-Phe₅, *O*-Me-L-Tyr₇, and L-Leu₁₁. Interestingly, loading the respective T domains with structurally related amino acids did not alter the substrate conversion rate of the hydroxylase *in vitro*. Additionally, skyllamycin also contains the non-hydroxylated residue Leu₁₀. Incubation of the L-Leu loaded T₁₀ domain with P450_{sky} yielded no hydroxylated product, again pointing out the importance of the T domain for substrate recognition of the *trans* acting hydroxylase.^[42] Subsequently, the observed substrate promiscuity towards the loaded amino acid was employed by replacing *O*-Me-L-Tyr on T₇ by an array of nitrogen-containing aromatic inhibitors. Some of them were also accepted as substrates,

enabling detailed kinetic studies indicating that the substrate recognition of the hydroxylase is driven by the interaction with the T domain, but high affinity is generated by coordination of the substrate within the binding pocket. This study finally led to the first successful crystallisation of a tailoring hydroxylase bound to its substrate T domain.^[43] The structural data revealed that most amino acid residues within the protein-protein interface of the T domain are conserved throughout the skyllamycin BGC. Therefore, the authors therefore concluded that slight alterations in the three-dimensional shape of the T domain alter its interface and thereby generate the crucial selectivity of P450_{sky}.

The substrate specificity of *trans* acting NHDMs was first discussed when the crystal structure of CmlA was published. The authors could not deduce any specificity-conferring residues in the binding interface with the *in silico* docked T domain. Additionally, the overall surface composition of the N-terminal domain of NHDMs is highly variable throughout the enzyme family. Only the tips of protruding arms are conserved and might therefore contribute to an interaction. However, in computed docking scenarios, the T domain is too distantly located from them. The authors thus hypothesized that the specificity of CmlA is conferred by interactions with the entire NRPS module and not just the T domain.^[59] However, in common protein-protein interactions, surface patches of at least 600 Å² or 17 amino acids per binding partner are buried by the convergence of the proteins and the interfaces are at least partly constituted of hydrophobic residues to drive the association of the binding partners by evading contact with hydrogen.^[61] Hence, the single conserved hydrophilic residues on the protruding arms might contribute to the interaction and its specificity, but cannot solely drive the affinity towards the cognate NRPS module.

The *trans* acting NHDM CmlA was further investigated in a recent publication. Its native substrate is the L-PAPA-loaded T domain of chloramphenicol biosynthesis, but it was also able to replace the native NHDM of teicoplanin biosynthesis, Tcp25, *in vitro*. In this special set-up, CmlA accepted a vast array of halogenated Tyr-derivatives, loaded onto the AT₆ didomain construct of Tcp11, demonstrating the promiscuity of *trans* acting hydroxylases towards the T domain-bound substrate. In the next step, the authors constructed a chimeric AT didomain from the A₆ domain of teicoplanin biosynthesis and the T₆ domain of kistamicin biosynthesis. Kistamicin is structurally related to teicoplanin but belongs to a different class of glycopeptides. One major difference is the missing β-hydroxy group at the L-Tyr residue that originates from module six. Nevertheless, incubation of CmlA with the chimeric AT didomain led to the formation of hydroxylated products in decent yields. In a next step, the T domain was replaced with T₇ of the skyllamycin BGC, which is a native substrate of the *trans* acting CYP450 P450_{sky}. However, in this case, no formation of hydroxylated products was observed after incubation with CmlA.^[60]

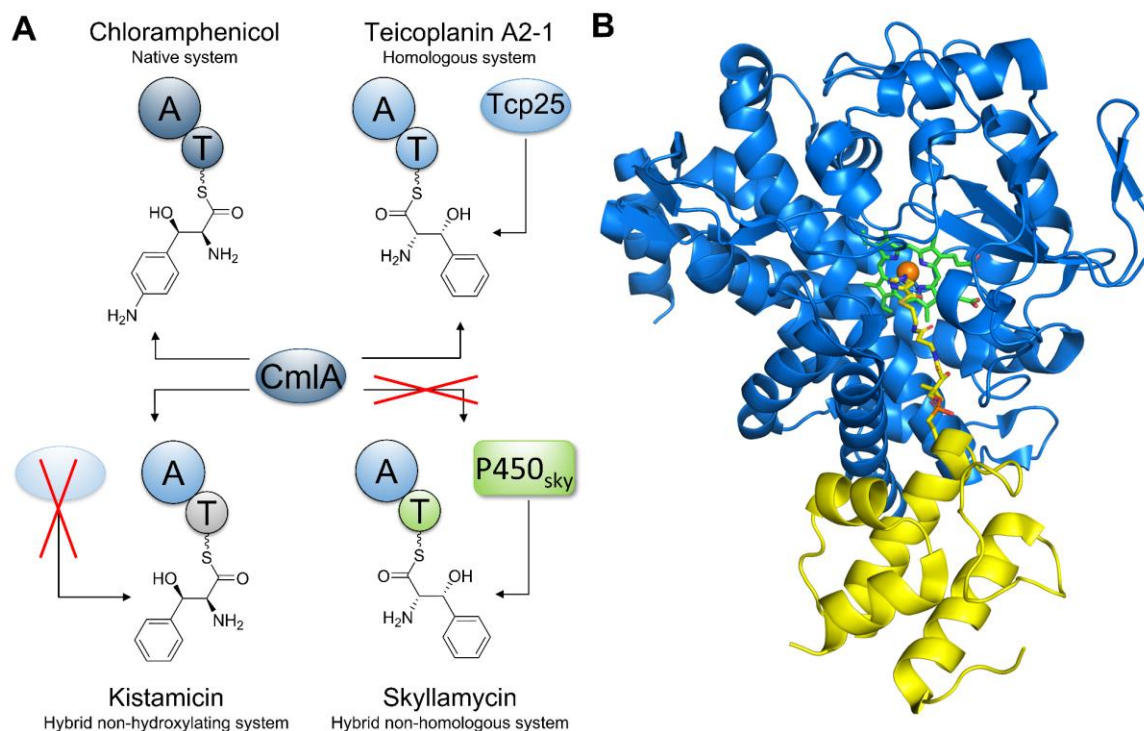


Figure 1.5: Substrate recognition of *trans* acting hydroxylases in NRPS. **A** The prototype NHDM CmlA natively β -hydroxylates L-p-aminophenylalanine in chloramphenicol biosynthesis. It functionally substitutes the homologue Tcp25 from teicoplanin biosynthesis *in vitro*. A chimeric construct of the A domain from teicoplanin biosynthesis and the T domain from kistamicin biosynthesis is also targeted by CmlA *in vitro*. The T domain of skyllamycin biosynthesis, which is natively targeted by the cytochrome P450, P450_{sky}, is incompatible and cannot be targeted by CmlA. **B** The only known structure of a *trans* acting hydroxylase in NRPS in complex with its cognate T domain (yellow ribbons, front) was obtained from P450_{sky} (blue ribbons). The phosphopantetheinyl (PPant) moiety and the heme group are displayed as yellow and green sticks, respectively with the iron core as brown sphere. The hydroxylase exclusively contacts residues in direct neighbourhood of the PPant attachment site (PDB ID: 4PXH).^[43]

Taken together, data about the mechanism of recognition of *trans* acting hydroxylases towards their substrate NRPS module are scarce, making it impossible to deduce general rules.^[22] Nevertheless, the above-mentioned publications give valuable information. First, the NHDM CmlA does not recognize the substrate T domain of the CYP450 P450_{sky}, suggesting a non-compatible mechanism of recognition for both hydroxylase systems. Secondly, whilst P450_{sky} is able to recognize the amino acid-loaded T monodomain, CmlA seems to depend on the amino acid-loaded AT didomain, in which the T domain is exchangeable. This indicates that CmlA, in contrast to P450_{sky}, does not recognize its substrate by the T domain, but by the A domain.

1.3 Lipid II-binding antibiotics

Investigations into the structure and biosynthesis of the lipid II-binding antibiotics hyeiptin and lysobactin were key aims of this thesis. For the general comprehension of their promising activity against problematic bacterial strains, a brief introduction into the biosynthesis of cell wall components in Gram positive strains is given here. Insights into the individual compounds are described subsequently.

The bacterial cell wall is a complex structure with a variety of functions. Aside from the capsule, it displays the outer layer of Gram-positive bacteria with direct contact to the environment. Thus, it provides essential structures for bacterial viability as a protectant against hostile environments, it is responsible for the overall shape of the organism, withstands turgor of up to 1.5 MPa, and provides structures for the interaction with nearby organisms, which may also display a virulence factor.^[62] Furthermore, the unique building blocks of the cell wall and the associated proteins expose receptor sites for viruses, antibiotics and immune cells.^[63] Gram-positive bacteria exhibit cell walls with a thickness of 30 to 100 nm, in contrast to Gram-negative bacteria. Here, a reduced cell wall of few nm is located between an inner and an outer membrane,^[62] where the latter is exposed to the environment and inherits most of the cell wall's functions.

In all bacteria, the cell wall is mainly composed of a peptidoglycan (PGN) meshwork and teichonic acids (TA). The structure of PGN is mostly conserved in all bacteria. A single unit features a disaccharide of β -1,4-linked *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc), which is a variant of GlcNAc with D-lactate bound to C₃ via an ether bond. The carboxyl group of the lactate is amidated with a pentapeptide, also called the peptide stem. The PGN layers of the cell wall are constructed by repeating linear units of the disaccharide and crosslinking of the stem peptides between the different layers.^[64] Variations of PGN mostly occur in the stem peptide, whereas the glycans are conserved across all bacterial species.^[65] In Gram positive bacteria, the stem peptide is composed of the following amino acids, starting from the lactate residue with possible variations in almost all positions: L-alanine, D-isoglutamic acid (d-iGlu), L-lysine (which is linked to the γ -amino group of the latter), D-alanine, D-alanine. In most Gram-negative bacteria, L-Lys is substituted by *meso*-diaminopimelic (*mDAP*) acid.^[66] Their additional amino groups are utilized for cross-linking, such as a pentaglycine bridge in *S. aureus*.^[67]

Cell wall biosynthesis is organized into three steps (see Figure 1.6): (1) The peptidoclycan is synthesized in the cytoplasm. Starting from uridine diphosphate-*N*-acetylmuramic acid (UDP-MurNAc), which is synthesized from fructose-6-phosphate by several cytosolic enzymes. This precursor is converted to UDP-GlcNAc by the enzymes MurAB by etherification with D-lactate, following sequential addition of the tripeptide Ala-iGlu-*mDAP* (MurC-E). The terminal Ala-Ala dipeptide is synthesized separately and bound to the carboxyl group of *mDAP* by MurF.^[68] Binding

of the precursor to undecaprenyl pyrophosphate ($C_{55}PP$) on the cytosolic leaflet of the membrane with release of UDP yields lipid I,^[69] which is converted to lipid II by MurG by β -1,4-glycosylation with GluNAc.^[70] (2) Lipid II is then flipped from the inner to the outer leaflet of the membrane, the exact mechanism is still unknown.^[71] (3) Lipid II is incorporated into the existing PGN layers by different kinds of penicillin-binding proteins, catalysing the necessary transglycosylation and transpeptidation reactions. $C_{55}PP$ is released, dephosphorylated to undecaprenyl monophosphate ($C_{55}P$) and transported past the membrane. In the cytoplasm, it is recycled to $C_{55}PP$ for the next round of PGN biosynthesis.^[72]

The $C_{55}PP$ - $C_{55}P$ system is also used to translocate TAs and saccharides across the cell membrane.^[73] TAs are integrated into the cell wall either as lipoteichonic acids (LTA), which are still linked to the undecaprenyl unit as a membrane anchor, or as wall teichonic acids (WTA) to mediate specific features to the cell wall. TAs display a strong negative charge due to their constitution from phosphodiester linked units, which is crucial for their versatile roles. For example, teichonic acids of both types protect against environmental stress and harmful cationic antibiotics.^[62] An important role is also the regulation of autolysins, an enzyme class that decomposes the PGN meshwork for cell division and cell wall recycling.^[74] Another $C_{55}PP$ - $C_{55}P$ dependent pathway is the biosynthesis of the capsule, which represents the outermost layer of some Gram-positive strains. It is composed of polysaccharides and/or polypeptides to form a gelatinous slime around the bacterium. The capsule's composition and structure are highly variable between different taxa, but it is often considered as a pathogenic factor, as it protects against phagocytosis and complement-mediated lysis by the immune system.^[62,75]

The cell wall as such does not exist in eukaryotic organisms, so a precise targeting of this unique structure should provoke only mild side effects. One prominent example that exploits this process are β -lactam antibiotics, which specifically inhibit penicillin binding proteins, responsible for the cross linking of PGN units. The class of β -lactam antibiotics evolved in nature, so the respective resistance mechanism coevolved and rapidly spread with the worldwide clinical use of penicillin and its derivatives. β -lactam resistance is easily mediated, for example by alterations of the proteinogenic target structure.^[76] A more perseverative antibiotic activity is archived by molecules that directly target PGN precursors like lipid II and $C_{55}PP$, as the nonproteinogenic target structure cannot be altered without major restraints regarding its functionality. Due to its central role in PNG-, TA-, and capsule biosynthesis, combined with its overall low abundance in the cell, $C_{55}PP$ is generally regarded as an auspicious target for antibiotic development.^[77] Additionally, enzymes throughout the whole biosynthetic machinery need to coevolve to accept new substrates and yield an altered product. The glycopeptide vancomycin was isolated and described in 1954 from *Amycolatopsis orientalis* and exhibits exceptional activity against many problematic Gram-positive strains. Inevitably, resistance against glycopeptides coevolved 10,000s of years ago.^[78]

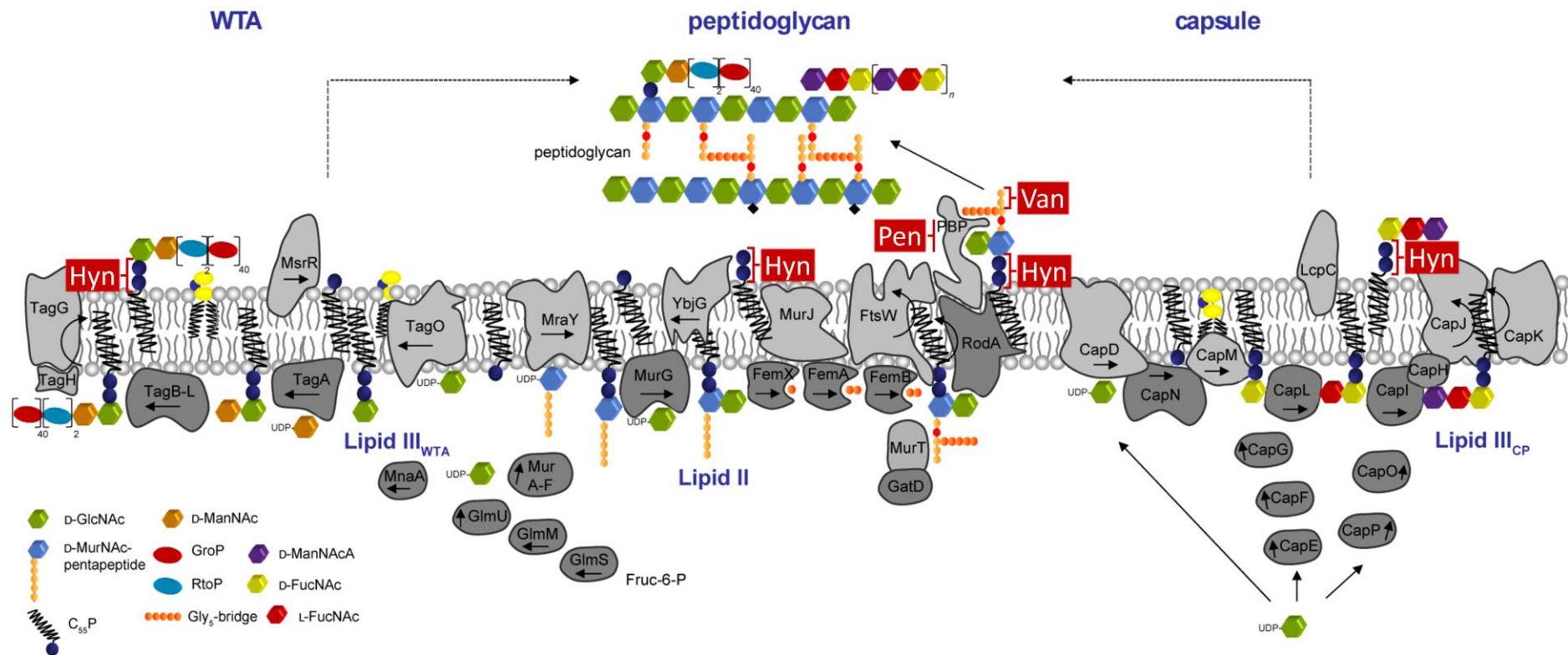


Figure 1.6: Biosynthesis of lipid II dependent cell wall components wall-teichonic acids (WTA), peptidoglycan, and capsule in *S. aureus*. Targets of discussed antibiotics are indicated. Hyn: hypeptin (also: teixobactin and lysobactin), Van: vancomycin, Pen: penicillin. CP: capsular polysaccharide, GlcNAc: *N*-acetylglucosamine, NurNAc: *N*-acetylmannosamine, ManNAcA: *N*-acetylmannosaminuronic acid, FucNAc: *N*-acetylglucosamine, GroP: glycerol phosphate, RtoP: ribitol phosphate, Gly₅: pentaglycine, UDP: uridine-5'-diphosphate, Fruc-6-P: fructose-6-phosphate. Figure taken from Wirtz, Ludwig *et al.*,^[34] with kind permission from Kevin Ludwig.

Nevertheless, efficient resistance is only achieved by alteration of the targeted D-Ala-D-Ala terminus of the PGN pentapeptide to D-Ala-D-Lac or D-Ala-D-Ser, which requires the energy-consuming translation of at least five enzymes.^[79] The worldwide spread of vancomycin-resistant strains to concerning levels thus took roughly 30 years.^[80] Lipid II-targeting is a widespread mechanism in nature to handle bacterial invasions, emphasized by a plethora of secondary metabolite classes like depsipeptides, lantibiotics, defensins, and bacteriocins.^[81] Interestingly, resistances apart from the use of glycopeptides have not yet been observed. In 2015, the depsipeptide teixobactin from *Eleftheria terrae* was reported to target lipid II, making it an effective antibiotic, also against vancomycin resistant Gram-positive strains.^[82] Unlike other described lipid II-binding antibiotics, teixobactin does not disrupt the cell membrane or the cell wall *per se*, but initiates a complex cascade that finally leads to lysis of bacterial cells, even in stationary growth phase. Further investigations revealed that teixobactin primarily binds to the pyrophosphate and the *N*-acetyl muramic acid moiety of lipid II to form μm -large clusters on the membrane surface, thereby withdrawing this low abundant intermediate from all dependent biosynthetic pathways. Interestingly, teixobactin exhibits only moderate affinity towards lipid II in anionic membranes, which is a native state due to WTAs and the clusters are formed within hours, suggesting a more complex mechanism of action than originally expected.^[83,84]

Despite intensive research, no resistance towards teixobactin has yet been reported, raising hope for the development of new antibiotics with similar mechanism of action to re sharpen the knife for the global fight against bacterial infections.

1.3.1 Hypeptin

The antibiotic hypeptin was first isolated in 1989 from a *Pseudomonas* strain. Despite its promising activity against Gram-positive strains, its structural characterization was not fully published and since then no further research on hypeptin was conducted.^[85] It is a cyclic octadepsipeptide with a four-membered ring, composed of three R-configured and four β -hydroxylated amino acids. Due to its structural similarities with teixobactin, hypeptin was already proposed to be a lipid II-binding antibiotic,^[86] but this was not further investigated until its recent rediscovery in a *Lysobacter* strain. The in-depth investigation of its biosynthesis and structure elucidation was part of this doctoral thesis and is described in detail in section 3.1. The activity, as well as the mechanism of action of hypeptin was investigated by Kevin Ludwig (working group of Prof. Tanja Schneider, Institute of Pharmaceutical Microbiology, University Hospital Bonn) and reported in a joint publication.^[34] The results of his work are summarized here.

The already reported excellent activity of hypeptin against all Gram-positive bacterial strains was reproduced. Noticeably, it was also active against problematic nosocomial strains, including

methicillin-resistant *Staphylococcus aureus* (MRSA), vancomycin resistant *Enterococcus* (VRE), daptomycin resistant *S. aureus* (DAPR) and against mycobacteria. Gram-negative strains such as *E. coli* and *Pseudomonas* were all resistant, except for the susceptible *E. coli* strain MB5746 with its permeable outer membrane, indicating that this additional barrier mediates resistance. Hypeptin effectively killed bacteria in early and, in contrast to many other antibiotics, also in late exponential growth phase. This effect is limited to bacterial cells, as human cell lines, and red blood cells were not affected by hypeptin and showed lytic tendencies only after supplementation of several MIC folds. The targeted pathway was identified to be the cell wall biosynthesis by a screening with reporter strains and phase contrast microscopy, showing membrane blebs, which originate from a disrupted cell wall. Additionally, accumulation of the cell wall precursor UDP-MurNAc was detected in hypeptin-treated cells, indicating the inhibition takes place in late-stage cell wall biosynthesis. *In vitro* assessment of single biosynthetic steps from cell membrane preparations of *Micrococcus luteus* revealed that hypeptin traps C₅₅PP-containing intermediates by complex formation in a ratio of 2:1 (hypeptin:intermediate). It was shown that the C₅₅PP structure is crucial for complex formation, but strong association is mediated by the first sugar moiety, whereby the exact nature of the sugar is not relevant. These results resemble the published data about teixobactin that reported fast dimerization of the antibiotic prior to target association, with noticeable contribution of the sugar moiety to the association with lipid II, validated by solid-state nuclear magnetic resonance (ssNMR) measurements. These findings suggested that hypeptin influences the biosynthesis of teichonic acids and capsule intermediates that also rely on the C₅₅PP-mediated transport across the membrane. The effect on WTAs was supported by reduced cell lysis of an *atlA* deficient strain upon treatment with hypeptin. Natively, liberation of the major PGN-decomposing autolysin AtlA is tightly regulated by binding to WTAs, so it was assumed that hypeptin impairs TA production and thereby liberates toxic concentrations of autolysins.

The effect on at least two target pathways probably causes synergistic effects. Liberation of autolysins weakens the overall cell wall structure that cannot be repaired due to blocked *de novo* biosynthesis of PGN. Further mechanisms like the impairment of capsule biosynthesis are probable, but were not considered for further research.

1.3.2 Lysobactin

Lysobactin and katanosin B were almost simultaneously described by two individual research groups in 1988 after isolation from *Lysobacter* sp. ATCC53042 (lysobactin)^[87] and *Cytophaga* PBJ-5356 (katanosin B).^[88] Both compounds are almost identical. They are described as depsipeptides, composed of a nine-residues macrocycle, coupled to a lipophilic linear dipeptide. Two of the amino acids are R-configured and three are β -hydroxylated. The reports differ in the exact configuration of a threonine residue. In lysobactin, it is assigned to be L-*allo*-Thr (2*S*,3*S*), whereas it is D-*allo*-Thr (2*R*,3*R*) in katanosin B. In most publications this difference is neglected, so lysobactin and katanosin B are used as synonyms. As publications about katanosin B from *Cytophaga* sp. PBJ-5356 are scarce, all given data refer to lysobactin from *Lysobacter* sp. ATCC53042 with validated (2*S*,3*S*)-Thr assignment.

Lysobactin exhibits exceptional activity against strains of resistant Gram-positive bacteria, but does not perturb growth of Gram-negative strains like *E. coli*.^[87] Similar to hypeptin, it kills bacteria in stationary growth phase by induction of cell lysis, without harming human red-blood cells at reasonable concentrations. In investigations of a binding target, lysobactin was shown to directly bind to C₅₅PP- containing intermediates of cell PGN and TA biosynthesis such as lipid I, lipid II, and lipidII_A. The latter resembles lipid II, but harbours β -1,4-linked GlcNAc and MurNAc without any peptides. In contrast to teixobactin and hypeptin, no homodimerization is required, so lysobactin binds in a 1:1 ratio to its target, without subsequent formation of higher-order aggregates.^[89] Despite the promising antibiotic and cytotoxic parameters, lysobactin never entered clinical trials due to questionable cytotoxicity in primarily conducted mouse models and decisions based on economic considerations.^[90]

The proposed lysobactin BGC *lyb* is composed of seven genes, *lybA*, *lybB*, and *orf78-82*. *LybA* and *lybB* encode for NRPS with four and seven modules, respectively that harbour two interesting features. First, the threonine- recruiting module eight contains an additional C domain that was postulated to catalyse C₃ epimerization to yield (2*S*,3*S*)-Thr as a building block. Secondly, *LybB* is terminated by a tandem TE domain. Biochemical investigations revealed that TE1 is solely responsible for thioester cleavage of the mature peptide from the adjacent T domain by macrolactonization to the hydroxyl group of the phenylserine moiety. TE2 was characterized as a type II TE domain that unspecifically cleaves thioester-bound substrates off T domains, to remove shunt products and prevent obstructions of the assembly line. It was observed that the linking region between TE1 and TE2 is prone to cleavage mediated by unspecific peptidases, raising the hypothesis that TE2 is natively cleaved off the mature enzyme to act as free type II TE domain.^[35]

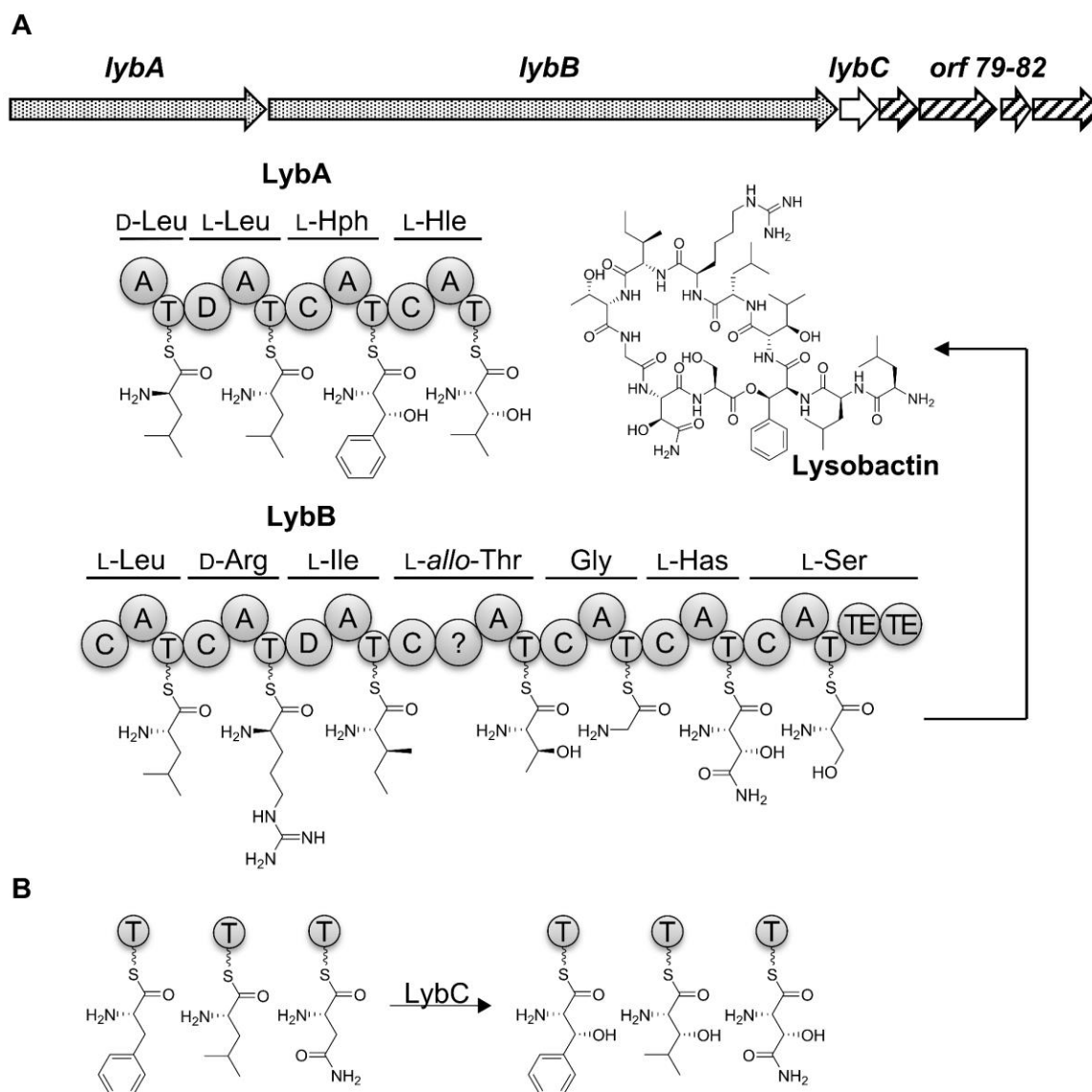


Figure 1.7: Biosynthesis of Lysobactin. **A:** Organization of the *lyb* biosynthetic gene cluster (dotted: NRPS; white: auxiliary biosynthetic enzymes, dashed: resistance and transporter-related genes) and architecture of the encoded non-ribosomal peptide synthetase (Hph: (2*S*,3*R*)-3-hydroxyphenylalanine; Hle: (2*S*,3*R*)-3-hydroxyisoleucine; Has: (2*S*,3*S*)-3-hydroxyasparagine; A: adenylation domain, C: condensation domain, D: dual condensation/epimerization domain; T: thiolation domain; TE: thioesterase domain). **B:** The non-heme diiron monooxygenase LybC putatively β -hydroxylates T domain-bound L-Phe, L-Leu and L-Asn in the respective modules.

The remaining genes, *orf78-82* were only investigated bioinformatically. *Orf79-82* were annotated as resistance mediating and transporter-related enzymes. Interestingly, *orf78* was initially annotated as β -lactamase, so the authors deduced it to be responsible for self-resistance, but the accountable β -lactam fold is the core motive to the superfamily of metallo- β -lactamases (MBL), catalysing oxidative reactions.^[55] By now, *orf78* is generally considered to encode for a *trans* acting NHDM, thus, *orf78* is reannotated to *lybC*.^[60] The character of the putative hydroxylase LybC poses several questions, as it would be responsible for the β -hydroxylation of L-Leu, L-Phe, and L-Asn in modules three, four, and ten of the NRPS. The coordination of these structurally diverse amino acids in the

substrate binding pocket demonstrates general promiscuity towards the amino acid substrate, suggesting it as a promising tool to introduce hydroxylation events in artificially constructed NRPS assembly lines. Additionally, lysobactin is composed of four leucine residues, but only the third residue is specifically recognized as substrate by LybC, indicating that the structural features for substrate recognition are not encoded by the substrate amino acid. These proposed unique characteristics of LybC make it a promising research target. Their validation would greatly contribute to the comprehension of in *trans* acting NHDM and give further implications for their utilization in artificially constructed NRPS systems.

1.4 FR900359

1.4.1 Structure and activity

The chromodepsin FR900359 (FR) was first described in 1986 after isolation from *Ardisia crenata*.^[91] This evergreen plant is mostly cultivated for ornamental purposes, but was also used in traditional Chinese medicine to treat various ailments like respiratory tract infections, menstrual disorders and tonsillitis.^[92] FR is a cyclic octadepsipeptide, mostly composed of non-proteinogenic amino acids. The cyclic part harbours D-phenylactic acid, *N*-methyl-L-dehydroalanine, L-alanine, *N*-methyl-L-alanine, *N,O*-dimethyl-L-threonine, and two β -hydroxy-L-leucines (Hle). From the latter, one residue is additionally *N*-acylated, whereas the hydroxy group of the second is esterified to the side chain *N*-propionyl-(2*S*,3*R*)-3-hydroxyleucine (*N*-Pp-Hle). The overall structure, especially the macrocycle pointed towards a bacterial or fungal origin of FR. Indeed, in 2015 the actual producer was identified to be the endosymbiotic bacterium “*Candidatus burkholderia crenata*”, which is predominantly located in the leaf nodules of the plant.^[93,94]

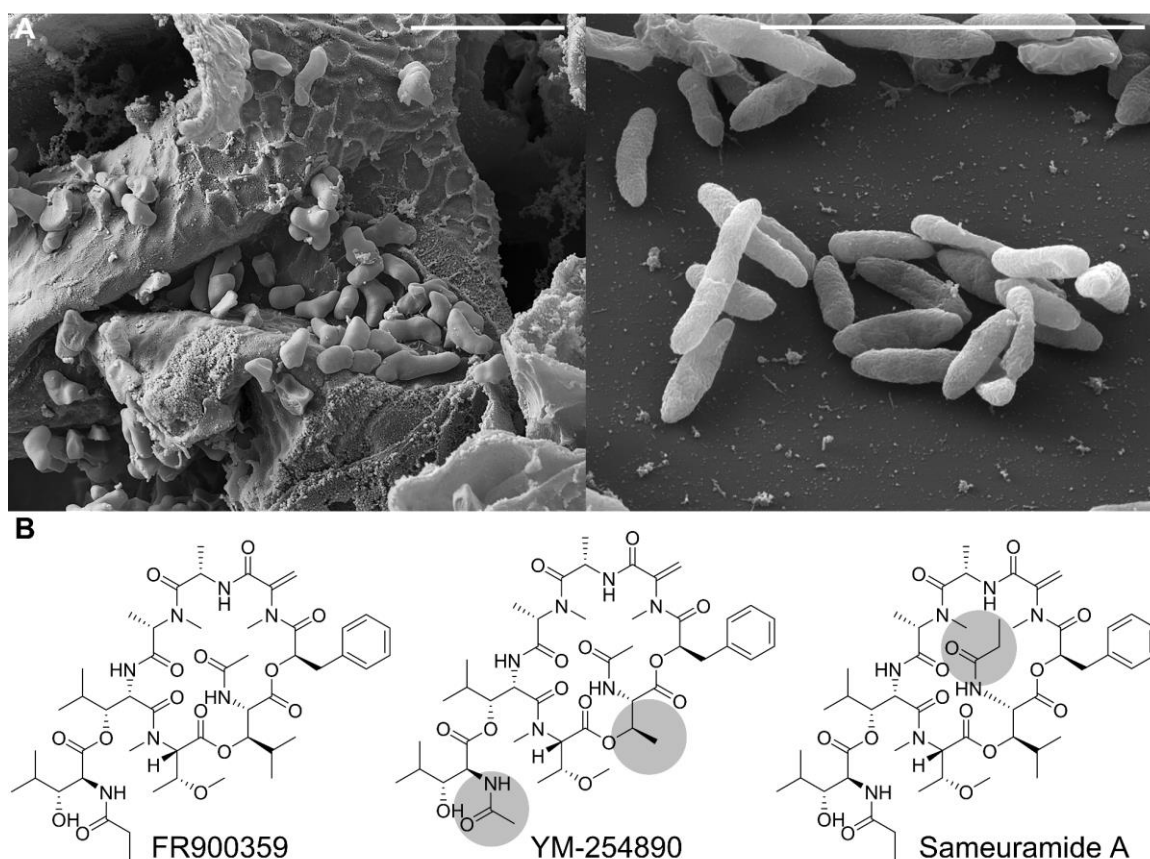


Figure 1.8: The chromodepsins. A Left: The natural producer of FR900359 (FR), “*Candidatus Burkholderia crenata*” inhabits intercellular spaces in the leaf nodules of the host plant *Ardisia crenata*. Right: The FR producing *Chromobacterium vaccinii* on LB-agar. Scale bars represent 5 μm . Pictures kindly provided by Eye of Science. B: Structures of described chromodepsins. The structural differences to FR are highlighted in grey.

By this time, the analogous chromodepsins YM-254890 from *Chromobacterium* sp. and Sameuramide A were described (Figure 1.8).^[95] The only differences are the substitution of *N*-acyl-Hle by *N*-acyl-L-threonine in YM-254890 and the length of the acyl residues. Interestingly, FR was also found to be produced by the soil bacterium *Chromobacterium vaccinii* MWU205.^[13] The access to an autarkic and easily cultivatable bacterial producer of FR facilitated the *in vitro* characterization of its biosynthetic enzymes, genome editing, and precursor feeding to generate intermediates and derivatives. Since then, the compounds FR2-5 were isolated and characterized, but feature-based MS/MS networking already indicated the occurrence of 30 natural derivatives in the extracts from *C. vaccinii* and *A. crenata*. Most of them are only produced in trace amounts.^[96]

The target structure of FR was rapidly identified as the α subunit of Gq proteins.^[97] These proteins consist of two further subunits (G β q, G γ q) and are associated with the cytosolic leaflet of transmembrane G protein coupled receptors (GPCR). Once an activating ligand binds to the outer part of a GPCR, conformational rearrangements induce the exchange of guanosine diphosphate (GDP) to guanosine triphosphate (GTP), which is bound to G α of the associated protein. This reduces the overall stability of the heterotrimeric complex and stimulates the dissociation of the G $\beta\gamma$, which in turn initiates intramolecular signal cascades. Once FR binds to G α , it impedes the GDP/GTP exchange by enhancing the protein's rigidity and thus inhibits downstream effects. FR inhibits Gq mediated signalling efficiently (IC₅₀=0.45 μ M) and selectively, as the activity of the analogous G proteins Gs, Gi, and G12 is not influenced.

1.4.2 Utilization as a pharmacological tool, clinical and ecological function

Investigations into the signalling pathways of specific G proteins are hindered by the lack of appropriate small molecule modulators. For Gs, the cholera toxin, and the small molecule suramin are used as tools, but both harbour disadvantageous properties.^[98] The discovery of FR and YM, and their commercial availability facilitated investigation of the role of G α q in diverse pathways, regarding immune response,^[99] ossification,^[100] and GPCR-mediated signalling in general.^[101] G α q is coupled to diverse GPCRs, distributed across the whole body of vertebrates, invertebrates and many more. Thus, FR has high potential as a pharmacological research tool but the exact extent remains to be determined.

On the other hand, the ubiquity of its target narrows the clinical spectrum of FR. *In vivo* and *in vitro* studies with FR and YM demonstrated their stark influence on the cardiovascular system by a drastic reduction of both blood pressure and antithrombotic effects.^[93,102] Due to the small therapeutic window, these approaches were stopped and researchers continued to explore the effects of topic application. Consequently, FR was shown to inhibit G α q-mediated bronchoconstriction that is initiated by diverse GqPCRs, making it feasible to treat the symptoms of asthma.^[103]

Additionally, it halted Gαq-dependent growth and metastasis of uveal melanoma in mouse models.^[104] Lastly, the thermogenesis in brown and beige adipose tissue is Gαq-regulated and therefore susceptible to FR manipulation, suggesting possibilities to treat obesity and diabetes.^[105]

Members of the chromodepsins have been found in organisms of diverse ecological systems such as plants,^[91] soil bacteria^[13] and didemnid ascidians.^[106] This raised questions about their actual ecological relevance for their respective host. The strong effect, mediated by a ubiquitous target, pointed towards a defensive strategy, *e.g.* the plant *A. crenata* accumulates FR in its leaf nodules to devitalize predators. This theory was supported by feeding experiments with the stink bug *Riptortus pedestris*, which showed a dose-dependent killing after a delay phase of four to six days.^[93] The influence of FR, produced by *C. vaccinii* on soil-dwelling organisms like nematodes is subject to current research.

1.4.3 Biosynthesis

The FR BGC *frs* was identified on an extrachromosomal plasmid of the leaf symbiont “*Ca. Burkholderia crenata*”^[94] and its function unambiguously verified by heterologous expression in *E. coli*.^[93] The *frs* BGC of *Chromobacterium vaccinii* MWU205 only differs in the GC content and the length of the intergenetic regions.^[13] Due to the minor relevance of the *frs* BGC from “*Ca. Burkholderia crenata*” for biochemical characterization, the term *frs* refers to the BGC of *C. vaccinii* MWU205 in this thesis.

Five of the eight *frs* genes encode for NRPS (*frsA*, *frsD-G*), two for tailoring enzymes (*frsC*, *frsH*) and one for a MbtH-like protein (*frsB*). MbtH-like proteins (MLP) are often necessary to maintain the structure and catalytic activity of A domains in NRPS. Bioinformatic analyses of the NRPS were in line with the conferred structure of FR and were able to predict the origin of the highly decorated non-proteinogenic amino acids. Modules four, six, and eight harbour MT domains, which correspond with the presence of the respective methylated amino acids. Interestingly, module eight includes an A domain, interrupted by two MT domains, which is a scarcely described phenomenon.^[28] The C domains of FrsA and FrsD were annotated as C_{starter} domains, which catalyse *N*-acylation of the respective L-Leu residue. Analysis of the specificity-conferring code of module three in FrsE did not suggest any amino acid. However, the tailoring enzyme FrsC showed similarities to malate and L-lactate dehydrogenases, so it was deduced that it catalyses the formation of L-phenylactic acid (PLA) from phenylpyruvate, an intermediate in phenylalanine biosynthesis. PLA would be substrate to the A domain in module four and subsequently epimerized by the encoded E domain. The additional tailoring enzyme, FrsH exhibited only minor homology to any described enzyme, but detailed analysis suggested it to be a homologue of the *trans* acting NHDM

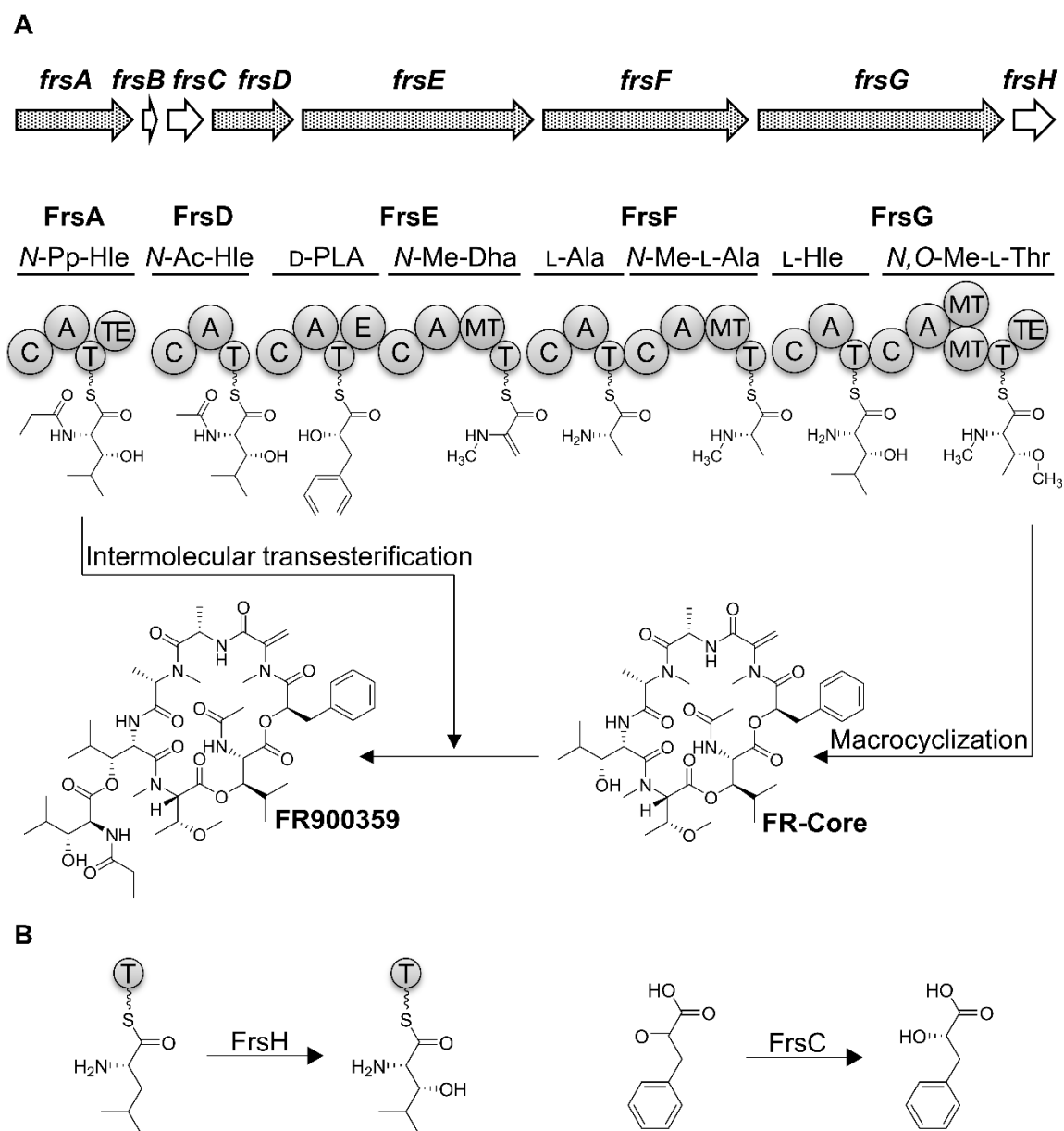


Figure 1.9: Biosynthesis of FR900359 (FR). **A** Organisation of the *frs* biosynthetic gene cluster (grey: NRPS; white: auxiliary genes) and architecture of the encoded non-ribosomal peptide synthetase (A: adenylation domain, C: condensation domain, E: epimerization domain; MT: methyltransferase; T: thiolation domain; TE: thioesterase). The TE domain of FrsG catalyzes macrocyclization of the bound linear peptide to release the cyclic intermediate FR-Core. The final biosynthetic step is then conducted by the TE domain of FrsA, transesterifying the side chain to FR-Core. (*N*-Pp-Hle: *N*-propionyl-(2*S*,3*R*)-3-hydroxyleucine; *N*-Ac-Hle: *N*-acetyl-(2*S*,3*R*)-3-hydroxyleucine; PLA: phenylactic acid; Dha: dehydroalanine). **B** Formation of hydroxylated building blocks. FrsH β -hydroxylates T domain-bound L-Leu in the modules one, two, and seven. FrsC generates L-PLA from free phenylpyruvate.

CmlA. Thus, FrsH would be responsible for the installation of β -hydroxy groups in the L-Leu residues of module one, two, and seven. Interestingly, two type I TE domains are encoded in the BGC, separating the NRPS into the two independent systems FrsA and FrsDEFG. This led to the assumption that FrsDEFG are responsible for the biosynthesis of the cyclic intermediate peptide FR-Core, which is finalized by a transesterifying reaction, mediated by the specialized TE domain of FrsA. Finally, the biosynthesis of Dha in module four remained enigmatic. The specificity-

conferring code of the A domain hints towards the recruitment of L-Ser, which requires subsequent dehydration to Dha. Dual condensation/dehydration domains were already described,^[107] but bioinformatic analysis of the respective C domain in module five gave no hints towards such a catalytic activity.

During the time period of this thesis, several members in the working group of Prof. Dr. G. König and Dr. Max Crüsemann devoted their work to biochemically decipher the single steps of FR biosynthesis and to confirm the described bioinformatic hypotheses. Dr. René Richarz investigated the origin of Dha, in the course of which he established a knock-out and complementation system for *C. vaccinii*, which contributed heavily to most other projects within the group.^[13] Dr. René Richarz and Sophie Klöppel validated the proposed activity of FrsC (manuscript in preparation). Dr. Cornelia Hermes demonstrated the biosynthesis of the side chain by FrsA together with FrsH, and its transesterification to FR-Core by the specialized FrsA TE domain.^[13] The characterization of FrsH was the focus of this work and is described in detail in section 3.3.

2 Aims of the study

Non-ribosomal peptide synthetases (NRPS) produce a major group of natural products with significance for human society as drugs or toxins. The biosynthesis of several clinically used drugs, as well as a plethora of unnamed compounds include members of the family of *in trans* acting non-heme diiron monooxygenases (NHDM). The abundance of this enzyme family mismatches the little effort that has so far been devoted to its characterization. Previously, only one member has been experimentally characterized. Within this thesis, NHDMs of three different NRPS biosynthetic gene clusters (BGC) shall be characterized to deepen the overall knowledge about this underreported enzyme family. The results, together with further proposed experiments outlined below are expected to shed light into the biosynthesis of three complex non-ribosomal peptides.

2.1 Hypeptin

The cyclic depsipeptide hypeptin exhibits potent antibacterial activity and consists of in total four β -hydroxylated amino acids with differing stereoconfiguration. Hypeptin was recently isolated from *Lysobacter sp.* K5869, and prior to this work the respective genome has been sequenced. To shed light on the biosynthesis of hypeptin, which is not known to date, the respective BGC shall be identified and analysed *in silico* to detect interesting features. Subsequently, the origin of four β -hydroxylated amino acids shall be investigated. Therefore, the two *in trans* acting hydroxylases HynC and HynE, as well as the respective NRPS modules will be heterologously expressed and reconstituted *in vitro*. The substrate specificity of the hydroxylases and their cognate modules will help to revise the configuration of hydroxy groups in hypeptin. which will then be verified by in-depth NMR analyses.

2.2 Lysobactin

The BGC of the potent lipid II binding antibiotic lysobactin from *Lysobacter sp.* ATCC53042 encodes for the putative NHDM LybC, which is proposed to β -hydroxylate three structurally diverse amino acids *in trans* to the NRPS system. This project will focus on the verification of the proposed function of LybC, which was not proven so far. Thus, the respective NRPS modules shall be heterologously expressed and the activity of the embedded A domain tested *in vitro*. Active NRPS modules will then be utilized to assess the activity of LybC.

2.3 FR900359

This project shall investigate the mechanism of substrate recognition of the NHDMs. So far, several clues point towards a crucial contribution of the A domain that has not been researched yet. The biosynthesis of the side chain of FR900359 (FR) shall serve as model for the investigation thereof. FR is a cyclic depsipeptide, produced by the soil bacterium *Chromobacterium vaccinii*, which exhibits exceptional inhibitory activity and selectivity towards Gαq proteins. The monomodular NRPS FrsA and the NHDM FrsH are together responsible for the biosynthesis of the FR side chain *N*-propionylhydroxyleucine (*N*-Pp-Hle). To reconstitute this biosynthesis *in vitro*, different constructs of FrsA will be generated and their adenylation activity assessed. Subsequently, the proposed ability of FrsH to generate *N*-Pp-Hle, together with the respective construct of FrsA will be verified in an *in vitro* side chain assembly assay.

Once this assay is established, the obligatory protein-protein interaction between the two responsible enzymes will be qualitatively and quantitatively analysed by size-exclusion chromatography and isothermal titration calorimetry. To get deeper insight into the driving forces of the interaction, a crystal structure of FrsH shall be generated. For this purpose, a native form of FrsH will be isolated in exceptional purity, submitted to crystallization trials and the grown crystal used for X-ray diffraction spectrometry. The thus obtained structure of FrsH will be used for *in silico* docking to the A domain of FrsA (FrsA_{1A}) to detect possible binding interfaces.

Additionally, the exchangeability of NHDM from different biosynthetic systems shall be tested by complementation of an *frsH*-deficient strain of *C. vaccinii* with *hynC* and *lybC*. The ability of the complemented strains to produce FR will be recorded to verify the functional substitution of FrsH by the homologous NHDM HynC and LybC. The evidence, which arises from these results, will permit to compare computationally generated three-dimensional structures of the hydroxylases to detect functionally conserved surface patches, which might facilitate the interaction with FrsA.

Lastly, in-depth analyses of the *in silico* docking results and the structural comparison of the homologue NHDMs will highlight distinct residues, which most likely contribute to the interaction of the enzymes. To assess their relevance for the interaction, the respective residues will be rendered by mutagenesis in FrsA and FrsH. The hereby generated enzymes will be assessed for their ability to generate hydroxylated product in the side chain assembly assay *in vitro*.

3 Results

3.1 Hypeptin

The isolation and subsequent cultivation of bacterial and fungal strains in the laboratory is hampered by the great plate count anomaly. This term describes the problem, that only 1 % of the microorganism, which are observable with the microscope, actually form visible colonies on agar plates or grow densely in liquid media.^[108] To overcome this issue, the working group of Prof. Dr. Kim Lewis developed a method to isolate and cultivate bacterial strains in their native environment. The resulting iChip technology enabled growth of about 50 % of soil bacteria.^[109] The potential of this method was demonstrated by the isolation of the previously uncultivated β -proteobacterium *Eleftheria terrae*, which finally led to the isolation of the promising antibiotic teixobactin.^[82] In the same screen, another extract showed promising activity against gram-positive reporter strains. Activity-guided fractionation of the extract by Paul Barac and Carina Marx, followed by LC-MS and NMR analyses identified hypeptin, previously isolated by Shionogy & CO in 1989,^[85] as the responsible bioactive compound. Genomic DNA of the strain K5869 was isolated and submitted to Eurofins Genomics for sequencing *via* PAC-BIO SMRT cell technology by Paul Barac and Dr. Max Crüsemann. The assembly yielded a single, circular scaffold of 5,986,034 bp with a GC content of 69 %.

Hypeptin is a cyclic depsipeptide, consisting of eight amino acids. Apart from the cyclic structure, the high amount of six non-proteinogenic amino acids suggests a non-ribosomal origin for hypeptin. In particular, the high abundance of four β -hydroxylated amino acids with different configuration and side chain structures poses questions about their exact formation. In this section, key steps of hypeptin biosynthesis were investigated with bioinformatic tools and *in vitro* characterization of heterologously expressed enzymes. The resulting data were published in a collaborative publication, together with the working group of Prof. Dr. Tanja Schneider from the Institute for Pharmaceutical Microbiology.^[34]

3.1.1 Phylogenetic analysis of 16S-rRNA DNA

The genus of procaryotic species is commonly identified by phylogenetic analysis of its 16S-RNA sequence. Its prevalence in all organisms, combined with the low evolutionary pressure makes it a universal molecular clock to reveal the relationship of unclassified bacterial organisms. The genome of *Lysobacter sp.* K5869 harbours two identical gene loci coding for 16S-RNA. Analysis of these genes with the EzBioCloud^[110] revealed a very high sequence identity with genes from different *Lysobacter* strains (Table 3.1), with the highest identity towards *L. enzymogenes*. The construction

of a phylogenetic tree, supported by 100x bootstrapping further corroborated this result (Figure 3.1). Unfortunately, the phylogenetic tree also shows that the individual taxa of the *Lysobacter* genus are mostly constituted from single strains, making the classification by 16S-rRNA error-prone. More data about the phenotype and catalytic or metabolic activity of the examined strain are needed to reliably classify its genus.^[111] Thus, it was not further classified and named *Lysobacter* sp. K5869.

Table 3.1: Closest relatives of the newly identified strain *Lysobacter* sp. K5869, according to comparison of 16S-RNA DNA sequences against EzBioCloud database.

Rank	Name	Accession	Pairwise Similarity [%]
1	<i>Lysobacter enzymogenes</i> ATCC 29487	jgi.1095734	99.780
2	<i>Lysobacter enzymogenes</i> C3	CP013140	99.727
3	<i>Lysobacter firmicutimachus</i> PB-6250	KU593484	98.841
4	<i>Lysobacter gummosus</i> KCTC 12132	AB161361	98.023
5	<i>Lysobacter antibioticus</i> ATCC 29479	CP013141	97.960

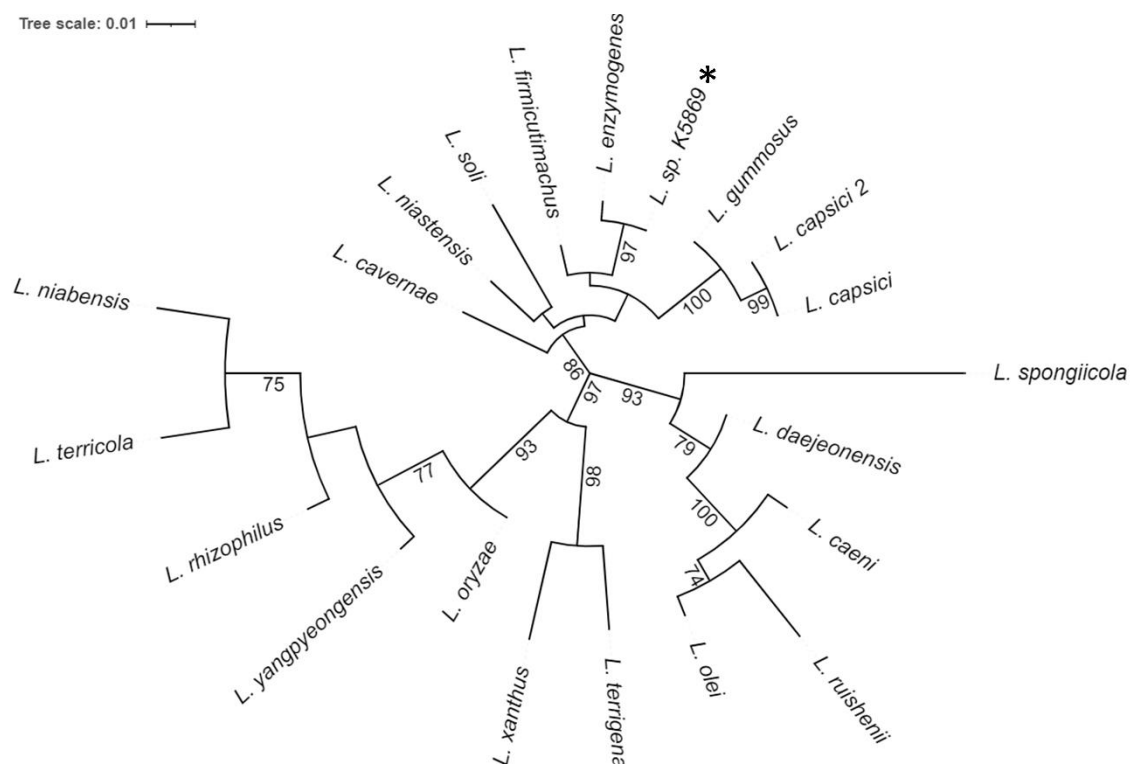


Figure 3.1: Phylogenetic tree of *Lysobacter* 16S RNA sequences. Sequences of the 20 closest related strains of *Lysobacter* sp. K5869 (asterisked) were extracted from the EzBioCloud database and subtracted to further analysis. The clades are supported by 100 iterations of bootstrapping. All values >70 are displayed. *L. sp. K5869* closest relative belongs to the genus *enzymogenes*, but the clade needs more branches to verify its classification.

3.1.2 Description of the biosynthetic potential of *Lysobacter* sp. K5869

To analyse the modified genome scaffold for biosynthetic gene clusters (BGC), it was submitted to antiSMASH.^[112] AntiSMASH is a free online platform that utilizes and links several bioinformatic tools to find, annotate and analyse secondary metabolite BGCs within an uploaded DNA sequence from bacterial, fungal or plant origin.^[112] The whole genome sequence of *Lysobacter* sp. K5869 was uploaded to antiSMASH for analysis. The detection strictness for finding BGC was set to relaxed.

AntiSMASH detected 12 regions with at least one BGC that span in total over 644,053 bp (10,8 % of the genome). The genome harbours three NRPS, two PKS, three lanthipeptide, two bacteriocine and one hybrid PKS/NRPS, arylpolyene, butyrolactone, lipolanthine, and thiopeptide BGC each. One of the NRPS-BGCs shows 100 % identity with the bicornutin A1/A2-synthase gene from *Xenorhabdus budapestensis* (NCBI Acc. No: JX424818.1) and the hybrid PKS/NRPS shows 87 % identity to the heat-stable antifungal factor (HSAF) BGC from *Lysobacter enzymogenes* (EF028635.2). The arylpolyene BGC has 50% identity and an overall similar organisation as the xanthomonadin I BGC from *Xanthomonas oryzae* pv. *oryzae* (AY010120.2). Out of the other nine BGC, no high similarity to known BGCs was detected.

These findings are in line with the genome analysis of other *Lysobacter* strains. All known species of this genus harbour a genome of ~6 Mbp that contains 12-16 BGCs, of which more than 50 % are NRPS or NRPS-hybrids.^[113]

3.1.3 Description of the *hyn* biosynthetic gene cluster

Given the structure of hypeptin, it was suspected to be synthesized by an NRPS. One annotated NRPS system harboured 8 modules encoded in two consecutive genes, which corresponds to the number of amino acid residues in hypeptin. The automated analysis of its A domains by antiSMASH predicted the assembly of Ala-Leu-X-Asn-Asn-Tyr and Leu-Ile. These predictions are consistent with the overall amino acid sequence of hypeptin Ala-(D-Leu)-(D-Arg)-(OH-Asn)-(OH-D-Asn)-(OH-Tyr)-(D-OH-Leu)-(Ile). The bioinformatic analysis of the C domains revealed dual condensation/epimerisation-domain structure in module three, four, and six, which corresponds nicely to the D- configured amino acids Leu₂, Arg₃, and OH-Asn₅. Interestingly, module five harbours two C domains. Both are annotated as ^LC_L domains by NaPDos^[114] and antiSMASH (Figure 3.2).

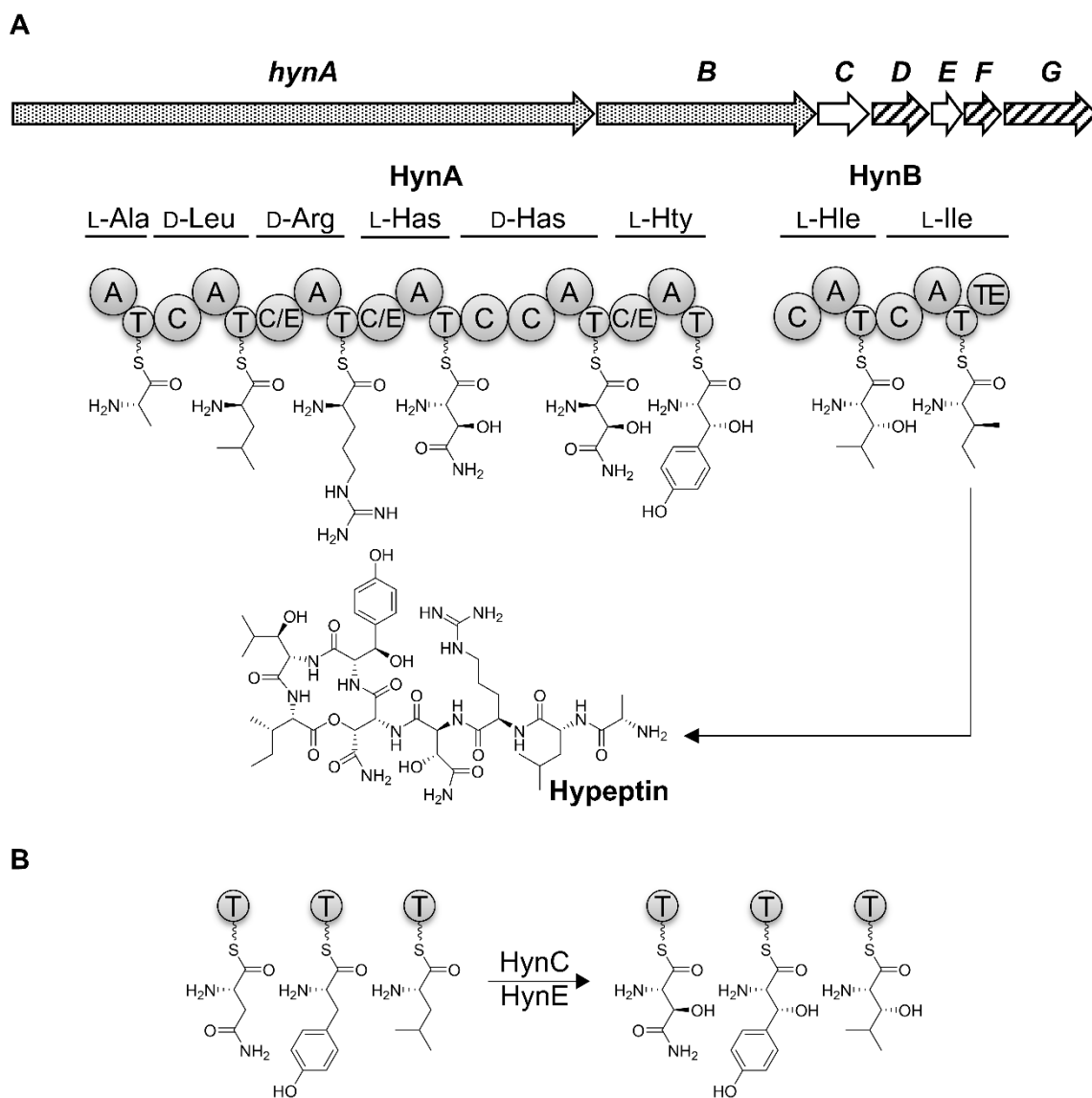


Figure 3.2: Biosynthesis of hypeptin. **A:** Organization of the *hyn* biosynthetic gene cluster (grey: NRPS, white: auxiliary biosynthetic enzymes, dashed: resistance and transporter-related genes) and architecture of the encoded non-ribosomal peptide synthetase (Has: 3-hydroxyasparagine, Hty: 3-hydroxytyrosine, Hle: 3-hydroxyisoleucine, A: adenylation domain; C: condensation domain, C/E: dual condensation/epimerization domain, T: thiolation domain, TE: thioesterase domain). **B:** The non-heme diiron monooxygenase HynC and the iron- and α -ketoglutarate dependent hydroxylase HynE hydroxylate the T domain-bound amino acids L-Asn, L-Tyr, and L-Leu. The elucidation of their substrate specificity was one aim of this study.

The NRPS is located in a region with at least two other BGCs that were assigned to be a single BGC by antiSMASH. The borders of the BGC responsible for hypeptin biosynthesis were determined manually for a more precise description and characterization. BLASTp analysis of the NRPS revealed very high sequence identities to corresponding sequences from the strain *Lysobacter psychrotolerans* ZS60, which was isolated from soil in Xinjiang, China in 2019.^[115] The publicly available genome of this strain was submitted to antiSMASH and revealed an identical architecture of seven genes encoding for the two core NRPS enzymes and five further proteins. These genes were set to be responsible for the biosynthesis of hypeptin and termed the *hyn* BGC

(see Table 3.2). Further genes up- and downstream of the *hyn* BGC of *Lysobacter sp.* K5869 have no homologues within the genome of *L. psychrotolerans* ZS60 and show significantly lower identities to genes from other species (*hynG* +1: YCII- related domain protein [*Lysobacter gummosus*] (62/78) ALN90542.1; *hynA* -1: DUF4019 domain- containing protein [*Lysobacter gummosus*] (63/76) WP_083512732.1). This indicates that the *hyn* BGC was exchanged via horizontal gene transfer between the species. Consequently, the genes *hynA* and *hynG* display the borders of the *hyn* BGC, spanning over 35596 bp.

Table 3.2: Top BLAST hits of the single proteins from the *hyn* BGC.

Name	Size (aa)	Annotated Function	Closest Homologue (ident [%]/simil [%])	Accession No. of the Homologue
HynA	6491	NRPS	Amino acid adenylation domain-containing protein [<i>Lysobacter psychrotolerans</i> . ZS60] (80/87)	WP_123088455.1
HynB	2447	NRPS	Non- ribosomal peptide synthetase [<i>Lysobacter psychrotolerans</i> ZS60] (82/88)	WP_123088454.1
HynC	531	Non-heme diiron monooxygenase	MBL fold metallo hydrolase [<i>Lysobacter psychrotolerans</i> ZS60] (92/96)	WP_123088453.1
HynD	608	ABC-transporter related protein	ATP-binding cassette domain-containing protein [<i>Lysobacter psychrotolerans</i> ZS60] (81/89)	WP_148041022.1
HynE	302	α -ketoglutarate dependent oxygenase	TauD/TfdA family dioxygenase [<i>Lysobacter psychrotolerans</i> ZS60] (83/91)	WP_123088451.1
HynF	401	RND family efflux transporter MFP subunit	Efflux RND transporter periplasmic adaptor subunit [<i>Lysobacter psychrotolerans</i> ZS60] (72/83)	WP_123088450.1
HynG	1028	AcrB/AcrD/AcrF family protein	Efflux RND permease subunit [<i>Lysobacter psychrotolerans</i> ZS60] (80/88)	WP_123088449.1

The correct three dimensional structure and catalytic function of some A domains depends on the interaction with small MbtH- like proteins (MLPs).^[8] These chaperone like proteins are often, but not always, encoded within the BGC of the respective NRPS.^[21] None of the genes of the *hyn* BGC encodes for an MLP, so a BLASTp analysis of the whole genome of *Lysobacter sp.* K5869 was performed against the MLP FrsB from the biosynthesis of FR900359. The search revealed one homologue within the genome that is also annotated as MLP by antiSMASH. The gene was named *hynMLP*. Another kind of enzymes that are crucial for the correct function of NRPS are type-II thioesterases (TE). These often neglected tailoring enzymes non-specifically cleave thioester-bound

residues of T domains. They hardly influence the function of NRPS during common elongation cycles, but are important in instances of an erroneously loaded substrate. If an A domain incidentally recruits the wrong amino acid, the NRPS machinery halts due to the proofreading activity of the respective C domain. Here, the unspecific type II TE is necessary to unblock the shunt that otherwise renders the whole NRPS useless.^[5] The genome of *Lysobacter sp.* K5869 harbours one type-II TE that was named *hynTE*.

3.1.4 Cloning and heterologous expression of *hyn* genes

The first bioinformatic investigations gave strong evidence that the *hyn* BGC is responsible for the production of hypeptin. To verify this hypothesis, the catalytic activity NRPS domains were investigated *in vitro*. The modules four, five, six, and seven were hypothesised to be responsible for the integration of hydroxylated amino acids into the peptide (Figure 3.2), so the study focussed on the heterologous expression and *in vitro* reconstitution of these four modules. The resulting module constructs are truncated parts of an integrative, larger protein and known to be difficult to be expressed in *E. coli*. To obtain soluble and functional proteins, different constructs of each module, all coexpressed with or without *hynMLP* were tested. Subsequently, the activity and substrate specificity of the putative tailoring hydroxylases HynC and HynE were assessed *in vitro*. To investigate the activity of the additional C domain in module 5, *hynTE* was also cloned and heterologously expressed as described below.

All relevant constructs for this work are summarized in Table 3.3.

Table 3.3: Cloned and heterologously expressed constructs used for *in vitro* investigation of the *hyn* BGC.

Protein	Molecular Weight [kDa]	Tag	Coexpressed With
HynA ₄ AT	67.6	C-terminal 6x His	pG-KJE8
HynA ₄ C _{don} ATC _{acc}	117.7	N-terminal 6x His	
HynA ₅ CAT	122.0	N-terminal 6x His	HynMLP
HynA ₆ AT	70.3	C- & N-terminal 6x His	
HynA ₆ CAT	117.5	N-terminal 6x His	HynMLP
HynB ₇ AT	66.9	C-terminal 6x His	
HynMLP	16.8	-	NRPS modules
HynC	63.2	N-terminal 6x His	NRPS modules
HynE	38.8	N-terminal 6x His	
HynTE	31.7	N-terminal 6x His	

3.1.4.1 Cloning and heterologous expression of NRPS modules

For the generation of expression constructs, sequential cloning was used. The selected DNA segment was amplified via PCR, with specific primers (Table 8.3) from previously extracted gDNA of *Lysobacter sp.* K5869. The resulting fragment was restricted with specific, sticky-ends-forming, endonucleases to integrate it into the multiple cloning site of the plasmid pET28a *via* ligation. *hynMLP* was integrated into the second multiple cloning site of pCDFDuet-1. The ligation product was transformed into the *E. coli* cloning strain α -Select Silver (α SS) for multiplication. Correct cloning and ligation were verified via analytical restriction digest and subsequent Sanger sequencing. For expression of the NRPS genes, the plasmids were transformed into *E. coli* expression strain BAP1 to ensure *in vivo* phosphopantetheinylation of the T domain. For all NRPS constructs, an additional strain, containing pCDFDuet-1::*hynMLP* was created for co-expression with *hynMLP*. Overexpression was started with 0.4 mM IPTG to induce the formation of the respective protein, which was then purified with Ni-NTA based affinity chromatography. The different fractions of the chromatography were analysed *via* SDS-PAGE. In a first attempt, the AT didomains of the modules 4, 6, and 7 were cloned and heterologously expressed. The proteins HynA₆AT and HynB₇AT were obtained in soluble form and in sufficient amounts. HynA₄AT was clearly overexpressed by the *E. coli*, but remained in the insoluble fraction after sonification. To overcome this obstacle, HynA₄AT was co-expressed with the chaperon plasmid pG-KJE8, in addition to HynMLP. This led to successful purification of HynA₄AT.

During this work, several research articles dealing with heterologous expression of NRPS modules were published. Most of these reports suggest larger constructs, because there is growing evidence that the catalytic activity of A domains depends on the interaction with adjacent domains.^[5] For genetic engineering of NRPS, the XU- and XUC-concept was published, where the authors propose a loop in the centre of C domains to be an intersection for recombination events.^[116,117] Additionally, it was reported that constructs with N-terminal tag are superior to constructs with C-terminal tag in regards of yield, solubility and catalytic activity.^[118,119] Regarding these new data, the constructs HynA₄C_{don}ATC_{acc}, HynA₅CAT, and HynA₆CAT were cloned and heterologously expressed with N-terminal hexahistidine tag, each with and without HynMLP. The yields of the overexpressed proteins were sufficient for further *in vitro* testing. Figure 3.3 depicts SDS-PAGEs with elution fraction of all constructs used for final *in vitro* testing. Coexpression with HynMLP did not result in any observable change in solubility or mobility during electrophoresis.

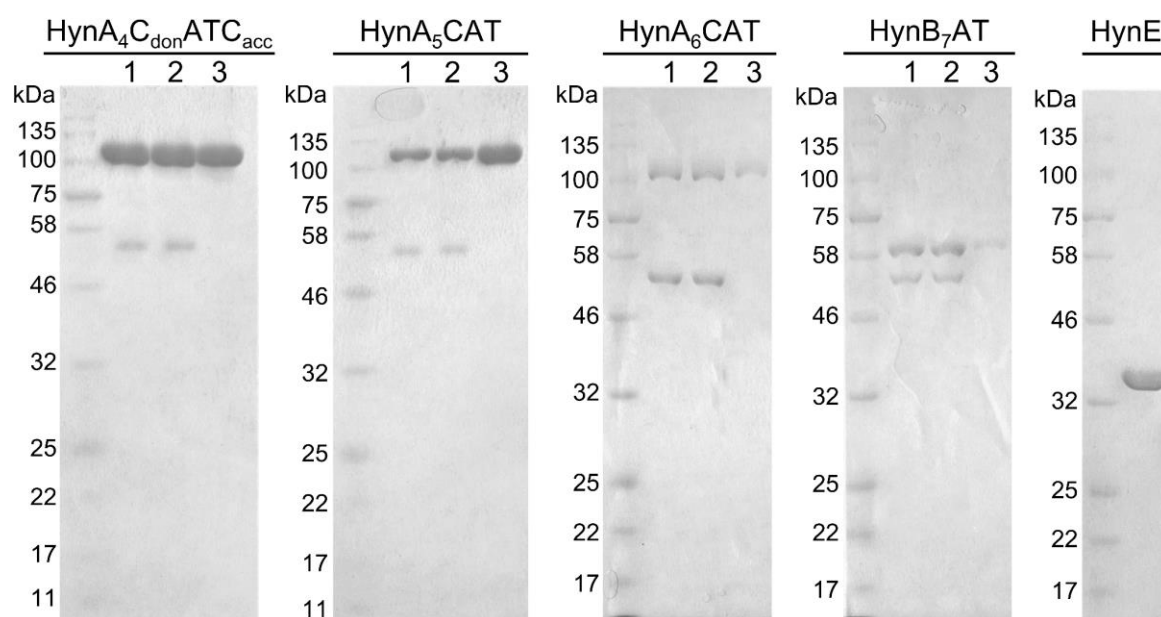


Figure 3.3: SDS-PAGE analyses of heterologously expressed enzymes from the *hyn* BGC. All NRPS constructs are coexpressed with HynC (1, expected size: 63.2 kDa), HynC E376D (2, expected size: 63.7 kDa) and alone (3), as used for the *in vitro* assays. HynA₄C_{acc}ATC_{don} expected size: 117.7 kDa; HynA₅CAT expected size: 122.0 kDa; HynA₆CAT expected size: 117.5 kDa; HynB₇AT expected size: 66.9 kDa; HynE expected size: 38.8 kDa.

3.1.5 *In vitro* activity of A domains

To assess the functionality of all successfully cloned, expressed, and purified NRPS constructs, the catalytic activities of the integrated A domains were evaluated with the $\gamma^{18}\text{O}_4$ -ATP exchange assay.^[120] The principle of this assay is based on the fact that the activation of the amino acid is an equilibrium reaction. As substrates, a defined concentration of a specified amino acid and ATP, isotopically labelled with four heavy oxygen atoms ($^{18}\text{O}_4$), are added. If the A domain is functional and accepts the amino acid as substrate, it forms the activated amino acid-adenylate and $^{18}\text{O}_4$ -pyrophosphate (PPi). In the reverse reaction, the amino acid-adenylate and non-labelled PPi, which is added in excess, is converted into the amino acid and light, non-labelled ATP. To conclude, if the A domain is catalytically active and accepts the added amino acid as substrate, heavy $\gamma^{18}\text{O}_4$ -ATP is converted into light $\gamma^{16}\text{O}_4$ -ATP. The outcome of the assay is quantified as the ratio of the two ATP isotopes ($m/z = 514$ against $m/z = 506$) in MALDI-TOF based mass spectrometry, given in [%] ATP-exchange (see also section 5.8.1).

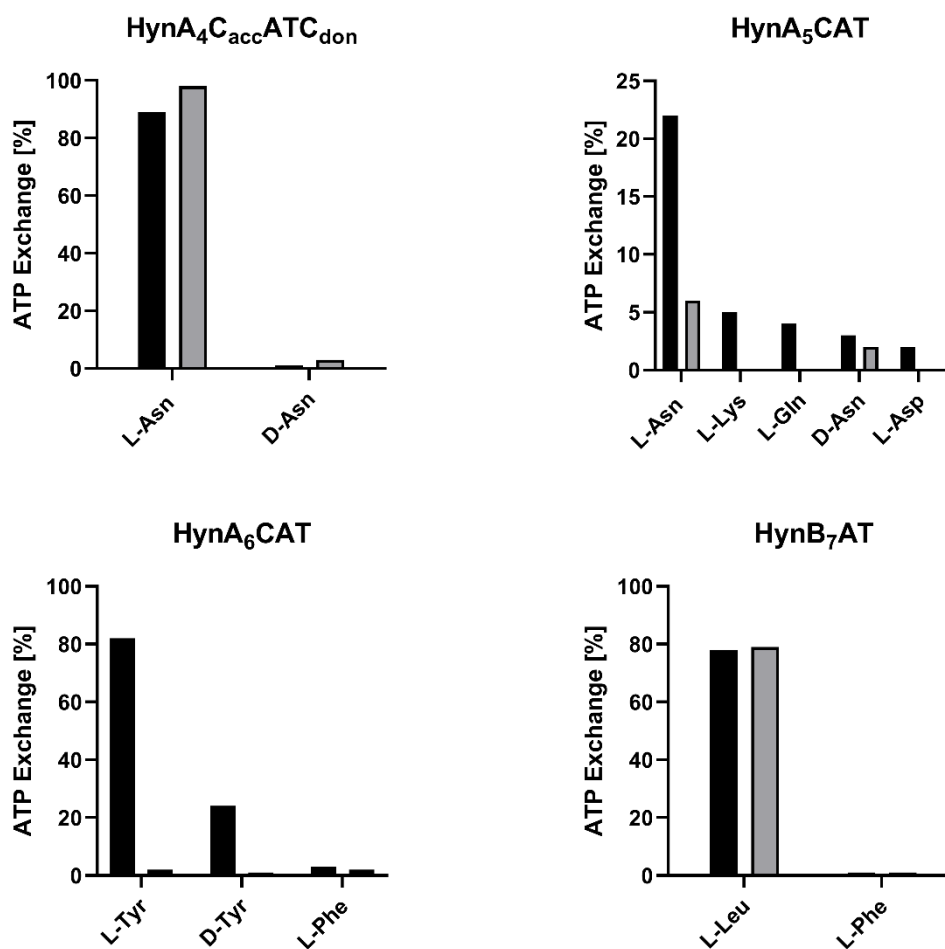


Figure 3.4: Activity of A domains in the *hyn* BGC. The A domains of constructs responsible for the integration of hydroxylated amino acids were tested for their substrate specificity with the $\gamma^{18}\text{O}_4$ -ATP exchange assay. All constructs were assayed with (black bars) and without (grey bars) coexpression of the MbtH-like protein hynMLP.

The smaller constructs, HynA₄AT and HynA₆AT showed no activity in this assay. In a first attempt, the heterologous expression and protein purification protocols were modified to promote the correct folding during translation of the proteins and the stability after purification. Additionally, both constructs were coexpressed with diverse chaperone plasmids (pG-KJE8, pKJE7, pG-Tf2, and pTf16). To eliminate any possible interferences of the A domain with the hexahistidine tag,^[118] the master student Isabelle Stritzinger performed a thrombolytic removal of the tag after purification. However, none of these attempts led to reconstitution of the catalytic activity of these constructs (data not shown), so they were not further investigated.

The larger construct, HynA₄C_{don}ATC_{acc} clearly activated L-Asn with nearly quantitative ATP exchange of 98 %, without coexpression of HynMLP. With HynMLP, the ATP exchange is slightly reduced to 89 %. This might result from a lower protein concentration in the assay, as the presence of HynMLP was not taken into account when calculating the molar absorbance of the proteins at $\lambda=280$ nm, which is the crucial factor for concentration measurements. The necessity of an upstream

interaction partner for some A domains to maintain the adenylating activity has been demonstrated in fungal L-lysine biosynthesis, prototypically represented by NPS3 of *Ceriporiopsis subvermispota*. Here, the monomodular NRPS NPS3 recruits L- α -amino adipic acid, which is subsequently released upon reduction to a semialdehyde. Interestingly, the adenylating activity heavily depends on an adjacent rudimentary C domain that lost its catalytic function, together with major parts of its structure.^[121]

HynA₅CAT only moderately activated L-Asn (22 %), when coexpressed with HynMLP. Without HynMLP, the activity was nearly abolished (6 %), which clearly depicts the dependency of the A domain on the chaperone. Up to a certain extent, A domains may accept amino acids that are structurally related to their actual substrate.^[13] To rule out the possibility that this is the reason for the moderate activity of HynA₅CAT, the assay was also performed with D-Asn, L-Asp, L-Gln, and L-Lys as substrates, but none of these substrates led to higher ATP exchange. These results suggest that, despite its slow activity, L-Asn is the preferred substrate of the A domain in module 5. Low activities of A domains might have diverse origins. In the biosynthesis of hormaomycin, the measurable activity of A domains was strictly dependent on the addition of DTT to the assay mixture. The results were then refined by additional purification steps of the proteins after Ni-NTA based affinity chromatography.^[122] Another reason might be the use of a suboptimal module construct, as A domain activity heavily depends on the interaction with adjacent domains. In teicoplanin biosynthesis, the turnover of an A domain was three times faster, when the naturally adjacent C domain was present in the assay.^[5] The fifth module of the *hyn* BGC harbours two C domains. The first is a common ^LC_L domain, whereas the function of the second C domain is enigmatic (see section 3.1.10). The heterologously expressed construct, HynA₅CAT harbours only the second C domain, so it might be possible that the actual ^LC_L domain is necessary to boost the A domain's activity.

The A domain of HynA₆CAT alone showed no activity at all. However, after coexpression with HynMLP, a high ATP exchange of 82 % with the predicted substrate L-Tyr was observed. Similar to the respective results from HynA₅CAT, this shows the dependency of the heptepin synthetase from HynMLP. The amino acid, D-Tyr was also activated in relatively high amounts (24 %), which should not affect the biosynthesis *in vivo*, as the non-proteinogenic amino acid is not abundant in the cell. This phenomenon has also been observed for FR900359 biosynthesis. Here, the A domain of the first module FrsA does not diminish between the naturally occurring substrate L-Leu and its stereoisomer.^[13] The structurally closely related amino acid L-Phe was not accepted as substrate by HynA₆CAT, reflecting the high substrate specificity towards naturally abundant amino acids.

Interestingly, the small construct HynB₇AT showed exceptional activity with the proposed substrate L-Leu with and without HynMLP (78 % and 79 %), respectively, whereas all other modules needed

to be expressed as larger constructs to reconstitute the A domains activity *in vitro*. The reason for the high activity of HynB₇AT in comparison to HynA₄AT and HynA₆AT remains enigmatic, as the constructs have all been designed in the same manner.

In conclusion, for the modules four, five, six, and seven, which are responsible for the integration of hydroxylated amino acids in the biosynthesis of hypeptin, differently designed constructs were heterologously expressed. For each of the four integrated A domains, the proposed substrate specificity towards the respective proteinogenic amino acid could be confirmed *in vitro*. The constructs of modules five and six need to be coexpressed with HynMLP for successful reconstitution.

3.1.6 Cloning and heterologous expression of *hynC* and *hynE*

The bioinformatic analysis of the *hyn* BGC revealed the presence of two encoded tailoring hydroxylases HynC and HynE, which were hypothesized to be responsible for the formation of hydroxylated amino acids in hypeptin. The occurrence of more than one hydroxylase in a BGC is most often associated with different kind of hydroxylation reactions. A well described example is the biosynthesis of glycopeptides. Here, the structurally diverse crosslinking of the aromatic side chains is mediated by up to four CYP450-hydroxylases.^[123] In the case of teicoplanin, the additional NHDM Tcp25 is responsible for β -hydroxylation of tyrosine, which is crucial for subsequent glycosylation.^[60] Another interesting example is the generation of β -hydroxyglutamic acid in the biosynthesis of kutznerides, where the two α KG hydroxylases KtzN and KtzO both target the same PCP-bound glutamic acid, resulting in two different products with opposite stereoconfiguration at the hydroxylated carbon atom.^[49] On the other hand, if different residues within a peptide are hydroxylated in the same manner, a single hydroxylase is often sufficient, like in the biosynthesis of FR900359,^[93] lysobactin,^[35] and skyllamycin.^[42] In each of these biosyntheses, a single enzyme is responsible for the hydroxylation of three amino acids that are in the latter two cases structurally different (see also chapter 1.3.2).

HynC and *hynE* were cloned into pET28a and heterologously expressed in the same way as described above and in section 5.6. As expression strain, *E. coli* BL21(DE3) was used. HynE could be obtained in high yield and purity after Ni-NTA based affinity chromatography (see Figure 3.3). HynC instead remained in the insoluble fraction after sonification. To overcome this obstacle, HynC was coexpressed with the chaperone plasmid pG-KJE8, which indeed helped to solubilize the protein. Unfortunately, HynC rapidly precipitated in the elution buffer after purification. An immediate change of the buffer system right after elution and constant chilling of the sample to 4 °C did not prevent the precipitation. In a first attempt to keep HynC soluble, all buffers necessary for its purification and subsequent handling were supplemented with additives that are reported to

support protein stability (10 % glycerol, 250 mM glycine, 0.01 % TritonX™, 0.05 % Sarcosyl NL, 0.5 M Urea, 3 mM TCEP),^[124] but the protein still precipitated after elution in each of these conditions. The reason for the difficult purification and subsequent precipitation of HynC might be that the interface of the hydroxylase is exposed to water without the interaction partner. The interaction of tailoring enzymes with their substrate T domain is often driven by hydrophobic interaction.^[22] Without the substrate protein, the interface residues might thus induce major rearrangements to evade the hydrophilic environment. To test this hypothesis, *hynC* was cloned and integrated into the first multiple cloning site of pCDFDuet-1::*hynMLP*, resulting in the plasmid pCDFDuet-1::*hynC_hynMLP*, which was used to coexpress *hynC* with each of the catalytically active NRPS constructs of HynA₄C_{don}ATC_{acc}, HynA₅CAT, HynA₆CAT, and HynB₇AT in *E. coli* BAP1. Strikingly, no precipitation of proteins was visible after elution in any of the samples. After 30 min at room temperature, a white precipitate was visible in the sample with HynA₆CAT, which could be circumvented by chilling on ice. SDS-PAGEs of the samples proofed the successful coexpression of the respective NRPS construct together with HynC (Figure 3.3).

With all necessary proteins isolated, the next aim was to investigate the activities of the hydroxylases HynC and HynE *in vitro*. First experiments performed by Master's student, Isabelle Stritzinger showed that the coexpressed enzymes HynB₇AT with HynC also yielded the product 3-hydroxyleucine (Hle) in the negative control, where the proteins were heat inactivated before addition of the substrates. We postulated that Hle was already produced during expression of the proteins *in vivo*, as all necessary enzymes and substrates for the reaction (coenzyme A, L-leucin, ATP, electron donor system) are most likely present in *E. coli*.

To disable the *in vivo* activity of HynC as a negative control, the enzyme was rendered inactive. In the case of CmlA, the prototype NHDM from chloramphenicol synthase, Jasniewski *et al.* postulated the glutamate residue 377 in the active site to be responsible for oxygen regulation during the catalytic cycle.^[58] In the reduced state, the carboxylic side chain of E377 is bidentately bound to the iron atom Fe1. Once the cognate T domain is attached and the substrate L-PAPA is coordinated within its binding pocket, it induces a shift of the E377 side chain, which results in a monodentate binding to Fe1. The now open binding site on Fe1 is then occupied by the substrate O₂, which is crucial for subsequent oxidation steps. To verify the hypothesis, Jasniewski *et al.* generated the mutant CmlA E377D. The shorter and less flexible side chain of the aspartate was only able to bind monodentately, leading to unregulated binding of oxygen and thus complete loss of function, whilst the structure of the enzyme remained identical to the wild type.^[58] A BLOSUM62-based protein alignment revealed the critical glutamate to be highly conserved in all NRPS-related NHDM (see Appendix Figure 8.2). These results were utilized for this thesis to generate an inactive, but presumably structurally identical mutant of HynC as negative control in

the coexpression experiments. The plasmid pCDFDuet-1::*hynC_hynMLP* was mutated *via* Gibson assembly to generate pCDFDuet-1::*hynC E376D_hynMLP*.

HynC E376D was coexpressed with the NRPS modules HynA₄C_{don}ATC_{acc}, HynA₅CAT, HynA₆CAT and HynB₇AT as described above for HynC. The expression yielded similar amounts of soluble protein and no difference between HynC and HynC E376D were distinguishable in SDS-PAGE (see Figure 3.3).

3.1.7 *In vitro* activity of HynE

To assess the function of HynE, we adapted the assay procedure from Singh *et al.*, where the origin of a 3-hydroxyaspartate residue in the siderophore Syringomycin E was investigated.^[52] Within the respective BGC, they found the gene *syrP*, which is homologous to the iron- and α -ketoglutarate-dependent dioxygenase TauD from *E. coli*.^[125] Singh *et al.* discovered that SyrP β -hydroxylates T domain bound aspartate and is unable to use the free amino acid as substrate. Free Fe and α KG are essential for the reaction. Fe is coordinated to the catalytic centre of the enzyme and oxidative decarboxylation of α KG provides succinate, CO₂ and a high-valent Fe(IV)-oxo intermediate, which is able to abstract a hydrogen atom from the primary substrate. The role of ascorbate in the catalytic cycle of Fe/ α KG dependent hydroxylases is not clear yet. In some cases, it is necessary to regenerate Fe after oxidation, but most often it just enhances the product yield without being obligatory,^[45,126] which is also the case for SyrP.^[52]

HynA₄C_{don}ATC_{acc}, HynA₅CAT, HynA₆CAT, HynB₇AT, and the hydroxylase HynE were overexpressed in *E. coli* BAP1 or *E. coli* BL21(DE3) with a hexahistidine tag and purified using affinity chromatography. Subsequently, each of the purified NRPS modules was mixed with HynE. As substrates, Fe(III), α KG, ascorbate, ATP, and the respective amino acid were added as described in section 5.8.3. If hydroxylation takes place, the *m/z* of the hydroxylated amino acid is detected in subsequent LC-MS analysis. For the negative control, HynE was heat inactivated, and should not yield any hydroxylated product. The expected *m/z* for each NRPS module is given in Table 3.4. Unfortunately, the authentic standards for the hydroxylated amino acids 3-hydroxyisoleucine (Hle), 3-hydroxyasparagine (Has), and 3-hydroxytyrosine (Hty) were not available commercially.

Table 3.4: Amino acids activated by the investigated NRPS modules, the hydroxylated product, and their *m/z* in negative mode.

NRPS module	Amino acid	<i>m/z</i>	<i>m/z</i> hydroxylated
HynA ₄ C _{don} ATC _{acc}	Asn	131.04	147.04
HynA ₅ CAT	Asn	131.04	147.04
HynA ₆ CAT	Tyr	180.06	196.06
HynB ₇ AT	Leu	130.09	146.08

Figure 3.5 A shows the single ion record (SIR) traces for the expected m/z of Has of the hydroxylation assay with HynE and HynA₄C_{don}ATC_{acc}. Strikingly, the mass of 3-hydroxyasparagine was detected in the assay, but not in the negative control. The peak in the chromatogram displayed an asymmetric shape with a “shoulder”. This might originate from the buffer system present in the mobile phase. The used pH of 7.4 is probably too close to the isoelectric point of the amino acid, so it is not protonated quantitatively. The different loading states of the molecules have a high impact on the chromatographic resolution, as the interaction with the stationary phase in a HILIC system is highly sensitive to ionic interaction.^[127] Trails to resolve the peak shape by optimizing the mobile phase conditions were unsuccessful because at lower pH, the amino acid was hardly detectable in positive nor in negative mode. A high pH would be optimal for the detection in negative mode, but most HPLC columns do not tolerate a pH over 8.5. Nevertheless, this result demonstrates the formation of 3-hydroxyasparagine by HynA₄C_{don}ATC_{acc} and HynE.

The LC-MS traces of the assay with HynA₅CAT and HynE are depicted in Figure 3.5 B. They are similar to those of HynA₄C_{don}ATC_{acc} with HynE, as the same amino acid gets hydroxylated. These results demonstrate the recruitment of asparagine by HynA₅CAT and its subsequent hydroxylation by HynE. Figure 3.5 C and D show the results of the hydroxylation assays of HynE with HynA₆CAT and HynB₇AT, respectively. No modified product amino acids were detectable, which demonstrates that HynE is only responsible for the hydroxylation of L-Asn in modules 4 and 5.

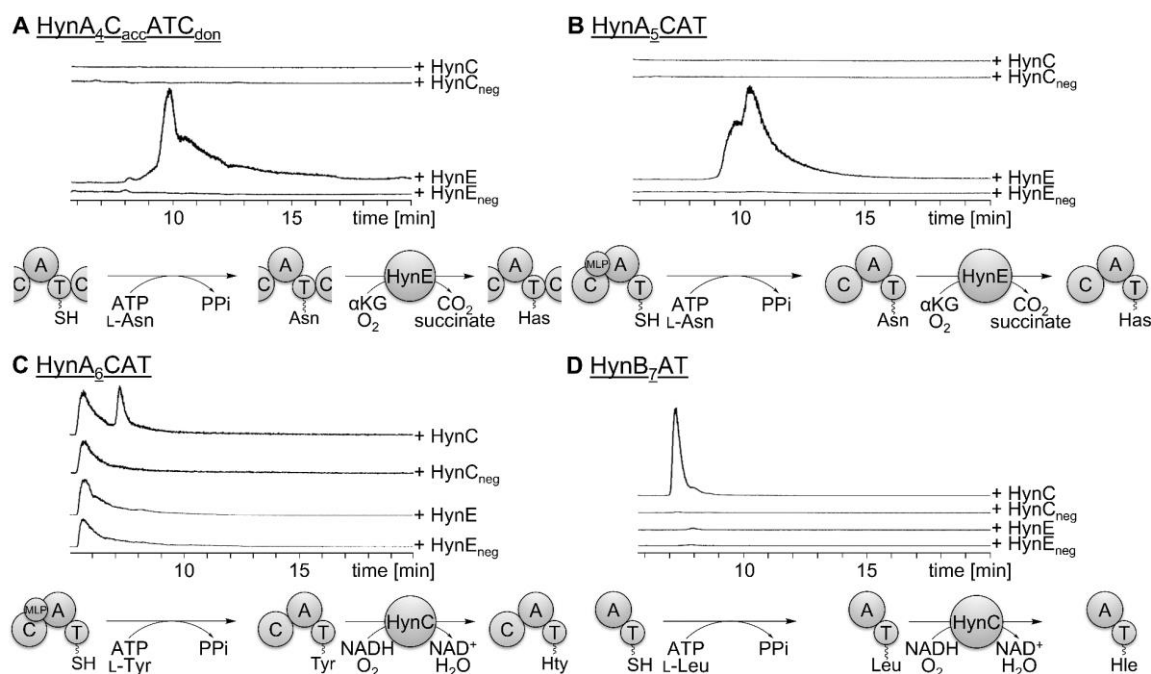


Figure 3.5: Activity of the hydroxylases HynC and HynE. The activity was tested *in vitro* with each of the amino acid-loaded NRPS modules as substrates and the hydroxylated product detected *via* LC-MS at the respective m/z . **A** Formation of 3-hydroxyasparagine (Has) by HynA₄C_{acc}ATC_{don} and HynE, $m/z = 147.0$. **B** Formation of Has by HynA₅CAT and HynE, $m/z = 147.0$. **C** Formation of 3-hydroxytyrosine (Hty) by HynA₆CAT and HynC, $m/z = 196.1$. **D** Formation of 3-hydroxyisoleucine (Hle) by HynB₇AT and HynC, $m/z = 146.1$.

3.1.8 *In vitro* activity of HynC

The assay to test the activity of the tailoring NHDM HynC was adapted from publications about the prototype NHDM CmlA. To hydroxylate the substrate amino acid that is covalently bound to the T domain of the cognate NRPS module, the diiron cluster of the hydroxylase needs to be transferred to the reduced state. For most NHDMs, this is achieved by addition of the cognate electron donor system,^[128] but for CmlA, no such system could be found within the genome. Instead, activation of the enzyme was achieved by chemical reduction, using sodium dithionite together with methyl viologen as electron transmitter. Interestingly, this activation was not reproducible in later studies, which led the authors to switch to a NADH/methyl viologen system.^[5,60]

HynC was heterologously coexpressed with each of the NRPS modules HynA₄C_{don}ATC_{acc}, HynA₅CAT, HynA₆CAT, or HynB₇AT in *E. coli* with a hexahistidine tag and subsequently purified using affinity chromatography. The purified proteins were then incubated with the respective amino acid, ATP, NADH, and methyl viologen (see section 5.8.2). As a negative control, HynC E376D was coexpressed with the NRPS modules in parallel and handled identically. The formation of the product amino acid was detected with LC-MS. Table 3.4 gives the expected masses of the hydroxylated and non-hydroxylated amino acids for each NRPS module.

Figure 3.5 **A** and **B** show the SIR traces of the hydroxylation assay of HynC with the NRPS modules HynA₄C_{don}ATC_{acc} and HynA₅CAT. Incubation of HynC with the NRPS modules and all necessary substrates does not lead to the formation of Has.

The LC-MS traces of the hydroxylation assay with HynC and HynA₆CAT are depicted in Figure 3.5 **C**. Here, a new peak with the expected m/z of Hty is visible. The detected amount of product was rather low and could not be enhanced in subsequent trials. This is probably due to the low solubility of tyrosine in water at the used pH,^[129] which strictly limits the amount of substrate for the A domain. Additionally, the proteins in this assay exhibit a low stability and tend to precipitate at room temperature. Lastly, the product Hty is sensitive to decomposition in acidic or alkaline solution.^[130] The conditions used to process the assay prior to LC-MS measurement are rather mild but might still contribute to the low product yield.

Figure 3.5 **D** shows the SIR of the hydroxylation assay with HynC and HynB₇AT. A clear peak with m/z of 146.1 indicates the formation of Hle in the assay. In the negative control, only trace amount of the product amino acid could be detected, which is comparable with the minimal basal activity reported for CmlA E377D.^[58]

3.1.9 Stereochemistry of hydroxylated amino acids

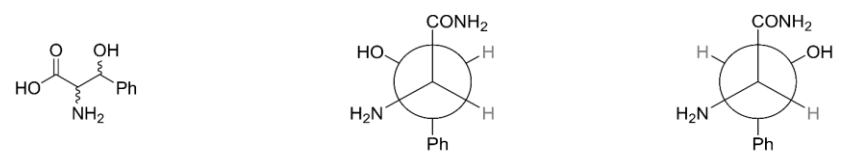
In the following section, the stereochemistry of the hydroxylation reactions in hypeptin biosynthesis will be discussed. The description and comparison of the β -hydroxylated amino acids is complicated by two facts: The absolute stereoconfiguration of C_3 , expressed in S,R -descriptors, depends on the priority of the neighbouring atoms. As the amino acids differ in their side chains, for Hle and Hty the priority is opposed to that of Has. Consequently, (2*S*,3*S*)-3-hydroxytyrosine and (2*S*,3*S*)-3-hydroxyasparagine have the same descriptors for the absolute configuration, but are sterically different at C_3 . This problem is solved with relative descriptors (*syn,anti*) by referring the configuration of C_3 to the configuration of C_2 , which always has the same neighbouring atoms in proteinogenic amino acids. Unfortunately, the C_2 -configuration of Has₅ is inverted by a dual epimerisation condensation domain during biosynthesis. Nevertheless, for the sake of clarity, the relative configuration, which is used to describe the stereospecificity of the hydroxylated C_3 , always refers to the 2*S*-configuration of the amino acids. As the C_2 -epimerisation takes place after hydroxylation and right before the complete assembly of the amino acid,^[5,42] it does not need to be taken into account when discussing the stereochemistry of the hydroxylation reaction.

At a first glance, the results of the *in vitro* hydroxylation assays with the two hydroxylases HynC and HynE complement each other and enable to coherently describe the biosynthesis of the hydroxylated amino acid residues in hypeptin. According to the obtained results, the NHDM HynC is responsible for the formation of Hty₆ and Hle₇, while HynE hydroxylates Asn₄ and Asn₅, resulting in two Has residues.

In the structure of hypeptin, published by Shoji *et al.*, the amino acids Has₄, Has₅, and Hty₆ are *anti* configured, whereas Hle₇ is *syn* configured.^[85] These configurations are not in line with the results of the *in vitro* hydroxylation assays described above, because according to this structure, HynC would hydroxylate Tyr₆ and Leu₇ in sterically opposed directions. Enzymes that do not perform chiral discrimination of their prochiral substrate are rare and involve a nonenzymatic reaction step.^[131] This is improbable in the case of HynC, as the catalytic cycle of NHDM is described in detail.^[57,132] Shoji *et al.* determined the stereochemistry of the hydroxylated amino acids by hydrolysis of hypeptin and subsequent chiral column HPLC-MS, compared to authentic standards. As Hty degrades during hydrolysis, its presence and structure were solely proven by NMR analysis, but no conferring data were published. To verify the reported structure of hypeptin, an in-depth NMR analysis of hypeptin was recorded by Dr. Stefan Kehraus (Appendix Table 8.2). The resulting data, combined with the results of the biochemical characterization of the enzymes, were then used to clarify the absolute configuration of Hty.

Vicinal *H,H*-coupling constants $^3J_{HH}$ depend on the relative configuration of the coupling protons, as they are directly influenced by the interplanar angle of the involved C-H bonds. If the carbon

atoms are differently substituted, they preferentially adapt a conformer with minimized steric interaction between the substituents. The resulting limited rotation leads to a greater or smaller $^3J_{HH}$.^[133] To determine the relative configuration of the stereocentres in Hty, $^3J_{HH}$ of H2_{Hty} and H3_{Hty} were compared with values from analogous compounds with already known configuration. Lin *et al.* used this method to determine the relative configuration in burkholderines. They observed a lower coupling constant in *anti* configuration, while in *syn* configuration it was higher.^[130] In hypeptin, the measured coupling constant was 7.2 Hz, which strongly suggests Hty to be *syn* configured (see Figure 3.6). According to the results of the $\gamma^{18}\text{O}_4$ -ATP exchange assay, the preferred substrate of A₆ is (2*S*)-Tyr and the module does not contain any epimerase domain, which allows to reassign the absolute configuration to (2*S*,3*R*)-Hty. Analysis of the ROESY NMR data of hypeptin by Dr. Stefan Kehraus support the configuration shown in Figure 3.2 (see Appendix Figure 8.20).^[134]



The figure shows three chemical structures. On the left is the 2D structure of 3-hydroxytyrosine with a phenyl group (Ph) at the 3-position. In the middle is a Newman projection for the *syn* configuration, where the CONH₂ group is at the top, the OH group is on the left, and the H₂N group is at the bottom-left. The Ph group is at the bottom, and H atoms are at the top-right and bottom-right. On the right is a Newman projection for the *anti* configuration, where the CONH₂ group is at the top, the OH group is on the right, and the H₂N group is at the bottom-left. The Ph group is at the bottom, and H atoms are at the top-left and bottom-right.

Configuration	<i>syn</i>	<i>anti</i>
$^3J_{H,H}$ literature values	4.6- 6.6 Hz	3.5- 4.8 Hz
Measured for hypeptin	7.2 Hz	

Figure 3.6: NMR analyses to characterize the relative configuration of 3-hydroxytyrosine (Hty) in hypeptin. Comparison of vicinal coupling constants with described compounds determine Hty to be *syn* configured

3.1.10 Bioinformatic analysis of HynA₅C₂

Bioinformatic analysis of the *hyn* BGC revealed two consecutive C domains in the fifth module of HynA. Both are annotated as ^LC_L Domains by antiSMASH,^[112] which is however doubtful. To get a deeper insight into the function of the domain, a phylogenetic tree was constructed with the trimmed dataset from Rausch *et al.*,^[135] and expanded by sequences of the more recently described I domains,^[53] X domains,^[136] and all C domains from the *hyn* BGC. I domains, as well as X domains were shown to be responsible for the recruitment of *trans* acting oxygenases in NRPS. Interestingly, HynA₅C₂ fits into the ^LC_L clade only with high distance. Hence, this domain probably belongs to a distantly related domain family. A brief BLAST search revealed few homologues of HynA₅C₂ in NRPS from various strains and also from well-known BGCs, like the synthetases of lysobactin,^[35] syringomycin,^[52] and teixobactin.^[82] When including 14 of the homologues into the phylogenetic analysis, they grouped together and formed a deeply rooted subtree within the ^LC_L clade (Figure

3.7). Apart from HynA₅C₂, all of them belong to modules incorporating threonine. For five of the corresponding BGCs, the absolute stereoconfiguration of the final peptide has been determined.^[35,52,82,137,138] In all these peptides, the configuration at C₃ of the respective Thr is inverted in comparison to the natural L-Thr. Hence, a C₃ epimerase function for this domain family is postulated.

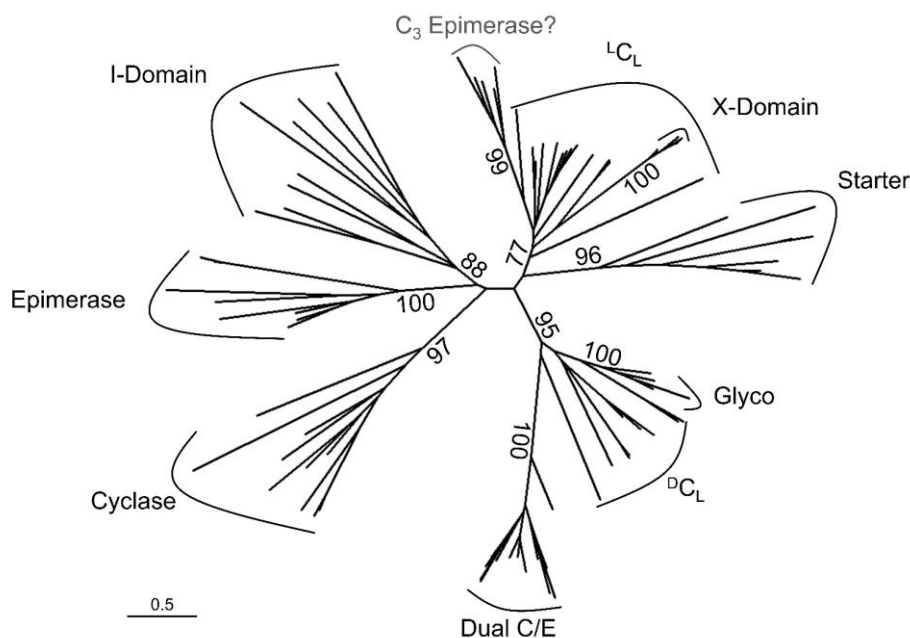


Figure 3.7: Phylogenetic tree of C domains from different datasets. The domains cluster depending on the specified function. HynA₅C₂ and homologs thereof form a new, distinct clade within the ^LC_L domains and probably catalyse C₃ epimerization (grey). Bootstrap values at key branches indicate the distinctness of clades. The scale bar represents 50 substitutions per 100 amino acids. Accession numbers of the sequence data used for this analysis are summarized in **Table 8.1**.

A detailed investigation of the primary structures of the putative C₃ epimerases revealed that here, the canonical HHxxxDG-motive is rendered to HHxxxDR. Similar to the X domains, which are responsible for the recruitment of tailoring CYP-oxygenases in glycopeptide biosynthesis, the bulky side chain of the mutated amino acid probably blocks the substrate channel to the acceptor side of the domain,^[136] leaving the substrate channel of the donor side as the only entrance to the active site. Another conserved residue within the family of C₃ epimerases is His6, which replaces an aliphatic amino acid in most C domains. In computed crystal structures,^[139] His6 is located in Helix α -1 with its side chain projecting towards active side (Figure 3.8). Analogously to the proposed mechanism of common E domains,^[140] the imidazole side chain of His6 might initialize the epimerisation by abstracting a proton from C₃. The active site of HynA₅C₂ and its closest homologue from *L. psychrotolerans* ZS60 contain a HRxxxDR-sequence, which would render the domains inactive. The identical C₃-configuration of Has₄ and Has₅ in hypeptin strongly support this theory.

Domain Family	2 nd res. in α -1	Consensus
C ₃ epimerase	His	H-H-x-x-x-D-R
^L C _L	aliphatic	H-H-x-x-x-D-G
^D C _L	aliphatic	H-H-x-x-x-D-G
X domain	Arg	H-R-x-x-x-D-D
I domain	variable	X-X-x-x-x-D-X
Starter	aliphatic	H-H-x-x-x-D-G
Glyco ^D C _L	aliphatic	H-H-x-x-x-D-G
Epimerase	aliphatic	H-H-x-x-x-D-G
Dual	aliphatic	H-H-x-x-x-D-H
Cyclase	aliphatic	D-X-x-x-x-D-A

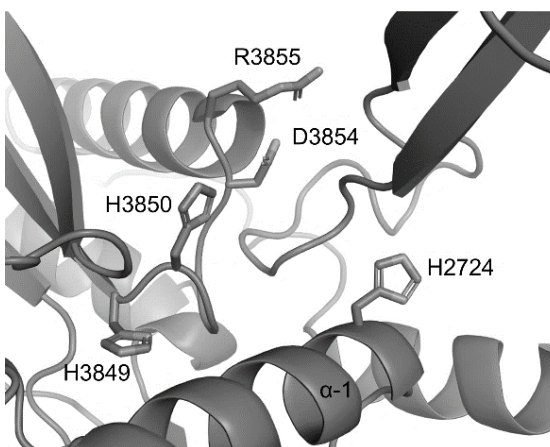


Figure 3.8: **Left:** Comparison of active site residues in different C domain families. All catalytically active domains contain an aliphatic amino acid residue at position 2 of helix α -1. The putative family of C₃ epimerases harbours a histidine at this position, which might initialize the epimerization reaction. **Right:** Homology-generated model of the putative C₃ epimerase from the BGC of lysobactin. The protein backbone is represented in cartoon style. View from the donor side into the active site. Amino acids of the consensus sequence and the second residue of α -1 (H2724) are visualized with side chains, showing the proximity of the unusual histidine to the active site.

3.1.11 *In vitro* activity of HynA₅C₂

To investigate the proposed C₃ epimerization activity of HynA₅C₂, the product of the *in vitro* hydroxylation reaction with HynE was derivatized with 9-fluorenylmethoxycarbonylchloride (Fmoc-Cl) to enable UV/Vis detection at $\lambda=263$ nm and examined with chiral column HPLC. As a negative control, the mutant HynA₅C₂ R201A with a HAx₃DR motive as consensus sequence was created. Heterologous expression of this protein in *E. coli* with subsequent NiNTA-based purification and SDS-PAGE analysis showed no difference in the expression pattern (data not shown). The aliphatic side chain of alanine at a crucial position of the active site reliably renders each kind of C domain inactive.^[24] Briefly, after *in vitro* hydroxylation by HynE, T domain bound amino acids were released by alkaline thioester cleavage or by the type II thioesterase HynTE, and subsequently derivatised with Fmoc-Cl at mild alkaline conditions. Afterwards, excessive Fmoc-Cl was precipitated by addition of 1-adamantanamine (ADAM) and the whole mixture lyophilized. Derivatized amino acids were solubilized in MeCN and then subjected to chiral column HPLC connected to a DAD, as described in section 5.8.5.

The analysis of the chromatograms was complicated by the high abundance of Fmoc-OH and Fmoc-Tris in the chromatograms. Switching the buffer system to HEPES and adjustment of the mobile phase helped to discriminate between smaller peaks. Nevertheless, the procedure still needs to be improved. Apart from Fmoc-OH and Fmoc-HEPES, the abundance of Fmoc-L-Asn from the substrate is verified with authentic standards. One additional, very small peak with a shorter retention time compared to Fmoc-L-Asn was detected in the assay with HynA₅C₂, and also with

HynA₅C₂ R201A. This might be Fmoc-Has, but its retention time did not vary between the assay and the negative control, indicating an identical stereochemistry of the product (Figure 3.9). Unfortunately, the signal was weak and the retention times were hardly reproducible due to fluctuation of the ambient temperature, so no further experiments were conducted. To verify the formation of Has, a HPLC device with column oven, coupled to an ESI-MS-based detection device should be applied. Once the formation and detection of Fmoc-Has is established, more mutations of HynA₅C₂ should be examined. As the bioinformatic investigations suggest the domain to be inactive due to R201, the mutant R201H might re-establish epimerisation activity. The mutation H6A might give hints regarding the role of H6 in the catalysis. Finally, the whole protocol would be simplified by switching to a NRPS module that is independent of a tailoring hydroxylase, for example module 8 of the lysobactin synthase.^[35]

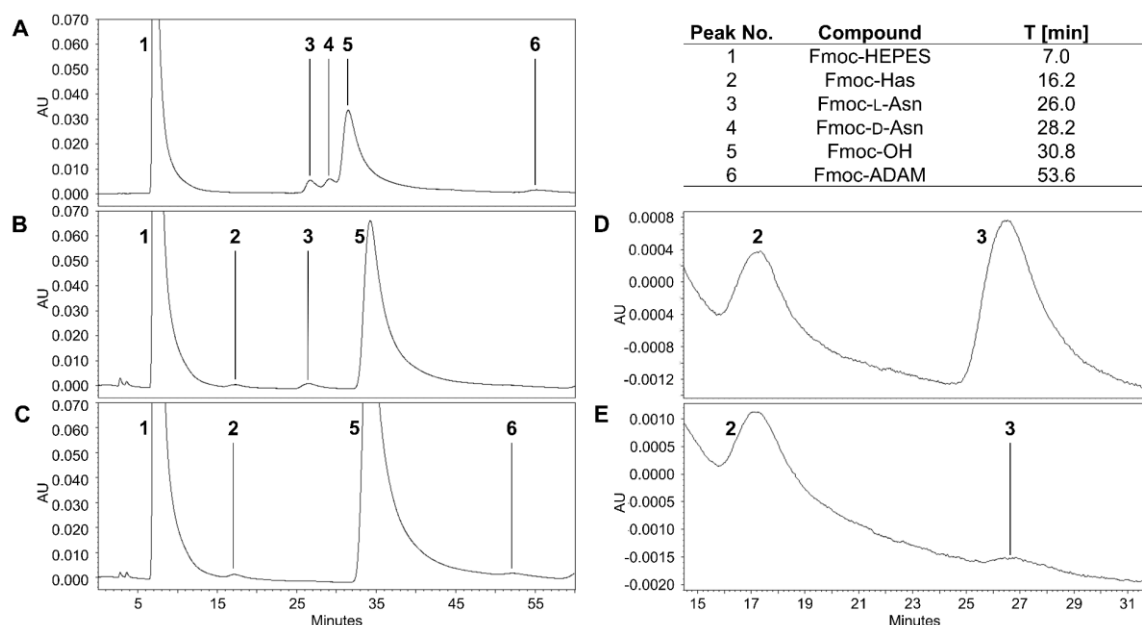


Figure 3.9: Characterization of HynA₅C₂ by chiral column HPLC analysis of hydroxylated amino acid products. After incubation of HynA₅CAT and HynE with all necessary substrates, the products were derivatised with Fmoc-Cl to facilitate UV/Vis-mediated detection at 263 nm. Shifts in the retention time of corresponding peaks were due to changes of the ambient temperature. **A** Blank measurement without proteinogenic assay components, supplemented with D-Asn show successful separation of D- and L-Asn. Peaks of Fmoc-HEPES and non-derivatized Fmoc-OH might interfere with product-related peaks. **B** Assay with HynA₅CAT and HynE leads to the formation of a new, more hydrophilic compound (peak 2), probably Fmoc-hydroxyasparagine (Fmoc-Has). **C** Assay with HynA₅CAT R201A and HynE also lead to the formation of Fmoc-Has (peak 2). Its retention time is not shifted, compared to the assay with non-mutated HynA₅CAT, so no epimerization took place. **D** Assay with HynA₅CAT and HynE, magnified to peaks 2 and 3. **E** Assay with HynA₅CAT R201A and HynE, magnified to peaks 2 and 3.

3.2 Lysobactin

The nonribosomally synthesised peptide lysobactin (also known as katanosin B) from *Lysobacter* sp. ATCC53042 is a potent lipid II-binding antibiotic. The biosynthesis of its core scaffold, consisting of eleven amino acids that form a nine membered macrocycle, has been described by Hou *et al.*^[35] Lysobactin comprises in total three β -hydroxylated amino acids, (2*S*,3*R*)-3-hydroxyphenylalanine₃ (=Phenylserine, Hph₃), (2*S*,3*R*)-3-hydroxyleucine₄ (Hle₄), and (2*S*,3*S*)-3-hydroxyasparagine₁₀ (Has₁₀). These features are very interesting from a biosynthetic point of view, as the amino acids are composed of structurally diverse side chains, but the BGC putatively encodes only for a single hydroxylase (*lybC*), which therefore displays a wide substrate specificity towards the T domains bound amino acids. Lysobactin also comprises two non-hydroxylated Leu residues (Leu₁ and Leu₅), so LybC also needs to be highly specific towards the cognate NRPS modules. Knowledge about LybC would make significant contribution to the general understanding of *trans* acting hydroxylases in NRPS, but despite the unrestricted availability of the producing strain, it was not characterized so far. The aim of this project was to verify the hydroxylating activity of the putative NHD LybC *in vitro* and to further investigate its substrate specificity. Most of the cloning, heterologous expression and *in vitro* testing was conducted by Isabelle Stritzinger during an internship and her master's thesis.

3.2.1 Cloning and heterologous expression of *lyb* genes

To investigate the hydroxylating activity of LybC, the hydroxylase and several constructs of the NRPS modules 1, 3, 4, and 10 were cloned and heterologously expressed in *E. coli*. Sequential cloning was applied to insert the specific gene into the plasmid pET28a, which was then transferred into the expression strain *E. coli* BL21 (*lybC*) or *E. coli* BAP1 (NRPS modules). From the NRPS modules constructs encoding either single A domains to test the adenylation activity, or AT didomains for subsequent investigation of the hydroxylation activity were cloned. The AT didomain constructs of module 3 and 4 did not show any adenylation activity, so larger constructs of these modules were also cloned and heterologously expressed.

Table 3.5 summarizes all relevant constructs of this project. All constructs of the *lyb* BGC were successfully cloned and heterologously expressed in *E. coli*. Most proteins could be obtained in decent amounts and purity after NiNTA-based affinity chromatography, except for LybA₃ATC, which was only detectable in minor amounts after elution.

Table 3.5: Cloned and heterologously expressed constructs used for *in vitro* investigation of the *lyb* BGC.

Protein	Molecular Weight [kDa]	Tag	Constructed by	Tested by
LybA ₁ AT	68.6	N-terminal 6x His	DW	DW
LybA ₃ A	58.4	N-terminal 6x His	IS	IS
LybA ₃ AT	67.1	N-terminal 6x His	IS	IS, DW
LybA ₃ ATC	107.2	N-terminal 6x His	IS	IS
LybA ₃ C _{acc} ATC _{don}	166.6	N-terminal 6x His	IS	IS
LybA ₄ A	58.3	N-terminal 6x His	IS	IS
LybA ₄ AT	67.1	N-terminal 6x His	IS	IS, DW
LybA ₄ CAT	116.4	N-terminal 6x His	IS	DW
LybB ₁₀ A	58.9	N-terminal 6x His	IS	DW
LybB ₁₀ AT	67.4	N-terminal 6x His	IS	DW
LybC	63.0	N-terminal 6x His	DW	DW

3.2.2 *In vitro* activity of A domains

The activity of the A domains within the NRPS modules was tested with the $\gamma^{18}\text{O}_4$ -ATP-exchange assay to verify the recruitment of non-hydroxylated amino acids, which implies that the hydroxylation takes place after thiolation of the proteinogenic amino acid onto the T domain. Additionally, catalytically active A domains are crucial for the investigation of the hydroxylase LybC. This assay measures the adenylating activity of A domains as ATP exchange in percent. For more detailed information about the assay, please refer to the section 5.8.1. The results of the assays are shown in Figure 3.10.

The single domain constructs, LybA₃A and LybA₄A both show a high adenylating activity towards their predicted substrates L-Phe and L-Leu, respectively. Interestingly, for the didomain constructs LybA₃AT and LybA₄AT, this activity depleted almost completely. These results contradict the common assumption that larger constructs of NRPS modules are beneficial for *in vitro* reconstitution, but some publications report lower adenylating activity of AT-didomains, compared to the single A domains.^[141] These findings are supported by structural data: crystal structures of multidomain constructs in various states of the catalytic cycle have shown that, after thiolation, the small subunit of the A domain (A_{sub}) swings open to pull the T domain out of the A domain's binding pocket. In this process, the T domain is pushed towards the upstream C domain, where it adopts its position in the acceptor binding site. Simultaneously to the subsequent condensation reaction, the A domain initiates the next catalytic round by adenylating another amino acid.^[16,29] In case of the AT didomain, the T domain cannot interact with any other binding partner than the A domain, making it unlikely for the latter to adopt the adenylating state. The high activity of LybA₃C_{acc}ATC_{don} and LybA₄CAT in the assay support this theory. Here, the T domain binds to the acceptor side after thiolation, which “freezes” the A domain in adenylation state.^[142] The results of LybA₃ATC, however show that the downstream C domain is not sufficient to restore the activity.

Here, the T domain might bind to the donor side of the C domain, but in contrast to hypothesis generated from crystal structures, this construct displays impeded adenylation.^[18]

LybB₁₀A and LybB₁₀AT both moderately activated the predicted substrate L-Asn. These results are comparable with those for HynA₅CAT (see section 3.1.5). Analysis of the construct LybB₁₀CAT would have given more insight into the dependency of the A domain's activity on adjacent domains, but the construct was not finished before abortion of the project.

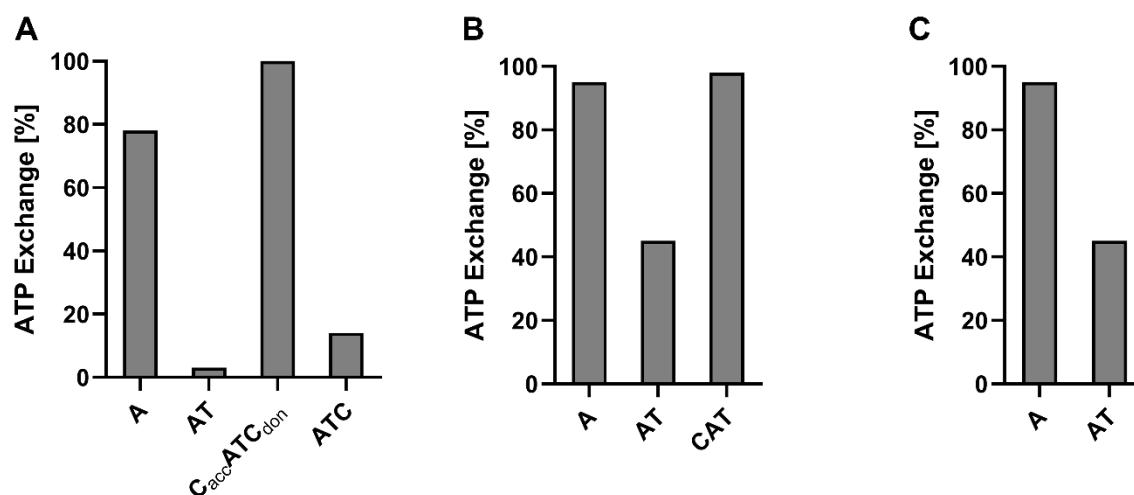


Figure 3.10: Activity of A domains in NRPS modules of the *lyb* BGC, which are responsible for the integration of hydroxylated amino acids. The NRPS modules were constructed in different domain architectures and the adenylation activities towards the predicted substrate amino acids were assayed with the $\gamma^{18}\text{O}_4$ -ATP exchange assay.^[120] **A** Constructs of LybA₃ assayed for adenylation activity with L-Phe as substrate. **B** Constructs of LybA₄ assayed for adenylation activity with L-Leu as substrate. **C** Constructs of LybB₁₀ assayed for adenylation activity with L-Asn as substrate.

3.2.3 UV/Vis spectrum of LybC

The putative NHDM LybC was heterologously expressed in *E. coli* BL21(DE3) and successfully purified, enabling further investigations *in vitro*. The diiron cluster of NHDMs absorbs light at a wavelength of roughly $\lambda=340$ nm to a certain extent, making it a characteristic feature of this enzyme family in UV/Vis-spectroscopy. This feature is visible in the ferrous state and diminishes upon reduction to the ferric state.^[54,143] Figure 3.11 shows the absorption spectra of LybC as obtained after NiNTA-based affinity purification (solid) and upon treatment with the chemical reducing agent sodium dithionite (dashed). The spectra are comparable with the ones of CmlA by Makris *et al.*,^[54] supporting the bioinformatically generated theory that LybC is a NHDM.

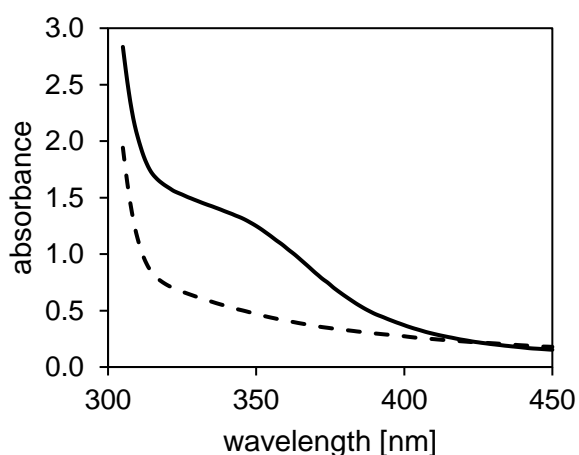


Figure 3.11: UV/Vis- absorption spectra of LybC. The enzyme is isolated from *E. coli* BL21(DE3) in ferrous state (solid), showing a typical feature at $\lambda=340$ nm. Transfer of LybC to the ferric state with sodium dithionite diminished the feature (dashed).

3.2.4 In vitro activity of LybC

The positive results of the $\gamma^{18}\text{O}_4$ -ATP-exchange assay with the NRPS module constructs LybA₃C_{acc}ATC_{don}, LybA₄CAT, and LybB₁₀AT, enabled testing of the activity of LybC *in vitro*. LybC and the NRPS modules were separately heterologously expressed in *E. coli*. After purification and buffer exchange, the enzymes were mixed and supplemented with the substrates NADH, methyl viologen, ATP, and L-Leu, L-Phe, or L-Asn, respectively, similar to the assay of HynC, described in section 3.1.8 (for method, see section 5.8.2). After incubation of the one-pot assays, the amino acids were released by alkaline thioester cleavage and detected with LC-MS. Unfortunately, no hydroxylated amino acids were detectable in neither of the assays (data not shown). After numerous attempts to detect LybC activity with all of the available module constructs, the project was discontinued.

3.3 FR900359

The nonribosomal peptide, FR900359 (FR) is a potent Gαq protein inhibitor. Its unique mechanism of action, combined with the strong affinity and selectivity towards its clinically relevant target, led to the detailed investigation by the Research Unit 2372.^[144] From a biosynthetic point of view, the FR BGC *frs*, exhibits several interesting features, among them the tailoring NHDM gene *frsH*. Its product enzyme, FrsH targets T domain-bound L-leucine, to synthesize in total three β-hydroxyleucine (Hle) building blocks, which are essential for the overall structure and the activity of the peptide. One of these is part of the (2*S*,3*S*)-(N-propionyl)-3-hydroxyleucine (*N*-Pp-Hle) side chain, which is assembled by the monomodular NRPS FrsA. Its A domain recruits L-Leu and tethers it onto the adjacent T domain, which is then recognized by FrsH, resulting in the Hle building block. Apart from the crucial A and T domain, FrsA contains a C_{starter} domain that links a propionyl residue, originating from propionyl-CoA, to Hle, and a TE domain that couples the assembled *N*-Pp-Hle side chain with FR-core in an intermolecular transesterification reaction.^[13] The simple architecture makes the FrsA-FrsH heterodimer an optimal example to biochemically investigate the complex nature of tailoring enzymes in NRPS. In this chapter, the results of such investigations are presented. First, the predicted activities of FrsH, as well as the A and C domain of FrsA were confirmed *in vitro*. Secondly, the protein-protein interaction between FrsH and FrsA, and the underlying recognition mechanism was further investigated.

3.3.1 Cloning and heterologous expression of *frs* genes

The enzymatic activity of FrsA and FrsH depend on each other, thus for the *in vitro* reconstitution, different constructs of *frsA*, as well as *frsH* were cloned and heterologously expressed in *E. coli*. The primary experiments were conducted in close cooperation with Dr. Cornelia Hermes, whose project focussed on the investigation of FrsA. All used constructs of FrsA, except FrsA₁AT, were assembled by her or by internship students under her supervision. For the crystallization trails, FrsH was expressed from pHis-TEV, which enables catalytic removal of the octahistidinyl tag of the purified protein by the tobacco etch virus-protease (TEV-protease), kindly provided by Dr. René Richarz. For the *in vivo* complementation, the *C. vaccinii* knock-out strain and the complementation plasmids were also generated and provided by Dr. René Richarz. All proteins used within this project are listed in Table 3.6.

Results

Table 3.6: Cloned and heterologously expressed constructs of the *frs* BGC, used for *in vitro* investigations.

Protein	Molecular weight [kDa]	Tag	Provided by
FrsA ₁ CAT	116.5	C- & N-terminal 6xHis	Cornelia Hermes
FrsA ₁ AT	57.8	N-terminal 6xHis	
FrsA ₁ A	55.2	N-terminal 6xHis	
	63.7	N-terminal 6xHis	
FrsH	62.7	N-terminal 8xHis	
TEV	26.8	N-terminal 6xHis	René Richarz

All listed constructs of the *frs* BGC were successfully cloned and heterologously expressed in *E. coli*. The proteins FrsA₁CAT, FrsA₁AT, and FrsA₁A contain an A domain were shown to depend on the coexpression with the MbtH-like protein (MLP) FrsB for adenylating activity. To ensure *in vivo* phosphopantetheinylation, the strain *E. coli* BAP1 was used for all constructs with T domains. All other proteins were expressed in *E. coli* BL21 (DE3). Figure 3.12 shows SDS-PAGES of the purified proteins.

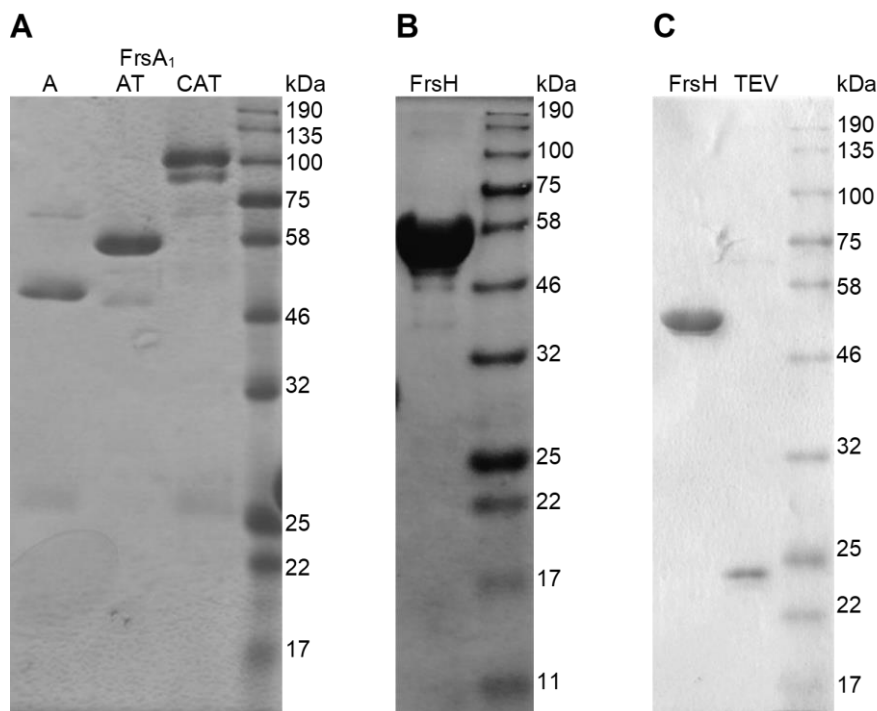


Figure 3.12: SDS-PAGE analyses of heterologously expressed proteins of the *frs* BGC. **A** Constructs of FrsA, purified from *E. coli* BAP1, all coexpressed with the MbtH-like protein FrsB, which is not visible due to its small size of 8.2 kDa. Expected sizes: FrsA₁A: 55.2 kDa; FrsA₁AT: 57.8 kDa; FrsA₁CAT: 116.5 kDa **B** FrsH with N-terminal hexahistidine tag, purified from *E. coli* BL21(DE3). Expected size: 63.7 kDa **C** FrsH with N-terminal 8x His tag (62.7 kDa) and tobacco etch virus-protease (26.8 kDa), which is used to cleave off the tag. Both purified from *E. coli* BL21(DE3).

3.3.2 *In vitro* activity of A domains

Published data about the family of *trans* acting NHDM in NRPS suggest T domain bound amino acids as substrate for the hydroxylase in contrast to the free amino acid. Thus, the successful reconstitution of FrsA was an essential prerequisite for the investigation of FrsH *in vitro*. In a first step, the adenylating activity of FrsA was verified with the $\gamma^{18}\text{O}_4$ -ATP-exchange assay.^[120] Cornelia Hermes performed this assay with the whole NRPS modules FrsA and FrsD. Interestingly, the A domains exhibit different substrate promiscuity, despite their almost identical primary structure.^[13] Apart from its expected substrate L-Leu, FrsA also accepted D-Leu, and, to some extent, Hle, whereas FrsD exclusively adenylated L-Leu.

The adenylating activity of the heterologously expressed constructs FrsA₁CAT, FrsA₁AT, and FrsA₁A was investigated by internship student Tobias Götzen. Subsequent to NiNTA-based affinity chromatography, he purified the enzymes by size-exclusion chromatography to separate possible impurities that might affect the results of the assay. Additionally, he tested the activity after one flash-freezing cycle. The results of his work contributed to our general knowledge about the stability of the proteins. Figure 3.13 shows his results of the $\gamma^{18}\text{O}_4$ -ATP-exchange assay with FrsA₁CAT, FrsA₁AT, and FrsA₁A.

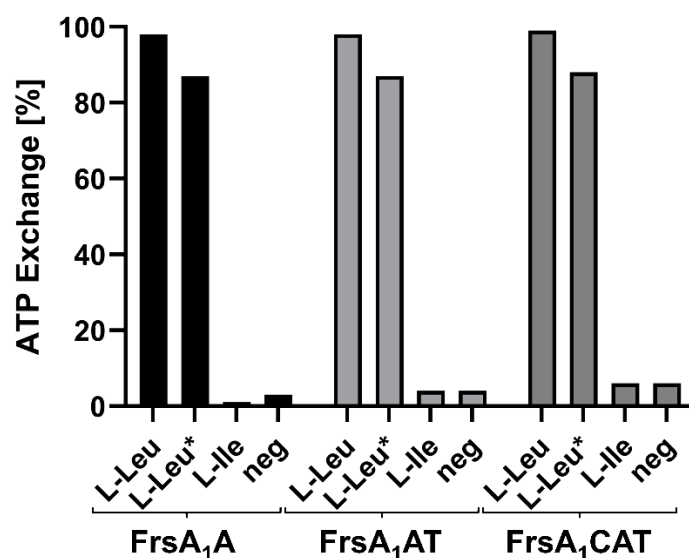


Figure 3.13: Adenylating activity of the A domain in FrsA. The NRPS module was heterologously coexpressed with the MbtH-like protein FrsB with different domain architecture and its specificity towards the substrate L-Leu tested in comparison to the homologue amino acid L-Ile with the $\gamma^{18}\text{O}_4$ -ATP exchange assay. In case of the L-Leu* sample, the enzymes were frozen in liquid nitrogen and thawed prior to use to assay their stability.

The results show the strong affinity of the A domain towards its predicted substrate, L-Leu, whereas L-Ile is not accepted at all. Cornelia Hermes already tested the adenylating activity of the whole FrsA construct, also after heterologous expression and purification from *E. coli* BAP1. She reported a rather loose substrate specificity, as the A domain also accepted D-Leu and Hle as substrates. For

L-Ile, she measured an ATP exchange of 9.5 %, which was a significant increase compared to the negative control (1.1 %).^[13] In contrast to her results, the A domain of the truncated constructs did not activate the non-native substrate L-Ile in comparison to the negative control. The differences to the results with the whole FrsA construct are probably due to the additional purification step. Minor proteinogenic impurities are known to affect the results of this assay if their catalytic activity relies on the hydrolysis of ATP. This was already reported for the case of HrmO1_A, the A domain of hormaomycin biosynthesis that is responsible for the recruitment of (3-nitrocyclopropionyl)-alanine. Here, additional purification steps via ion exchange chromatography removed false positive activities towards L-Thr and L-Trp.^[122] Moreover, the whole FrsA also harbours a C terminal TE domain, which might interfere with the A domain and thereby influence its substrate specificity. So far, only a single study reported the successful structural characterization of a complete terminal NRPS module. In this case, the TE domain is located alongside the A domain, coordinated by a small interface,^[19] that might influence its adenylating activity. However, it is unlikely that the observed substrate promiscuity of the FrsA A domain derives from the additional TE domain, which rather contributes to the native state of the enzyme with high specificity.

Flash-freezing did not seem to affect the adenylating activity of any construct. The slight reduction of the exchange rates probably originates from a reduced protein concentration due to addition of glycerol as cryoprotectant. These results count as proof of principle that the enzymes of the *frs* BGC are able to maintain their catalytic activity after at least one flash-freezing cycle, which facilitates the storage of overexpressed enzymes and simplifies working procedures.

Interestingly, the activity of the A domains within the different constructs is not rendered at all by the absence or presence of adjacent domains, which is in stark contrast to the results of the $\gamma^{18}\text{O}_4$ -ATP-exchange assay of the *lyb* BGC (see section 3.2.2).

3.3.3 UV/Vis spectrum of FrsH

Comparable to the NHDM LybC, FrsH harbours a diiron cluster and thus absorbs light at $\lambda=340$ nm in the ferrous state (see section 3.2.3). Figure 3.14 shows the UV/Vis spectrum of concentrated ferrous FrsH upon purification and after addition of methyl viologen as an electron transmitter and sodium dithionite as a chemical reducing agent in the ferric state. Upon reduction, the feature at 340 nm diminished. This result gave evidence that FrsH is in fact a NHDM, so its activity was further investigated *in vitro*.

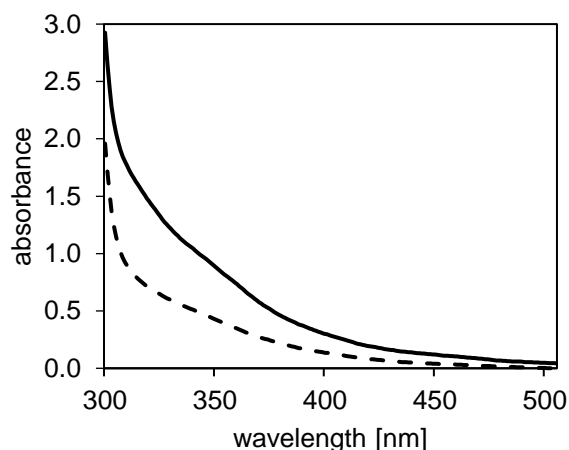


Figure 3.14: UV/Vis-spectrum of FrsH. The enzyme is isolated from *E. coli* BL21(DE3) in ferrous state (solid line), showing a small feature at $\lambda=340$ nm, which is typical for diiron coordinating enzymes. Chemical reduction of the diiron centre the ferric state (dashed line) by sodium dithionite in the presence of methylviologen diminishes the feature.

3.3.4 In vitro activity of FrsH

The proposed function of FrsH to generate Hle building blocks was investigated with the monomodular NRPS FrsA, which generates the FR side chain *N*-Pp-Hle. Together with Dr. Cornelia Hermes, an assay was designed to assemble the whole side chain in a one-pot reaction. Therefore, two assay principles were combined. The first described the functionality of the initiation module in surfactin biosynthesis *in vitro*. The authors discovered that the C_{starter} domain uses a CoA-bound acyl residue as substrate for the donor binding site to subsequently perform the common condensation reaction with the amino acids in the acceptor site.^[145] The second assay was used to characterize the prototype NHDM CmlA^[54] and was also deployed to characterize the NHDM HynC from hypeptin biosynthesis described further in section 3.1.8.

Briefly, for the side chain assembly assay, the purified enzyme FrsA₁CAT was mixed with L-Leu, ATP, and acetyl-CoA as substrates. FrsH was activated by incubation with NADH as reducing agent and methyl viologen as electron transmitter and added to the reaction. As negative control, heat-inactivated FrsH was used. After incubation, the product was released from the T domain by

alkaline thioester cleavage and analysed via LC-MS. The authentic standard *N*-Pp-Hle was chemically synthesized by Dr. Cornelia Hermes in cooperation with Dr. Jim Küppers (working group of Prof. Gütschow, University of Bonn). Figure 3.15 shows LC-MS traces of the extracted *in vitro* hydroxylation assay with FrsH and FrsA₁CAT. Strikingly, *N*-Pp-Hle was detected in the assay, unambiguously proving the proposed functions of FrsA and FrsH, which together synthesize the FR side chain.

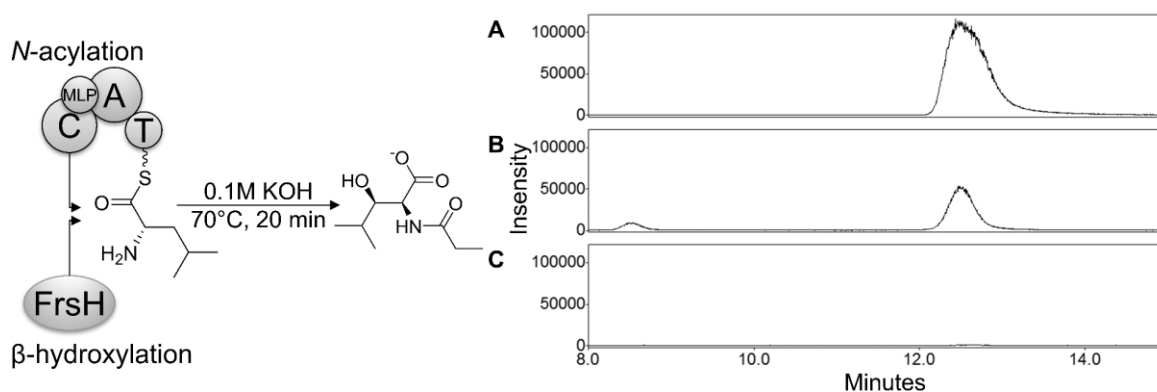


Figure 3.15: FR side chain assembly *in vitro*. The A domain of FrsA₁CAT binds L-Leu to the T domain, which is then further modified by FrsH and FrsA₁C. After incubation, the product *N*-Pp-Hle is liberated by alkaline thioester cleavage and detected via LC-MS ($m/z = 202.1$). **A** Chemically synthesized *N*-Pp-Hle (1 $\mu\text{g/mL}$). **B** Assay of FrsA₁CAT and FrsH, incubated with all necessary substrates. **C** Negative control of FrsA₁CAT and heat-inactivated FrsH, incubated with all necessary substrates.

Cornelia Hermes conducted further experiments to test the substrate specificity of the A domain and C domain of FrsA. Replications of the side chain assembly assay with D-Leu and L-Ile as amino acid substrates also yielded products with the same exact mass and retention time, but in lower yields, suggesting that the isomers of L-Leu are also accepted for side chain assembly, to some extent. In a next step, Cornelia changed the substrate for the C domain from propionyl-CoA to acetyl-CoA or butyryl-CoA, which led to the formation of (2*S*,3*S*)-(N-acetyl)-3-hydroxyleucine and (2*S*,3*S*)-(N-butyryl)-3-hydroxyleucine, respectively. The substrate promiscuity of the C domain *in vitro* was then utilized *in vivo*, to generate new FR-derivatives in larger scale by genetical engineering and precursor feeding.^[13]

For this project, these results are of utmost interest. It has already been reported that *trans* acting CYP450s and NHDMs accept a variety of amino acids, related to their natural substrate, but not the respective stereoisomers.^[42,60] In contrast, FrsH was able to produce a hydroxylated product from D-Leu, pointing out the inferior relevance of the T domain bound amino acid for substrate recognition of FrsH.

3.4 Investigations of the substrate recognition of FrsH

To date, hardly anything is known about the substrate recognition of *trans* acting hydroxylases in NRPS. Their nature as independent enzymes theoretically provides access to all NRPS modules and their bound substrate. Nevertheless, only specific, cognate modules are targeted. This poses questions about the underlying recognition mechanism, which is still elusive. So far, it is known that *trans* acting NHDM depend on the A domain of the targeted NRPS module,^[60] in contrast to *trans* acting CYP450s, which seem to solely target the T domain.^[42] This chapter describes the results gathered to elucidate the mechanism of substrate recognition of *trans* acting NHDM in NRPS. First, bioinformatic methods were applied to generate evidence from comparing the different primary sequences of NRPS modules in BGC that harbour NHDMs. Second, the protein-protein interaction of FrsH with FrsA was characterized *in vitro* by size-exclusion chromatography and isothermal titration calorimetry. Third, the already described NHDMs HynC and LybC were used to replace FrsH *in vivo*, allowing direct comparison of the enzymes. Lastly, structural data of FrsH were generated, and subsequently used for *in silico* docking studies, which guided the selection of candidate residues for mutagenesis studies to disrupt the interaction with FrsA.

3.4.1 Bioinformatic analysis on the recognition mechanism of *trans* acting NHDM

Bioinformatic analysis of the *rmy* BGC (NCBI Accession No: AL646053.1), responsible for the formation of ralsolamycin in *Ralstonia solanaceum*^[146] revealed the putative NHDM RmyC to be responsible for β -hydroxylation of L-Tyr in module 3. The amino acid alignment of all T domains from the *rmy* BGC, revealed that the T₃ and the T₄ domain are exactly identical and three further T domains share a sequence identity of >95% with the T₃ domain (see Appendix Figure 8.1). Nevertheless, RmyC only targets module 3. This fact gives additional support for the hypothesis that tailoring NHDMs do not recognize their substrate NRPS module by the structure of the T domain. Furthermore the NHDMs HynC and CmlA from hypeptin and chloramphenicol biosynthesis, respectively, were shown to distinguish the AT didomain of cognate modules as substrate, to hydroxylate bound L-amino acids (see chapter 3.1.8),^[34,54] leading to the conclusion that structural features, necessary for substrate recognition, have to be located within the A domain or the AT domain linker region. To identify candidate amino acid residues, a bioinformatic approach was applied. The primary structures of all AT didomains encoded in the BGCs of FR (NCBI Accession No: MT876545.1), hypeptin (CP072597.1), lysobactin (JF412274.1), chloramphenicol (FR845719.1), teicoplanin (AJ632270.1), and ralsolamycin (AL646053.1) were extracted and aligned to identify any residues that are conserved within the modules that are substrate to a NHDM. Unfortunately, from this approach, no specific residues could be identified (data not shown).

In the next approach, the alignments were further utilized to assess, if phylogenetic analyses facilitate to distinguish between A domains, that are responsible for NHDM recruitment and those that are not. The superfamily of C domains^[135] and, to some extent, TE domains^[13] were already shown to cluster dependent on their catalytic function, and not on of their hosts evolutionary relationship. A domains sometimes cluster depending on their substrate amino acid with high accuracy, but oftentimes, no prediction based on phylogeny is possible.^[147] To test if recruitment of tailoring NHDMs influences their phylogenetic clustering, the abovementioned alignment of A domains was used to construct a phylogenetic tree (see Figure 3.16). The branches of the resulting tree that are supported by high bootstrapping values, either separate due to the substrate amino acids, or the taxonomy of the host strains. Modules that recruit tailoring NHDM are distributed over the whole tree without observable pattern. The analysis was repeated with the interdomain linker sequences only, to eliminate the influence of the substrate amino acid. This led to a severe drop of bootstrapping support, but did not alter the distribution of NHDM recruiting modules throughout the phylogenetic tree (data not shown).

Taken together, the mechanism of recognition of tailoring NHDMs towards their substrate NRPS is not trackable by solely analysing the primary structures of domains which are cognate or non-cognate by using basic bioinformatic tools. This is in line with the results of Makris et al., who also reported the lack of prominent specificity-conferring structures on CmlA, as well as on the cognate NRPS module.^[59]

3.4.2 Size-exclusion chromatography

In size-exclusion chromatography (SEC), the stationary phase consists of an inert material that is not meant to interact with the mobile phase or the sample material. Instead, it is highly porous, whereby molecules, depending on their size, cannot enter all pores. This leads to a lower volume accessibility and faster elution of larger molecules. Thus, SEC is a simple method to detect stable protein-protein interactions (PPI). If the interacting proteins remain bound to each other as a stable dimer throughout the chromatographic separation process, described by a low offloading rate,^[61] they elute in a single fraction with a shorter retention time than each of the monomers. Alternating the set-up of the single chromatographic runs, i.e. pH or ionic strength of the buffer, may give rough indications about the stability, stoichiometry and driving forces of the interaction.^[40] Here, the method was used to generally confirm the interaction between the proteins FrsH and FrsA₁CAT. Both proteins were heterologously expressed and purified from *E. coli* BL21(DE3) or *E. coli* BAP1. The *holo* T domain was eventually loaded by incubation with L-Leu und ATP. The enzymes were then mixed in a molar ratio of 1:1 and, after necessary preparation steps, injected into the chromatographic system. As a reference, the single proteins were treated equally and also subjected to SEC. Eluting fractions were collected for subsequent SDS-PAGE analysis.

Figure 3.17 **A** shows the results of the SEC runs, conducted with FrsH and FrsA₁CAT, both purified from *E. coli* BL21(DE3). The combined proteins elute in a broad peak with two tips at roughly the retention times of the single proteins, indicating that the prosthetic group and the substrate amino acid are crucial for a stable interaction. A slight shift in the retention volume of FrsH, compared to the reference indicates a weak interaction of the proteins. This might either derive from partial loading of the T domain by native phosphopantetheine transferases of *E. coli in vivo*, leading to fully functional FrsA₁CAT after purification, or by FrsH displaying some residual affinity towards non-loaded FrsA₁CAT. In the next step, the expression host of FrsA₁CAT was switched to *E. coli* BAP1, to ensure complete phosphopantetheinylation of the T domain *in vivo*. Figure 3.17 **B** shows the chromatograms of the proteins FrsH and FrsA₁CAT alone and combined after incubation. Strikingly, the combined proteins elute in a single peak with a smaller retention volume, compared to the retention volumes of the single proteins. Analysis of the collected fractions *via* SDS-PAGE in Figure 3.17 **C** verified the reduction of FrsHs retention volume by roughly 500 μ l. This result visualizes the interaction of FrsH with its loaded substrate NRPS module FrsA. These results are in line with the observations of Haslinger *et al.* They titrated the CYP450 P450_{sky} with chemically synthesised substrates that mimicked either the single amino acid, the phosphopantetheine-bound amino acid or the T domain-bound amino acid. The resulting kinetic data showed that the coordination of phosphopantetheine-bound amino acid in the binding pocket is crucial for tight binding of the hydroxylase to the carrier protein.^[43] Here, the results of analytical SEC demonstrate a similar behaviour of the NHDM FrsH.

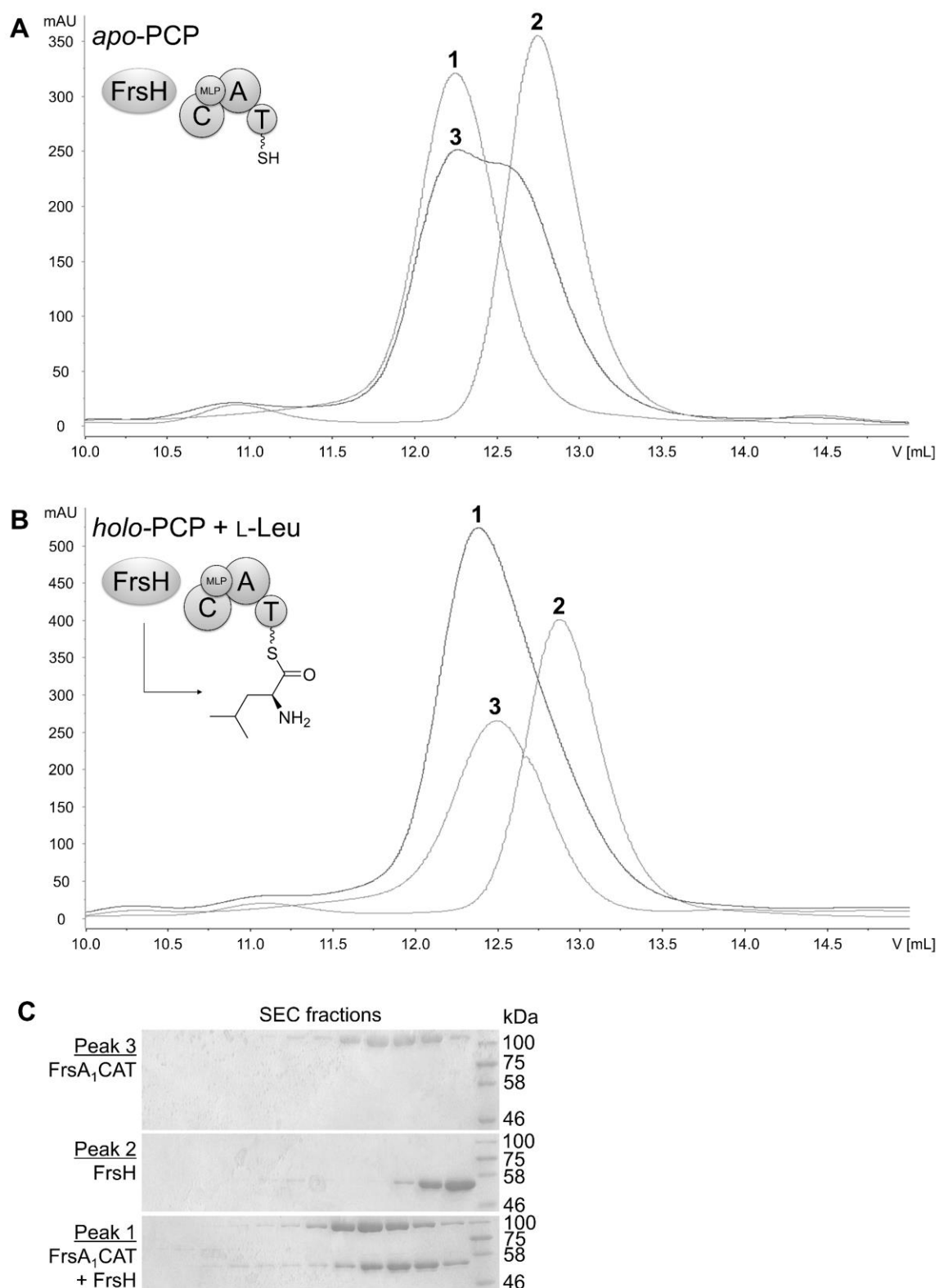


Figure 3.17: Overlay of size-exclusion chromatograms to visualize the interaction of FrsA₁CAT and FrsH. After purification from *E. coli*, the enzymes were incubated alone (1: FrsA₁CAT; 2: FrsH) or equimolar mixed (3), prior to sample preparation and injection into the chromatographic system. **A** FrsA₁CAT was purified from *E. coli* BL21(DE3) to prevent phosphopantetheinylation of the T domain and loading with the substrate L-Leu, leading to separate elution of FrsH and FrsA₁CAT. **B** FrsA₁CAT was purified from *E. coli* BAP1 in *holo* state and incubated with L-Leu and ATP to load the T domain with the substrate amino acid. The proteins elute in a single peak with shorter lower retention time, compared to the references. **C** SDS-PAGE analyses of collected fraction of the chromatographic experiments displayed in **B** verify the reduced retention volume of FrsH in the mixed sample.

Unfortunately, these results were not reproducible. In subsequent uniform experiments, FrsH and FrsA₁CAT always eluted separately without any shifts in the retention volume, compared to the single proteins. The gathered UV/Vis spectroscopy data (see chapter 3.3.3) indicated that FrsH was mostly purified in activated, ferric state from *E. coli*. This might lead to hydroxylation of the loaded L-Leu substrate and subsequent dissociation of FrsH from FrsA₁CAT, even before loading into the chromatographic system. To circumvent this difficulty, the inactive mutant FrsH E376D was generated. Analogously to CmlA E377D^[58] and HynC E376D (see chapter 3.1.8), the glutamate E376 in FrsH regulates the access of oxygen to the diiron cluster of the active site. The rationale to insert this mutation was that the diiron cluster is oxidated after purification, which renders the enzyme catalytically inactive, without negative impact on the affinity towards the substrate FrsA₁CAT. However, repeating the experiments with FrsH E376D and L-Leu-loaded FrsA₁CAT still resulted in two separate peaks without altered retention volume in comparison to the single proteins.

Table 3.7: Summarized conditions that were tested to visualize the interaction of FrsA₁CAT and FrsH by size exclusion chromatography. For each single change, an independent experiment was conducted.

	Buffer	Loading of T domain	Binding of FrsH
Standard protocol	50 mM Tris 150 mM NaCl 10 mM MgCl ₂ pH 7.5	15 μM FrsA ₁ CAT 1 mM ATP 1 mM L-Leu Incubation: 20 °C, 20 min	15 μM FrsA ₁ CAT 15 μM FrsH Incubation: 20 °C, 30 min
Changes	50 mM HEPES 150 mM NaCl 10 mM MgCl ₂ pH 7.5 50 mM Tris 25 mM NaCl 10 mM MgCl ₂ pH 7.5 50 mM Tris 500 mM NaCl 10 mM MgCl ₂ pH 7.5	+0.5 mM propionyl-CoA	15 μM FrsA ₁ CAT 30 μM FrsH 30 μM FrsA ₁ CAT 15 μM FrsH 15 μM FrsA ₁ CAT 15 μM FrsH E376D +5 mM DTT Concentrate the mixed proteins to 50 μM

In subsequent experiments, single conditions that are known to influence protein-protein interactions^[148] were varied in independent experimental trials: The ionic strength of the buffer was changed by a higher and lower concentration of NaCl, which is important if the interaction primarily relies on hydrophilic or hydrophobic contacts. The buffer system was switched to HEPES because Tris is known to interact with proteins to stabilize their native state and has a temperature-dependent pK_a .^[149] FrsA₁CAT was additionally supplemented with propionyl-CoA, for the case that the activity of the C domain influences recruitment of FrsH. Lastly, the molar ratio of the enzymes was changed from 1:1 to 2:1 and 1:2. The tested conditions are summarized in Table 3.7. Unfortunately, the interaction was not observable in any of the conditions tested. These results hint towards an interaction with a high offloading rate of FrsH from FrsA. Due to the transient dimeric state, the interaction is hardly detectable with SEC, which requires rather stable interactions for detection. In FR biosynthesis, FrsH is responsible for the hydroxylation of three T domain tethered L-Leu residues. The fast dissociation of FrsH from FrsA might thus contribute to the efficient assembly of FR, as a longer residual time at one module impedes the hydroxylation process at the latter two modules.

3.4.3 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is commonly used to characterize the interaction of two binding partners that can either be small molecules or proteins. Compared to other techniques like surface plasmon resonance or analytical ultracentrifugation, all thermodynamic parameters of the occurring reaction (ΔH , ΔS , ΔG), together with the dissociation constant K_D and the stoichiometric relation of the reactants, can be measured within a single experiment.^[150] Alternating the temperature of the experiments enables calculation of the heat capacity change (ΔC_p), which directly correlates with the dehydration of hydrophilic residues, driving the interaction.^[151] With the recent development of extremely sensitive instruments, even kinetic parameters can be recorded from the data.^[152] To measure an interaction, one of the binding partners is titrated into the other in small steps. Depending on its thermodynamic properties, the reaction consumes or emits energy that is quantified as a change of temperature. Once all binding sites of the reaction partner in the reaction chamber are saturated, the difference in temperature after additional titration steps converges to nearly zero. The slope of this conversion depends on the dissociation constant K_D , and the turning point within it enables calculation of the stoichiometry of the reaction. Based on the concentration of the reactants and the emitted heat after each titration step, the thermodynamic parameters ΔH , ΔS , and ΔG can be calculated.^[153]

FrsH and FrsA₁CAT were expressed and purified from *E. coli* BL21 or *E. coli* BAP1, respectively. After NiNTA-based purification, FrsA₁CAT was incubated with L-Leu and ATP to load the T

domain. To ensure high purity of the proteins, they were further subjected to preparative SEC. This step also transferred the proteins into the exactly identical buffer system, which is crucial for ITC measurements. The fractions of the single proteins were collected, pooled and concentrated to roughly 1 mM in the case of FrsH (~60 mg/mL) and 0.1 mM in the case of FrsA₁CAT (~12 mg/mL). ITC measurements were conducted on a MicroCal PAEQ-ITC instrument (Malvern Panalytical) in cooperation with Niels Schneberger from the working group of Prof. Dr. Matthias Geyer (Institute of Structural Biology, University Hospital Bonn). The measurements were hampered by the surprisingly low stability of FrsA₁CAT in the reaction chamber after injection of FrsH. This led to precipitation of the proteins, making it unfeasible to record data of the respective measurement. Additionally, the single injections yielded only low amplitudes in the differential power despite high titration volumes. Finally, data of a single experiment were of decent quality for further analysis with the triple software set of NITPIC,^[154] sedphat,^[155] and Gussi.^[156] The results are depicted in Figure 3.18. As mentioned above, the amplitude in the thermogram after each injection is rather low, compared to other binding studies of tailoring hydroxylases to NRPS.^[123] After each injection, the differential power first becomes negative, then rapidly deflects to positive values before dropping into slightly negative values again. The opposed amplitudes seem to cancel each other out, which might be a reason for the overall low heat of injection. The strong positive and negative deflections within a single injection indicate larger structural rearrangements prior or post binding of the two protein, which is a known obstacle of ITC measurements.^[61] This is a major drawback for the characterization of whole NRPS modules via ITC, as NRPS essentially rely on shuttling of the substrate between the domains by the highly flexible T domain. In resting state, the L-Leu loaded T domain either sticks to the A or the C domain. It is also probable that the T domain constantly switches between the two recipient domains in an equilibrium, balanced by the affinity towards each binding partner, as it has been observed in cryo-EM density maps of FmoA3 of JBIR-34 and -35.^[17] Yet it is improbable that, after thiolation, the loaded T domain dissociates from the A domains active site and stays unbound in solution.^[18] Once FrsH is titrated to FrsA₁CAT, it thus competes with the A and the C domain for interaction with the T domain, thereby inducing major rearrangements of the NRPS, which impedes the measurement of the actual binding reaction. In this case, the obtained data were the result of a so called “displacement experiment”, which is an appropriate method to determine the thermodynamic parameters of high affinity binding reactions if the binding parameters of the displaced reactant are already known.^[150]

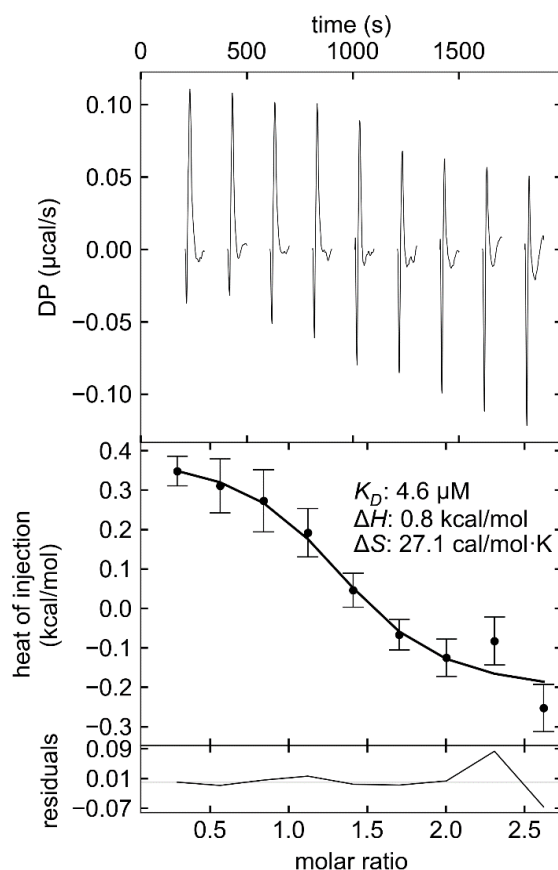


Figure 3.18: Isothermal titration calorimetry to determine the thermodynamic parameters of the association of FrsH to FrsA₁CAT. FrsH (0.982 mM) was titrated into FrsA₁CAT (0.071 mM) with a PEAQ-ITC device. Upper: Thermogram visualizing the measured differential power after each injection in dependency of the time. Middle and lower: Emitted heat per injection with residuals from the employed fit. The calculated thermodynamic parameters are given in the middle box. Graphical visualization was performed with Gussi.

Regarding these limitations, the measured parameters of the binding reaction of FrsH to FrsA₁CAT have to be evaluated carefully. Additionally, the association of *trans* acting enzymes to NRPS have rarely been investigated in a quantitative manner, so data for comparison are scarce. The calculation of ΔG , ΔH , and ΔS depend on the absolute values of differential power after each injection. Thus, these calculated values most likely do not correctly describe the association of FrsH to the T domain and will not be discussed. The observed K_D in low micromolar range (4.6 μM) is comparable with the association of *trans* acting CYP450s to recruiting X domains in glycopeptide biosynthesis.^[123,136] However, in skyllamycin biosynthesis, the CYP450 P450_{sky} binds directly to its cognate T domain with a much higher K_D of 89 to 93 μM .^[42,157] Despite all restraints, these data hint towards a different mechanism of interaction between the two in *trans* acting hydroxylating systems of NHDMs and CYP450s.

Taken together, it appears that SEC and ITC are not adequate methods to characterize the transient interaction of NRPS and *trans* acting enzymes. Especially the flexibility of the NRPS probably hampers the detection. Other methods like dynamic light scattering or surface plasmon resonance

might be more suitable for the detection and quantification of such interaction. A more sophisticated approach would include stabilization or trapping of the enzymes in interacting state, which is commonly accomplished to crystallize complexes in otherwise transient states. Haslinger *et al.* utilized this method to characterize the affinity of P450_{sky} towards its cognate T domain. Therefore, the *holo* T domain was loaded with a nitrogen-containing aromatic inhibitor that formed a stable complex with the central iron atom of the hydroxylase.^[43] The exact same inhibitor should be applicable to visualize the interaction of FrsH and FrsA₁CAT and might also facilitate the crystallization of the complex for structure elucidation.

3.4.4 Substitution of *frsH* in vivo

So far, the approaches to describe the interaction between FrsH and FrsA and to characterize the structural mechanism of interaction were only partly successful. Each approach gave informative hints about the nature of the interaction, but none yielded distinct evidence to unravel the underlying driving forces. During the course of this work, a study was published, describing the exchangeability of the native NHDM of teicoplanin biosynthesis (Tcp25) by CmlA from chloramphenicol biosynthesis *in vitro*.^[60] The compatibility of NHDM in unrelated NRPS systems indicates a shared mechanism of interaction, which gives useful hints about the interacting interfaces of the involved enzymes, as the necessary surface patches have to be functionally conserved. In this chapter, the exchangeability of the NHDM FrsH from FR biosynthesis with the homologous enzymes HynC and LybC from the synthetases of hypeptin and lysobactin, was assessed. In contrast to the former study of CmlA and Tcp25, the substitution of FrsH was conducted *in vivo*. In a first step, the gene *frsH* was knocked-out in the natively FR producing strain *C. vaccinii* MWU205 by René Richarz. The functionality of this knock-out system was already proven by deletion of the gene *vioA*, which disrupted the production of the purple pigment violacein and thereby enhanced the production titres of FR.^[13]

The *frsH*-deficient mutant of the native FR producer *C. vaccinii* MWU205, *C. vaccinii* Δ *frsH*, was kindly provided by René Richarz. Fermentation of *C. vaccinii* Δ *frsH*, butanolic extraction, and subsequent LC-MS measurements validated the abolishment of FR production in this strain, highlighting the crucial role of FrsH in the biosynthesis of FR (Figure 3.19). In a next step, *frsH* was cloned from *C. vaccinii* MWU205 into the pCv1. This plasmid originates from pMLBAD^[158] and was modified by René Richarz for utilization in *C. vaccinii*. As *C. vaccinii* MWU205 is unable to assimilate arabinose,^[159] the thereby inducible promotor region of pMLBAD, as well as the *araC* gene were replaced. Instead, the naturally occurring, quorum-sensing mediated promotor region of the violacein operon *vioP* was introduced. Additionally, the trimethoprim resistance cassette of pMLBAD was exchanged with *tetA*, which enables resistance against tetracycline and finally

resulted in pCv1. The constructed plasmid pCv1::*frsH* was inserted into *C. vaccinii* Δ *frsH* by biparental conjugation from *E. coli* NEB 10- β , moderated by the immobilized helper plasmid pTA-Mob,^[160] yielding the strain *C. vaccinii* Δ *frsH* + *frsH*. Strikingly, LC-MS measurements of the butanolic extracts from *C. vaccinii* Δ *frsH* + *frsH* validated the reestablishment of FR production. Despite lower product yields, this counts as proof of principle for the herewith established knock-out and complementation system in *C. vaccinii* MWU205. Analogously to the steps before, the genes encoding the NHDM HynC and LybC were used to complement *C. vaccinii* Δ *frsH*. The NHDM HynC from *Lysobacter* sp. K5869 was successfully characterized in chapter 3.1.8. It is responsible for the β -hydroxylation of L-Leu and L-Tyr in the biosynthesis of the antibiotic hyeptin.^[34] Despite their validated homology, FrsH and HynC share only a small overall sequence identity of 34%. Most of the identical residues are essential for the correct three dimensional structure and are conserved throughout the family *trans* acting NHDMs in NRPS.^[54] The putative NHDM LybC from *Lysobacter* sp. ATCC53042 belongs to the BGC of lysobactin (*lyb*), but its activity was not yet experimentally validated. Bioinformatic analyses suggest a broad substrate specificity, as is the only enzyme encoded in the *lyb* BGC, which is candidate for β -hydroxylation of L-Leu, L-Phe, and L-Asn.^[35] The attempts to characterize the enzyme *in vitro* within this work are described in chapter 3.2.4, but were not successful. Similar to HynC, FrsH and LybC share only 33% of sequence identity (see appendix Figure 8.2).

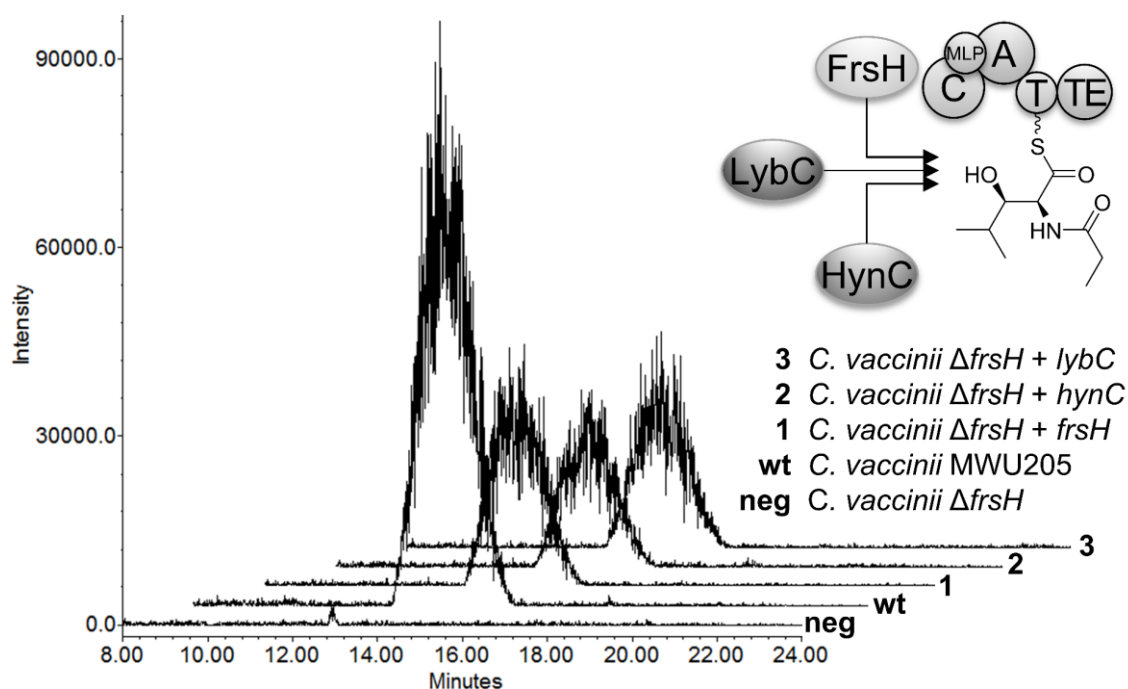


Figure 3.19: FR production in *C. vaccinii* MWU205 knock-out and complementation strains. Single ion record LC-MS traces at $m/z=1002.5$ (FR) validate the abolishment of FR production in the *frsH* deficient strain (neg), compared to the wild type (wt). Complementation of *C. vaccinii* Δ *frsH* with *frsH*, as well as genes of the homologous NHDMs *hynC* and *lybC* reconstitute FR production.

Introduction of pCv1::*hynC* and pCv1::*lybC* into *C. vaccinii* Δ *frsH* reconstituted FR production without constraints compared to the *frsH*-complemented strain. Apart from the successful characterization of LybC as a NHDM, these results demonstrate a compatible mechanism of interaction in the three hydroxylating systems of FR, hypeptin, and lysobactin biosynthesis. Thus, the necessary structural features of the hydroxylase and the cognate NRPS module, e.g., the interacting surface patches have to be functionally conserved. Together with the described exchangeability of Tcp25 with CmlA,^[60] these results strongly suggest a conserved mechanism of substrate recognition throughout the whole family of *trans* acting NHDMs in NRPS. Nevertheless, more specified investigations are needed to establish a generally applicable rule.

3.4.5 Structure elucidation of FrsH

Prior to this thesis, the interface of FrsH for the interaction with NRPS, which probably also harbours structures conferring the specificity towards cognate modules, was unknown. Bioinformatic approaches to extract the responsible residues from the primary sequence or even the three dimensional structure were already conducted in studies about CmlA, but did not yield satisfying answers.^[54,59] Apart from CmlA, no further structural data of *trans* acting NHDMs are reported, impeding the validity of any hypotheses that are generated from CmlA and intended to describe the whole enzyme family. Thus, trials to obtain a structure of FrsH by X-ray crystallography were conducted. To date, this is one of the most powerful, accurate, and reliable methods to elucidate proteinogenic structures.^[161] In a common crystallization trail, the protein of interest is isolated in supreme purity, ideally in native state and without any tag from preceding purification processes. The protein is then submitted to a screening approach to elucidate optimal conditions for nucleation and crystal growth. Subsequent optimization trails might be necessary to obtain a regular, pure monocrystal, which is inevitable for high-quality X-ray diffraction measurements.^[162] The crystallization trails of FrsH were conducted in cooperation with Niels Schneberger and Dr. Gregor Hagelücken from the working group of Prof. Dr. Matthias Geyer (Institute of Structural Biology, University Hospital Bonn). Niels Schneberger was responsible for the screening approaches to obtain a protein crystal, as well as all further preparation steps, X-ray diffraction measurements, data evaluation, and structure elucidation. Hence, only the preparative steps for protein crystallization are described here.

To elucidate the native structure of FrsH, the encoding gene *frsH* was cloned into the expression plasmid pHis8-TEV. This plasmid encodes for the recognition site of the tobacco etch virus protease (TEV) between the multiple cloning site and the N-terminal octahistidine tag.^[163] This set-up enables the specific removal of the tag post NiNTA-based affinity purification of the protein from the cell lysate of *E. coli* BL21(DE3). The successfully cleaved FrsH was then separated from the

protease and uncleaved FrsH by inverse Ni-NTA based chromatography where it already elutes by addition of minor concentrations of imidazole. In a final step, FrsH was further purified by preparative size-exclusion chromatography.

In a first attempt, the optimal stoichiometry of FrsH and TEV for His tag removal was elucidated. Therefore, the proteins were incubated overnight at 4°C in the ratio (FrsH:TEV) of 2:1, 10:1, and 100:1. After incubation, the proteins were submitted to inverted NiNTA chromatography, and the resulting fractions analysed via SDS-PAGE (Figure 3.20 A). Independently of the stoichiometry, only minor fractions of FrsH were not cleaved by TEV and are thus visible in the elution fraction, whereas the cleaved FrsH already eluted in the flow through and wash fraction. Remarkably, at the ratio of 100:1, the amount of TEV is too small to be detectable by Coomassie staining, thus, for all subsequent trials, the removal of the octahistidine tag was conducted by incubation of FrsH and TEV in a ratio of 1:100, thereby reducing putative contaminations that might get introduced with the TEV fraction.

The fractions of untagged FrsH in Figure 3.20 A already show a high purity. Nevertheless, the flow through and wash fractions of the inverted NiNTA chromatography were pooled and submitted to preparative size-exclusion chromatography (SEC) as a final “polishing” step. Here, minor proteinogenic contaminants, should get removed. Additionally, some proteins tend to form aggregates in solution that are not observable in SDS-PAGE analysis,^[164] but would impede the crystallization trails. The SEC was also used to transfer the protein into a “minimal” buffer system (25 mM Tris pH 7.5), to reduce any interferences with the crystallization buffers. Figure 3.20 C shows a representative chromatogram of FrsH without tag and Figure 3.20 B the SDS-PAGE analysis of the collected fractions. FrsH elutes after 13.4 mL in a distinct, well-shaped peak. Two minor impurities elute separately before and after the main peak. The fractions 23-29 of this chromatogram were pooled, concentrated to 10 mg/mL (166 µM) and transferred to Niels Schneberger for crystallization trials.

Several standard kits for crystallization buffer screenings (in total 384 conditions) were applied in a sitting drop vapor diffusion approach. In this set-up, 2 µl of the protein containing solution were mixed with 2 µl of the buffer on a plateau of a sealed chamber. The chamber additionally contains a larger reservoir of the buffer. Due to the higher concentration of buffer components in the reservoir, compared to the protein solution, which was diluted by addition of the protein, the water of the protein solution slowly evaporates. This leads to continuously increasing concentrations of the protein and buffer components. Eventually, the protein becomes supersaturated in solution and does not simply precipitate but nucleates as a crystal. Once a nucleus is formed, further protein molecules form additional layers of the growing crystal. Both phases, nucleation and crystal growth, heavily depend on, sometimes different, conditions and components of the surrounding buffer.^[162]

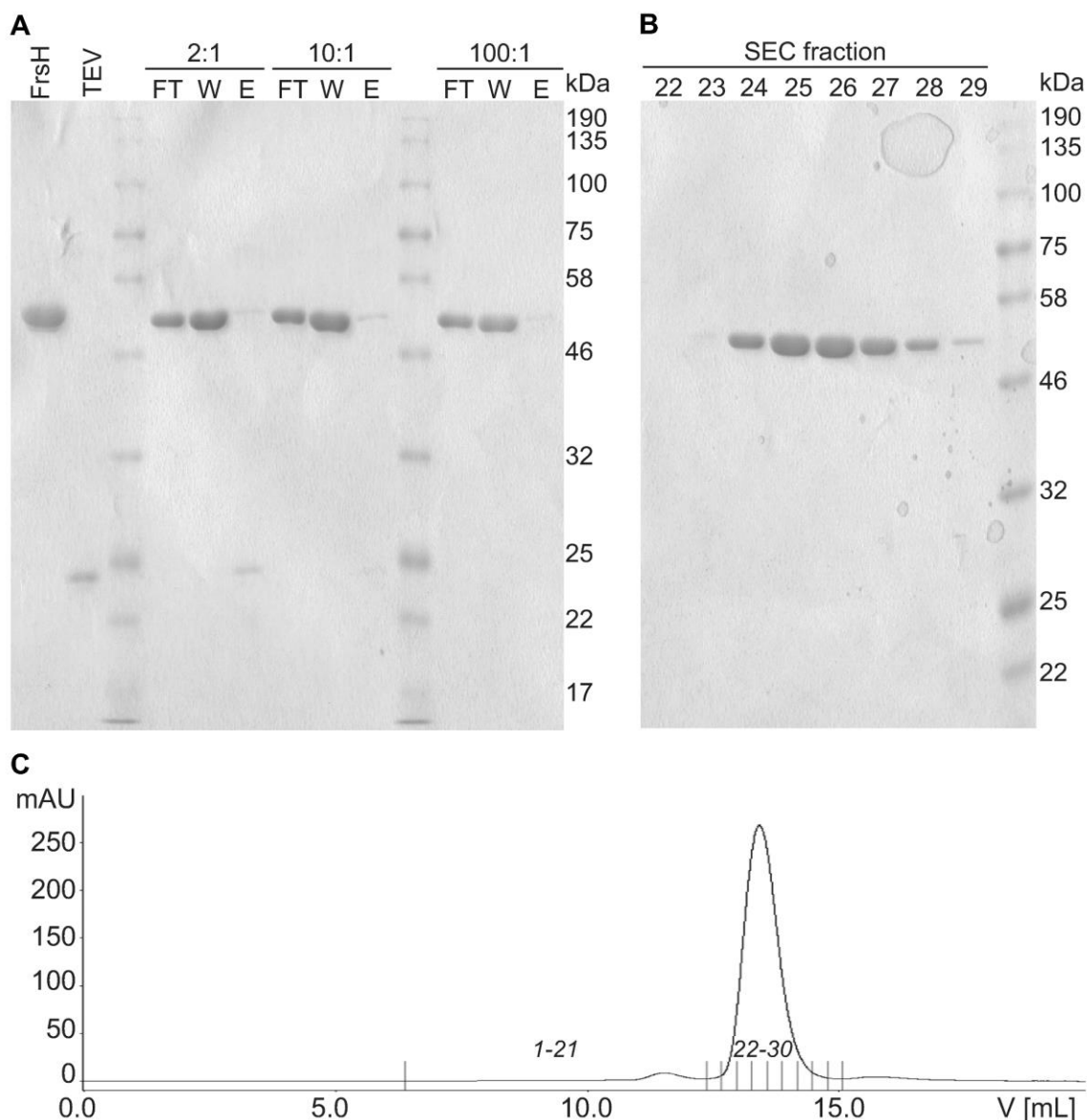


Figure 3.20: Purification of FrsH for crystallisation trials. **A** After NiNTA-based purification, FrsH (with tag: 62.7 kDa; without tag: 60.5 kDa) and tobacco etch virus-protease (TEV, 26.8 kDa) were incubated in different ratios to optimize conditions for His tag removal. After incubation, the proteins were subjected to NiNTA-based chromatography and the collected fractions, together with the as purified proteins analysed via SDS-PAGE (FT: flow through; W: wash 35 mM imidazole; E: elution 300 mM imidazole). **B** The FT and W fractions, containing untagged FrsH were further purified with size-exclusion chromatography. The collected fractions were analysed via SDS-PAGE. **C** Representative chromatogram of a preparative size exclusion chromatography run to purify untagged FrsH from the FT and W fractions. The collected fractions are displayed in italic and were analysed via SDS-PAGE in **B**.

Nucleation of FrsH was not observed in any of the conditions of the initial screens. Therefore, an additional screen was conducted. The buffer components of this screen based on the crystallization conditions of the NHDM CmlA from chloramphenicol biosynthesis (100 mM HEPES pH 7.5, 10-15% PEG 20000, 100 mM potassium acetate, 10% glycerol).^[59] Finally, after 2 weeks of incubation time at 4 °C, a square-bipyramidal FrsH crystal grew in 100 mM HEPES pH 7.9, 10.9% PEG 20000, 80 mM potassium acetate, and 1.36% glycerol (Figure 3.21). The protein crystal diffracted with a

RMDS of 2.7 Å in X-ray measurements. A model of FrsH was constructed by molecular replacement on basis of the reported structure of CmlA (PDB ID: 4JO0).

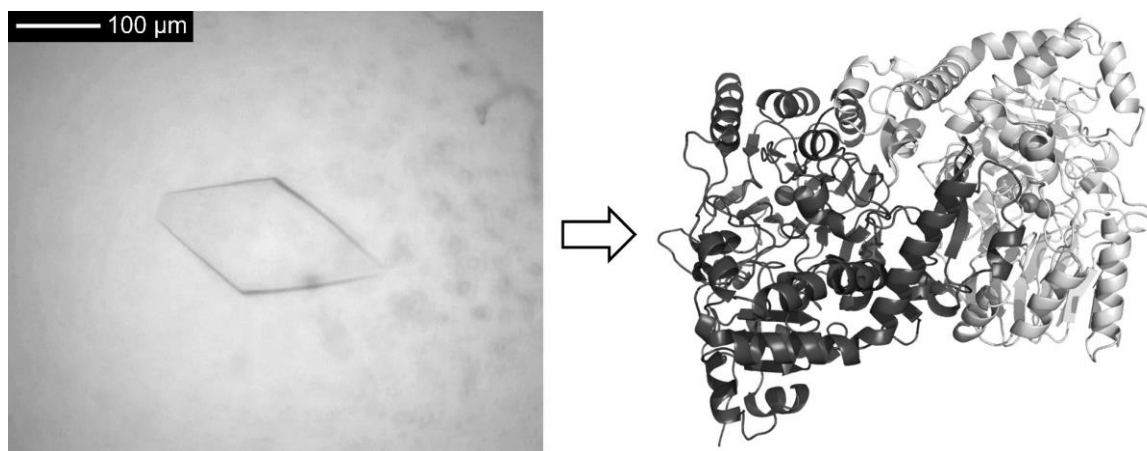


Figure 3.21: Crystallization of FrsH. **Left:** FrsH formed square-bipyramidal monocrystals (Picture provided by Niels Schneberger). **Right:** Crystal structure of FrsH, obtained by X-ray diffraction measurements of the crystal on the left. One structure is composed of two FrsH monomers (light and dark grey), each containing a dinuclear iron cluster (spheres).

3.4.6 Structure of FrsH

Despite their low sequence identity of 34 %, the overall structure of FrsH is nearly identical to the prototype NHDM CmlA. The monomeric units of FrsH and CmlA align with a RMDS of only 1.448 Å (Figure 3.22). The protein includes both a C- and a N-terminal domain. The N terminal domain of FrsH comprises a “L”-shaped structure of 60 x 36 Å, formed by rather distorted antiparallel β -sheets and α -helices. It is dominated by a long helix, which follows the general “L”-shape on the surface of the domain. A 75° kink in the helix is induced by Ala92 and Gly93, which disrupt the linear shape. The internal part of the domain is formed by two sets of antiparallel β -sheets and several α -helices. The larger part of the “L”-shape constitutes the dimerization arm of the protein, while the smaller part is packed against the C-terminal domain. A search for structural homologues of this domain in the protein data base with the DALI server (Protein structure comparison server),^[165] gave no significant matches, except for CmlA.

The C terminal domain is much larger (40 x 55 Å) and comprises an $\alpha\beta\alpha$ fold, which is typical for members of the MBL superfamily.^[55,166] A search on the DALI server yielded numerous Metal-dependent hydrolases, like the MBL H8A from *Thermotoga maritima* (PDB ID: 3X2Y) and L-ascorbate-6-phosphate lactonases, like the Mn²⁺-dependent lactonase UlaG from *E. coli* (PDB ID: 2WYL) as closest structurally related hits in the PDB. The core of the domain is formed by two large β -sheets, each consisting of seven strands, which are organized in parallel as well as antiparallel orientation. The β -sheets are flanked by three α -helices per side, whereas the helices, which interact with the N-terminal domain are poorly structured and truncated, compared to their

counterparts. A rather distorted loop extrudes from the ordered helices, it contributes to the formation of the substrate channel, interacts with the N-terminal domain, and is proposed to display a docking interface to the surface.^[59] Due to low sequence identities, members of the MBL superfamily are reliably classified by structural alignments. Nevertheless, sequence alignments identified a consensus sequence, which is primarily involved in the coordination of the dinuclear metal ion cluster, and displays a higher degree of conservation. It is defined as H-x-H-x-D-x/H-x₇₀-H-x₂₅-C/S-x₅₀-H.^[167] For FrsH, as well as all other members of the family of *trans* acting NHDH,^[54] the sequence is slightly rendered to H-x-H-x-D-H-x₇₆-E-x₂₅-D-x₂₆-E (see appendix Figure 8.2). The diiron cluster is located in the C-terminal domain at the internal tips of the β -strands, which is comparable to all other members of the MBL superfamily.^[55] Fe2 is coordinated by Asp311, His312, and an acetate residue, whereas Fe1 is coordinated by His307, His309, and Glu379. Furthermore, both iron atoms are μ -1,1-carboxylato and μ -hydroxo-bridged by Asp405 and hydroxo ion, respectively (Figure 3.23A), which is in line with the active site structure of CmlA in diferrous state.

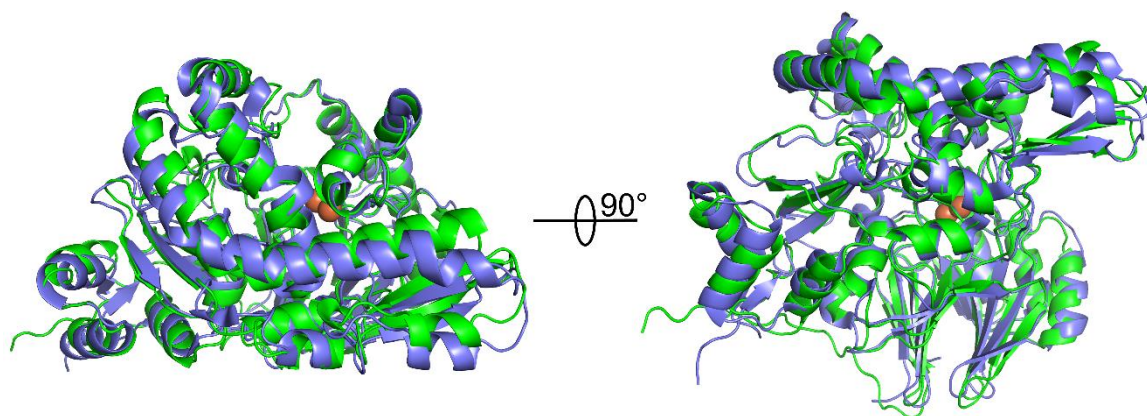


Figure 3.22: Structural alignment of monomeric CmlA (violet) and FrsH (green). The proteinogenic structures are displayed in cartoon style and the iron atoms are given as brown speres. The overall shape of the proteins, as well as the length and position of secondary structures is highly similar.

The binding pocket of FrsH has a size of roughly 10 x 9 Å and is connected to the surface by a 17 Å deep channel, which corresponds to the length of the phosphopantetheinyl moiety of the T domain that carries the substrate amino acid. The bottom of the pocket is defined by H262, Q310, W440, L441, L460, V494, and V495 (Figure 3.23B), which is nearly identical to the respective residues in CmlA (H258, Q310, W438, L439, L458, V494 and V495). For CmlA it was proposed that upon substrate binding, the side chain of Q308 adopts a new conformation to extend the size of the binding pocket and to coordinate the NH₂-group of L-PAPA.^[59] Interestingly, Q308 of CmlA is also present in FrsH (Q310). Upon binding of the hydrophobic substrate L-Leu, Q310 might get repulsed away from the active site, towards the hydrophilic residues Y26 or S314. The aliphatic part of Q309 might then be able to coordinate L-Leu in the enlarged binding pocket, together with

F181 and V494. If this hypothesis is true, the flexibility and amphiphilic character of Q309 would heavily reduce the substrate specificity of FrsH. In the native BGC, this does not affect the product formation because the accessible substrate of FrsH is tightly restricted by the specificity of the A domain in the respective NRPS module, which in contrast is specifically targeted by FrsH. On the other hand, the loose specificity of the *trans* acting enzyme towards the bound substrate facilitates the hydroxylation of diverse amino acids by just a single enzyme. This is exemplified by LybC in the lysobactin BGC, which, according to the bioinformatic hypothesis, hydroxylates the structurally diverse amino acids L-Leu, L-Tyr, and L-Ans.^[35]

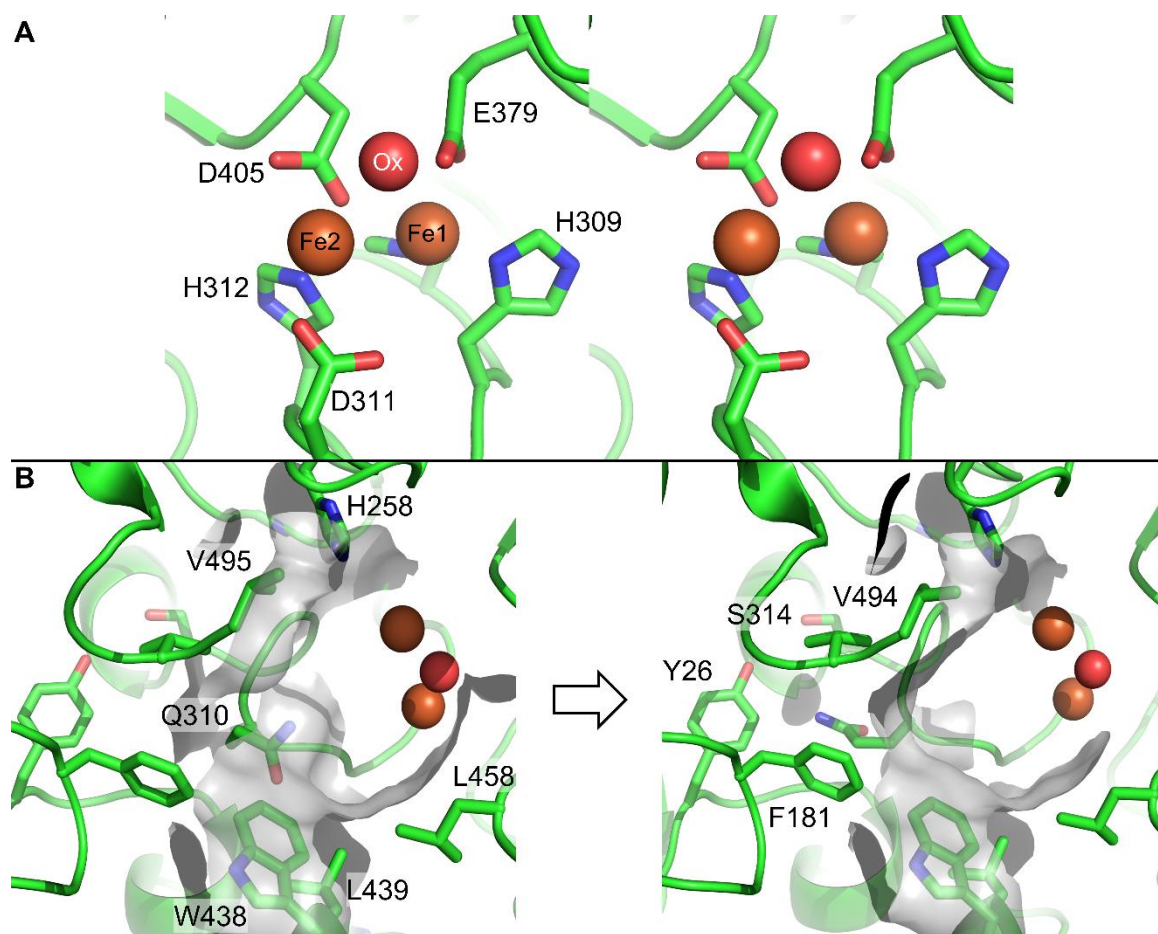


Figure 3.23: Cartoon model of the active site of FrsH. **A:** Stereoview of the diiron cluster (brown spheres). Side chains of coordinating residues are given in sticks and the oxo ion is depicted as red sphere. The Fe2-coordinating acetate residue, which originated from the crystallization buffer, was omitted for the sake of clarity. **B:** View through the substrate channel onto the diiron cluster. Binding pocket-defining residues are shown as sticks and their apparent surface illustrated as grey contour. Left: In the obtained crystal structure, the hydrophilic side chain of Q310 protrudes towards the diiron cluster. Right: Upon binding of substrate L-Leu, Q310 might get repulsed towards the hydrophilic residues S314 and Y26, resulting in a lipophilic environment in the binding pocket.

One crystallization unit of FrsH consists of two monomers, which interact tightly with each other. The entrances to the substrate channels of the monomers are located on a plane surface, which is formed by the interdigitation of the dimerization arms. Apart from the numerous contacts of the dimerization arms, the monomers hardly interact with each other, which leads to the formation of

a deep groove on the opposite side of the enzyme. The groove distinctly separates the C terminal domains of the monomers, which contribute only a single contact between K344 and D359 to the dimerization process (Figure 3.24).

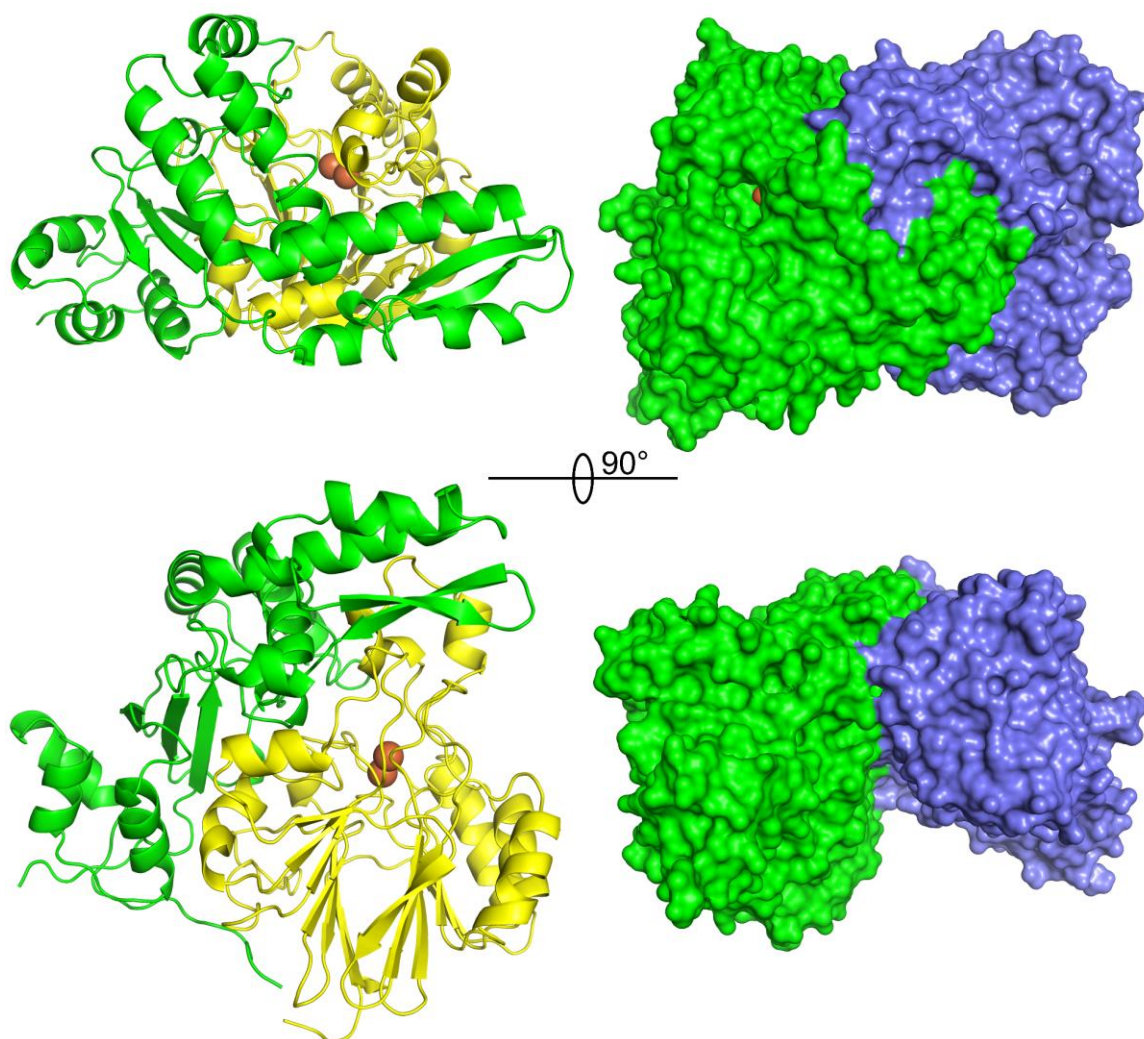


Figure 3.24: Structure of FrsH. **Left:** Monomeric structure of FrsH given as cartoon model. The structure of the N-terminal domain (green) is rather distorted and dominated by a long, kinked α -helix on the surface, while the C-terminal domain consists of a well-structured $\alpha\beta\alpha$ motif. **Right:** Dimeric FrsH rendered as solid structure, with differently coloured monomers. In the top view, the dimerization arm is clearly distinguishable. One iron atom of the active side (brown spheres) is visible through the substrate channel. The sideways view depicts the deep groove between the C-terminal domains of the monomers.

Taken together, the structure of FrsH resembles that of CmlA in all relevant sites, including the diiron cluster, the substrate binding pocket, the electron transport system, and the overall homodimeric structure. This was already postulated by sequence alignments, but could not be stated clearly due to the low sequence similarity, especially in the N-terminal domain. However, the elucidated structure of FrsH verifies the postulations and thus allows to draw conclusions from homology-generated models of other NHDM like HynC and LybC from the synthetases of hypeptin and lysobactin, respectively.

3.4.7 Dimeric character of FrsH

Similar to the prototype NHDM CmlA, one crystallization unit of FrsH consists of two monomers (chains A and B). The monomers are densely packed and bury an interface of roughly 2700 Å², which implicates a stable dimeric character of the enzyme in solution. This has also been shown for CmlA by native PAGE analysis.^[59] To verify the dimeric character of FrsH in solution, an analytical size-exclusion chromatography (SEC, see section 5.9.1) was performed. The retention volumes of proteins depend in a logarithmic manner on their apparent size. Hence, the retention volumes of a mixture of proteins with varying sizes was measured to set up a calibration curve. The mathematical function of this curve allowed to calculate the apparent size of FrsH from its retention volume to be roughly 80 kDa (Figure 3.25). This was bigger than the calculated size of a monomer (63.7 kDa), but also much lesser than that of a dimer (127.4 kDa). The correct sigmoidal calibration of the SEC column is feasible due to the globular shape of the proteins in the calibration mixture, which thus have a comparable hydrodynamic volume in solution. The obtained crystal structure of FrsH shows that the dimeric structure displays a handle-shape, while the monomer is rather formed like an “L”. In both cases, the structure is non-globular and thus impedes the comparability with the calibration standard, which might explain the altered retention volume.^[164] These data do not provide evidence to state the definitive dimeric or monomeric character of FrsH in solution. Further experiments, like blue native page^[168] or dynamic light scattering^[169] would allow a more precise estimation of the protein’s size, but were not performed due to the proven dimeric character of the homologue CmlA and the subordinate relevance of this matter for the overall project.

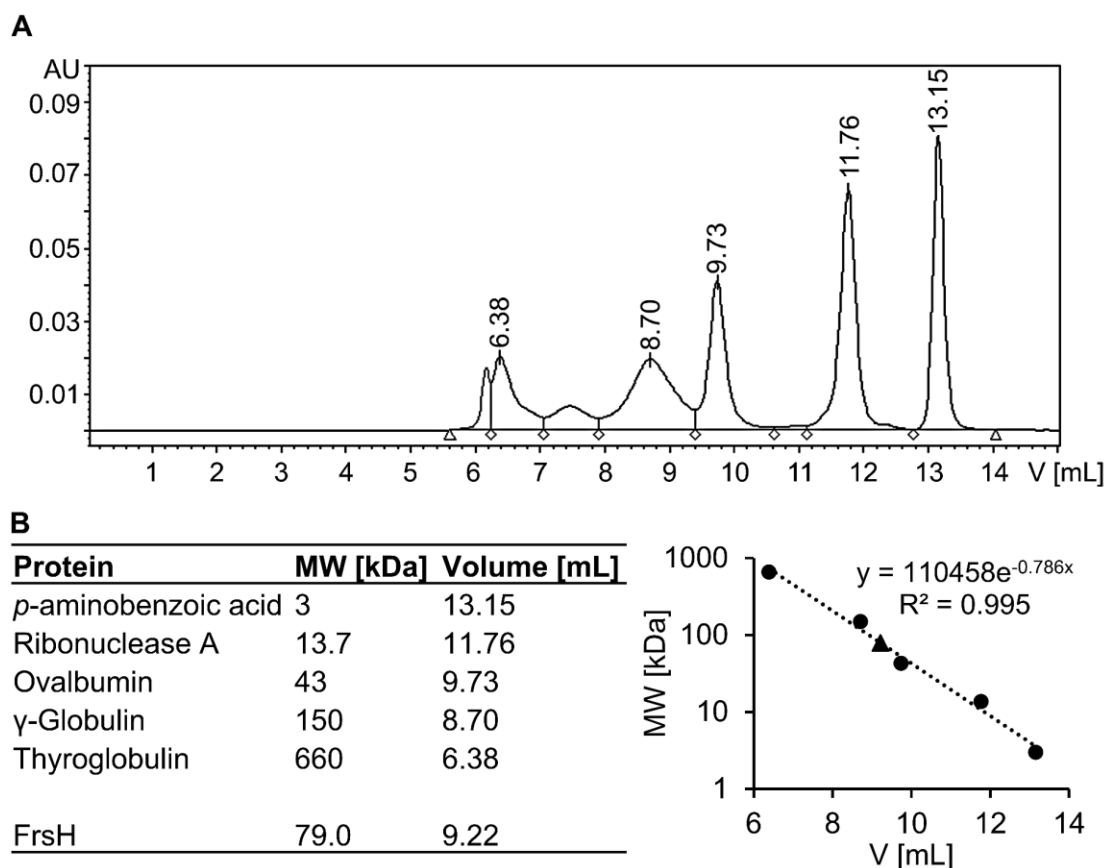


Figure 3.25: Investigation of the dimeric character of FrsH in solution. Size-exclusion experiments were conducted on a TOSOH TSKgel G3000SW_{XL}, coupled to a common HPLC device. Proteins were detected at a wavelength of $\lambda=280$ nm **A** Chromatogram of the calibration mixture. Retention volumes of the single proteins are indicated. **B** The known molecular weights (MW) of the calibration mixture and the measured retention volumes are given in the table. These values were used to set up a calibration curve (globes), which facilitates the calculation of the apparent molecular weight of FrsH (triangle) as a function of its retention volume.

In silico docking of FrsH and FrsA

The *trans* acting NHDM FrsH directly interacts with the substrate T domain, to hydroxylate the bound substrate L-Leu. In the previous chapters, bioinformatic, *in vivo*, *in vitro*, and structural analyses of the involved enzymes were applied to gather insight into the interaction and its driving forces. Together with published data,^[59,60] these results substantiate the hypothesis that the NHDM does not solely interact with the T domain, but also with surrounding parts of the NRPS module, probably the A domain. In this chapter, the interaction of FrsH with the A domain of FrsA was computed *in silico* to unveil essential residues for the interaction of both enzymes. The results provide a basis for subsequent experiments *in vitro*, which aim to functionally disrupt the postulated interface.

Generally, most algorithms for *in silico* protein-protein docking apply a similar approach. If no experimental data are provided to narrow the possible binding site, the proteins are handled as rigid structures. In a first step, the ligand protein is docked in a screening approach in all possible

orientations to each surface patch of the receptor protein. If a specific orientation is energetically favoured, moderate flexibility of the protein backbone and the side chains at the putative attachment site is allowed to further optimize the energetic value of the interaction.^[170] Due to the low probability that the calculated best hit actually displays the correct binding site, the top ten hits are indicated and have to be manually analysed for plausibility.

A domains in NRPS are commonly organized into two subdomains. The larger, N-terminal A_{main} (approximately 50 kDa) holds large parts of the active site and of the substrate binding pocket, which are complemented by the much smaller (~ 10 kDa) A_{sub} .^[8] A plethora of multidomain crystal structures gave deep insight into the interplay of both subdomains with their surroundings. While A_{main} is rigidly packed against the upstream domain (mostly C domain) and a possible MLP, the linker regions, which flank A_{sub} are highly flexible and account for substantial variability in the architecture of NRPS.^[17,18,29] This flexibility severely hinders the detection of protein-protein interfaces *in silico*, as the computational power for *in silico*-docking of proteins is insufficient to consider the high degrees of freedom in the backbone of three-dimensional structures.^[171]

In the case of FrsH interacting with FrsA, several rules were posed to level the impact of the flexibility on the docking computations and to reduce the surface, which is free for binding:

- (1) In all structures of A domains that do not adopt the conformation of thiolation state, A_{sub} does not contribute strongly to any interaction. Instead, it serves as a hinge to allow mobility of the T domain. Thus, it probably does not significantly contribute to the interaction of FrsA₁A and FrsH and will be disclosed in the docking approaches.
- (2) The T domain interacts with FrsH with the conserved serine located directly above its substrate channel.
- (3) The T domain is bound to A_{sub} , which allows a maximum distance to A_{main} of roughly 54 Å, corresponding to the spanned distance of LgrA₁ A_{sub} in condensation state (PDB ID: 6MFZ).^[29]
- (4) The upstream C domain interacts with the A domain. It occupies larger parts of the surface and does not significantly contribute to the interaction.
- (5) The MLP FrsB interacts with the A domain and does not significantly contribute to the interaction.

A structural model of FrsA₁ A_{main} was generated with iTasser,^[139] based on an homology alignment with the tridomain NRPS EntF from enterobactin BGC (PDB ID: 5T3D) (for complete structure, see Appendix Figure 8.3).^[172] The iTASSER server was used for this approach, as it computes highly accurate three-dimensional structures of proteins from the amino acid sequence, by combining template-based models from the protein data bank with a protein function database. In a next step, the complex of FrsA₁ A_{main} bound to FrsB was mimicked by superimposing FrsA₁ A_{main}

with the crystal structure of EntF, bound to its native MLB YbdZ (PDB ID: 5JA1) in PyMOL2. The protein chains of FrsA_{1A_{main}} and YbdZ were then merged into a single structure. The structure of the upstream C domain was not included to keep the necessary resources of the docking server on a reasonable level.

The generated structures of FrsA_{1A_{main}}/YbdZ (ligand) and FrsH (receptor) were docked to each other with the SwarmDock server,^[170] allowing free binding of the molecules. The SwarmDock server was chosen for a primary run, because it utilizes the best algorithm to model the interaction of two protein, according to the latest Critical Assessment of Predicted Interaction (CAPRI).^[173] One advantage of the server is the “democratic” clustering, which ranks the docking results in a specific voting scheme. In addition, the server automatically precomputes the uploaded structures, which spares the user energy-minimizing prior to uploading. Despite the superior performance in the CAPRI ranking, compared to other docking servers, the results of SwarmDock were profoundly analysed for their plausibility. In a performance test with a benchmark set consisting of 51 heterodimers, SwarmDock yielded an at least acceptable prediction in the top ten models for only 41.3% of the structures. The reliability of the predictions improved significantly when supplying the algorithm with experimental information about the interaction,^[173] but distinct information about the interface are scarce for the interaction of FrsH with FrsA_{1A_{main}}.

The top ten results according to “democratic” clustering were downloaded and analysed with PyMOL2. Only one of the models (109a, ranked No. 6) was excluded from further analyses, because it locates FrsA_{1A_{main}} directly on the substrate channel of FrsH, thereby violating rule No. (2). To generate a larger database, a homology model of LybC was generated with SwissModel,^[172] based on the structure of CmlA. The resulting structure was also uploaded to the SwarmDock server as a receptor and docked to FrsA_{1A_{main}} (ligand). Here, five of the top ten results (31c, 111c, 115d, 15b, 42a; rank No. 2, 4, 8, 9, 10) were excluded from further analyses due to violation of the above-mentioned rules. The remaining 14 models are discussed in the following two sections.

3.4.8 Putative docking interface of FrsH

The docking results of FrsA_{1A_{main}}/YbdZ to FrsH and LybC were aligned and superimposed to the initial model of FrsH, to get an overview of probable docking sites on the NHDMs. Their dimeric character hampered the interpretation of the resulting alignment, due to the duplication of possible binding sites. On a first glance, it seemed that FrsA_{1A_{main}} is docked to three different positions of the NHDM, which are reduced to two after considering the dimeric structure of the hydroxylase. In four of the 14 models, the binding interface is postulated to be comprised of the surface of both dimerization arms and located on the plane area between the entrances of the substrate channels. In this case, the homodimeric enzyme would display a functional monomer, as it could bind a single NRPS module at once. Nine of the remaining models proposed the binding site to be located on the C terminal domain of a monomeric unit. The putative interface is mainly composed of the superficial strands of the two β -sheets, together with the proximal α -helix, of the $\alpha\beta\alpha$ -motive (Figure 3.26 C).

The results of the *in vivo* substitution of FrsH by the homologues HynC and LybC (see section 3.4.4) indicate a similar mechanism of interaction of the enzymes. Hence, the interface for the interaction with FrsA_{1A_{main}} has to be at least partially conserved and might be detectable by comparing the surface composition of the enzymes. For this approach, SwissModel-generated homology models of HynC and LybC, together with the obtained structure of FrsH were loaded in PyMOL2. The electrostatic properties of each enzyme's surface were visualized with the "APBS Electrostatics" algorithm (Figure 3.26 A). By superimposition of the enzymes, single amino acids and patches with similar electrostatic composition could be manually detected. Due to the low sequence identity of the proteins, major parts of the overall surface differ heavily between the enzymes. Nevertheless, apart from the T domains binding site at the entrance of the substrate channels, four surface patches seem to be functionally conserved. First, within the dorsal groove, which is formed between the two C terminal domains of the monomers, several lipophilic amino acids with low electrostatic potential form a patch of 150 Å². Due to the location in the narrow groove on the opposed side of the substrate channels, a contribution of this interface to the association with FrsA can be ruled out. A second functionally conserved location is the outer edge of the N terminal domain, which displays a net positive charge in all three investigated enzymes. It was already proposed for CmlA that this charge might be responsible for the interaction with the cognate NRPS module or with an electron transport system, which is necessary to regenerate the active site after hydroxylation occurred.^[59] The two further functionally conserved sites actually coincide with the hot spots of the docking results: The surface between the entrances of the substrate channels is dominated by net negative charges, except for two protruding tips, originating from a distorted α -helix of the C terminal domain that display a positive charge on the surface. Lastly, the

superficial β -strands of the $\alpha\beta\alpha$ motive in the C terminal domain display an uncharged patch of varying size, formed by bulky, lipophilic amino acids like Leu and Met (Figure 3.26 B).

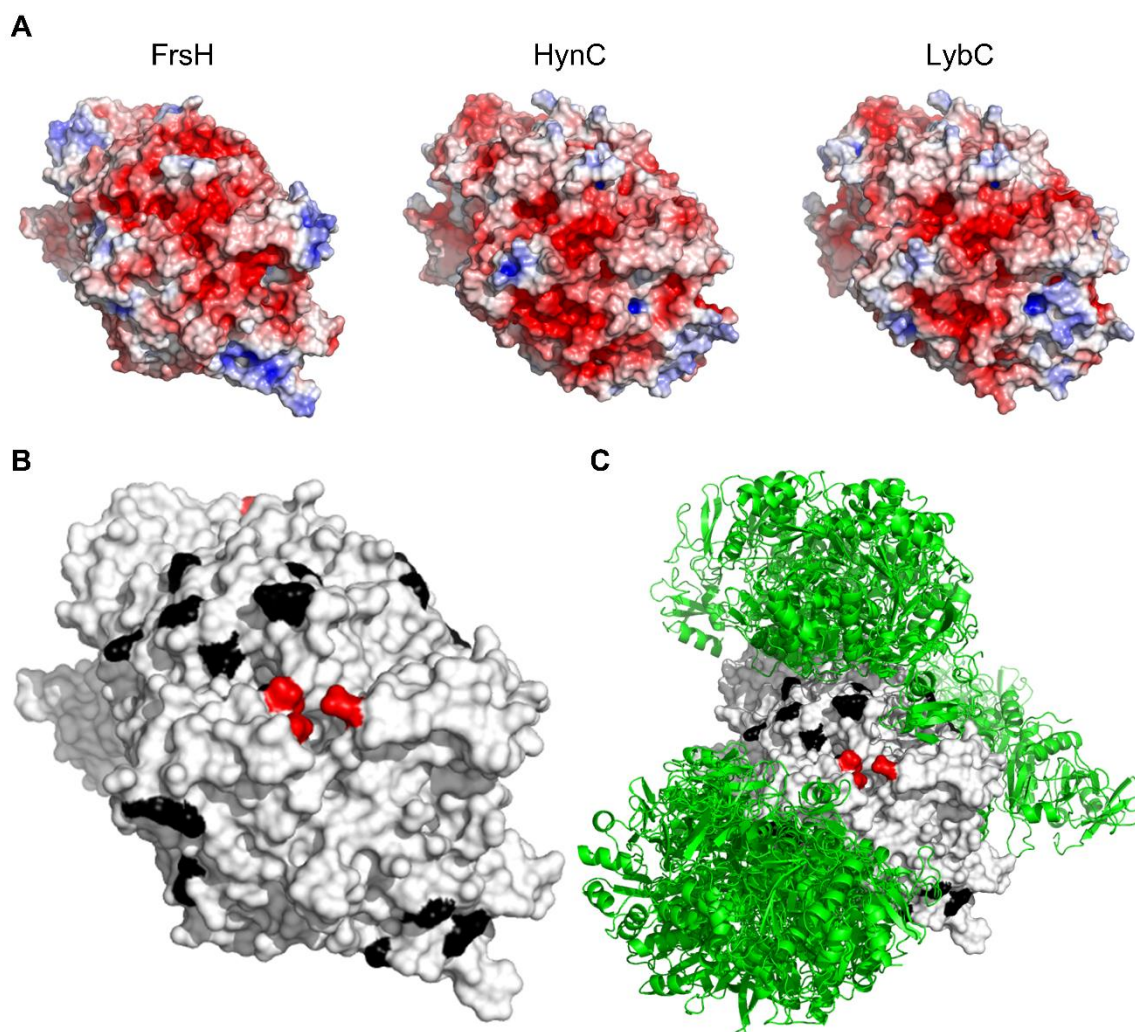


Figure 3.26: Structural analysis of NHDMs to determine possible interfaces to bind FrsA. **A** Structures of the NHDMs FrsH, HynC, and LybC, all in identical orientation with colour-coded electrostatic character of the surface (red: negative; blue: positive; grey: neutral). Direct comparison of the structures unveils functionally conserved surface residues. **B** Solid structure of FrsH visualizes the result of the analysis in A. Functionally conserved surface residues that do not belong to the substrate channel are marked in black. For better orientation, the entrance of the substrate channel is highlighted in red. **C** Resulting structures of the computational docking of FrsA_{1Amain} (cartoon style, green) to FrsH and LybC, aligned and superimposed to the structure shown in B. In almost all cases, the proposed binding interface covered a functionally conserved residue, whereas some functionally conserved patches were not included at all, indicating docking sites for other binding partners.

To further corroborate the predictions of SwarmDock, the computations were repeated on the ClusPro2.0 server,^[174] with FrsH as receptor and FrsA_{1Amain} as ligand. According to the latest CAPRI ranking, ClusPro predicts protein-protein interactions just as well as SwarmDock, when considering the top ten results and not just the very first one.^[173] The general workflow of this algorithm is similar to the one of SwarmDock, but it allows for adjustment of more parameters for the computing. Additionally, the resulting models are given in different ranks, depending on the

favoured kind of interaction (electrostatic, hydrophobic, VdW, or balanced), which is useful for the evaluation of specific classes of interaction. In the present case of a common protein complex, no specific residues should be favoured.^[175] Thus, the “balanced” results were considered for further analysis. In general, the docking results of ClusPro resemble these of SwarmDock (Figure 3.27). Five of the models predict the A domain to bind between the entrances of the substrate channels, yet the structures are spread wider about the surface. Two models display the A domain to bind to the superficial strands of the β -sheets of the C terminal domain, which is the favoured docking site according to SwarmDock. In the final three models, FrsA₁A_{main} is docked deep inside the distal groove between the C terminal domains of the monomers, which is no plausible docking site due to the abovementioned reasons.

Taken together, the *in silico* analyses, which were based on the structure of FrsH, yielded promising candidate interfaces for the association to FrsA₁A_{main}. In particular, the site constituted by the β -strands of the C-terminal domain is functionally conserved and was favoured by the docking algorithms. Thus, it was further examined by *in vitro* experiments (section 3.4.10).

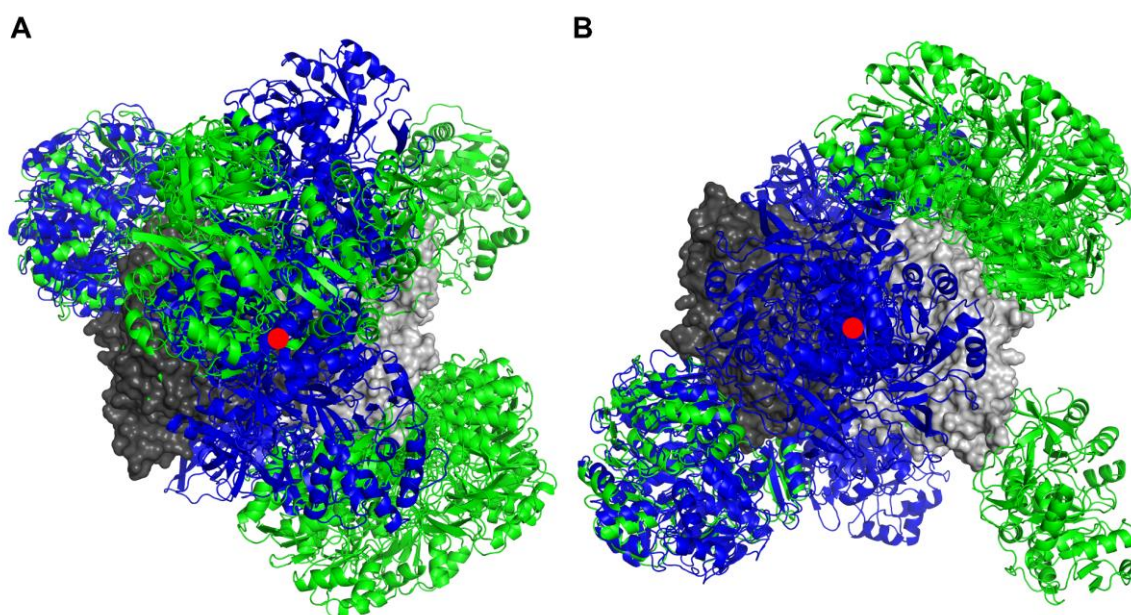


Figure 3.27: Comparison of the docking results from SwarmDock and ClusPro server. All results of the computational docking experiments of FrsA₁A_{main} to FrsH were aligned and superimposed to the structure of FrsH (solid, monomers given in light and dark grey). The structures of FrsA₁A_{main} (cartoon style) are given in green for SwarmDock results and in blue for ClusPro results. **A** View of the upper side of FrsH, the entrances of the substrate channels point towards the viewer. **B** Bottom view of FrsH. The red dot indicates the symmetry axis of the dimeric FrsH. Structures, which dock to one site of FrsH, can equally dock on the opposed site of the axis. Except for the structures on the bottom side of FrsH, both docking servers predict similar possible binding sites on FrsH.

3.4.9 Putative docking interface of FrsA₁A_{main}

To distinguish surface patches on FrsA₁A_{main}/YbdZ that were preferred for the binding to FrsH and LybC by the SwarmDock algorithm, the results of the docking runs were aligned and superimposed to the primary structure of FrsA₁A_{main}/YbdZ. Strikingly, almost all (twelve of 14) models predict the binding interface next to the MLP and the catalytic site of the A domain (Figure 3.28 **A**). The structures of FrsH are distributed over a surface area of roughly 2600 Å² (without the MLP), whereas the single structures occupy 870 Å² to 1300 Å², depending on the orientation of FrsH in the respective structure. In four of these models, the MLP contributes larger, yet not major, parts to interface. In the remaining two of the 14 structures, the docking site is modelled to the opposite side of FrsA₁A_{main}. Both structures are nearly identical and propose an interface of 860 Å².

Each of the models was manually reviewed to identify residues of FrsA₁A_{main}, which are essential in the specific docking position. Figure 3.28 **C** and **D** exemplify one of these examinations and Figure 3.28 **B** visualizes the overall result of the investigation. In total nine residues (S502, D508, V509, M555, W573, D585, R597, Q613, and F635) were identified as to playing a major role in the docking of at least one of the models. Their actual relevance for the interaction of FrsA and FrsH was further assessed *in vitro*.

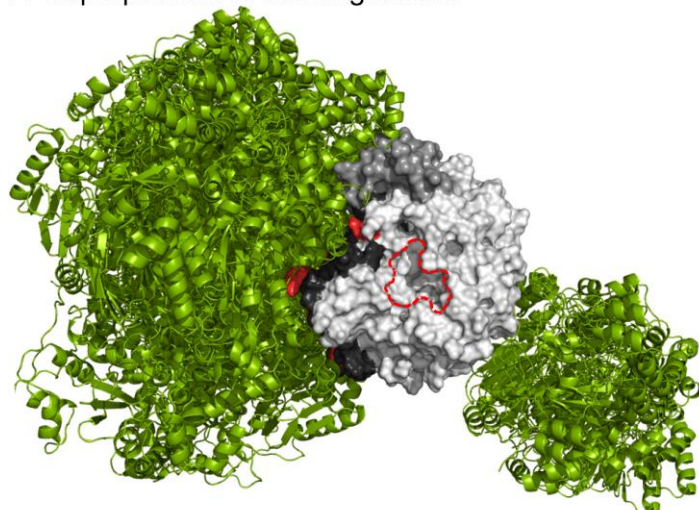
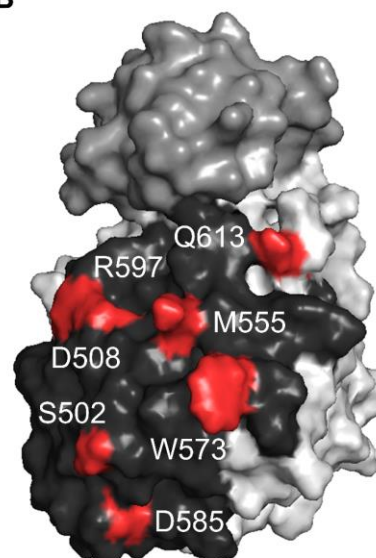
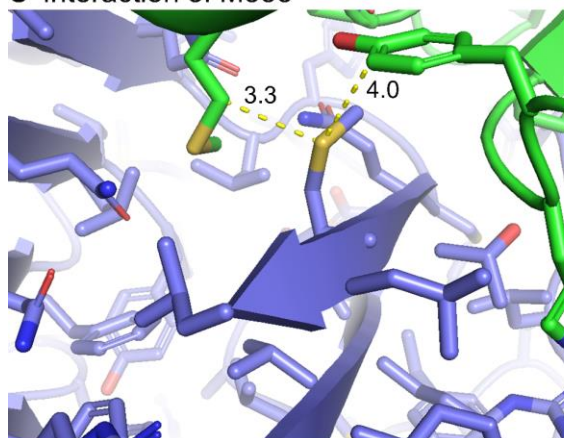
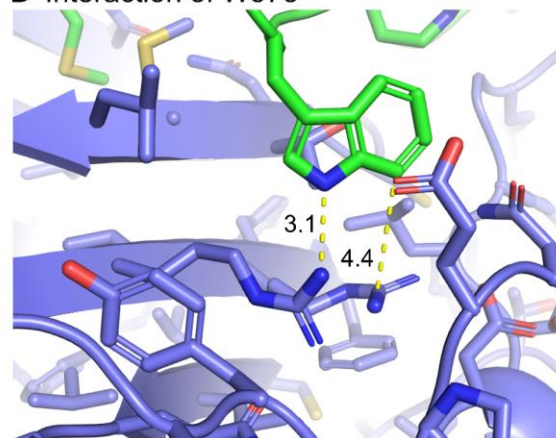
A Superposition of docking results**B****C** Interaction of M555**D** Interaction of W573

Figure 3.28: Binding interface of FrsA_{1A}main, predicted by SwarmDock. The structures of the NHDMs FrsH and LybC were docked to FrsA_{1A}main *in silico* to detect putative binding interfaces. **A** After exclusion of redundant results, the remaining models were superimposed to the structure of FrsA_{1A}main (solid, light grey), bound to the MLP (solid, dark grey). The NHDMs (cartoon, green) predominantly bind on a surface patch next to the MLP and the active site of the A domain (red dashed line). **B** Enlarged model of FrsA_{1A}main rotated by 90°. Surface residues that are part of the binding interface to a NHDm in at least one model are coloured black. Residues that particularly contribute to the interaction in the docking models are highlighted in red. **C** Interaction detail of FrsA_{1A}main (green) and FrsH (blue) that was predicted in several models. Protein backbones are depicted in cartoon style with amino acids side chains as sticks. FrsA M555 protrudes into a small lipophilic groove between the superficial β -strands of FrsH and is in close proximity to FrsH M299, which also forms a long-ranging S-aromatic interaction with FrsA Y572. Distance measurements between single atoms are indicated as yellow dashed lines and given in Å. **D** Interaction detail of FrsA W573 that protrudes from a backbone loop into a small hollow of FrsA to interact with FrsH R253 and FrsH R255 in a cation-pi interaction. Flexibility in the backbone loop of FrsA_{1A}main might allow a rather perpendicular orientation of the arginines to the plane aromatic atoms to further strengthen the interaction.

3.4.10 *In vitro* assay to verify the interaction of FrsH and FrsA₁A_{main}

So far, the interaction of FrsH with FrsA was not explicitly verified. It is plausible that during hydroxylation of the substrate L-Leu, FrsH has to interact with the T domain, which carries the to be hydroxylated amino acid. However, the interaction with the A domain was proposed on basis of bioinformatical analyses and general causalities. In this chapter, a method to quantify the association of FrsA₁CAT and FrsH *in vitro* was established to measure the impact of single mutations on the protein-protein interaction.

The side chain assembly assay, which is described in section 3.3.4, yields (2*S*,3*R*)-(N-propionyl)-3-hydroxyleucine (*N*-Pp-Hle) as the main product. Interestingly, the non-hydroxylated product (2*S*)-(N-propionyl)-leucine (*N*-Pp-Leu) was also always detectable in trace amounts. The proportion of hydroxylated to non-hydroxylated product is controlled by the gating mechanism of the upstream C domain. It prefers hydroxylated amino acids in its acceptor-site binding pocket to slow down the peptide assembly until hydroxylation occurred, thereby reducing the formation of shunt products *in vivo*.^[5] On the other hand, FrsH cannot hydroxylate the product of the C domain at all because the binding pocket, visualized in the crystal structure, harbours no apparent space for the acyl chain. Additionally, the primary amine is crucial for oxygen regulation upon substrate binding.^[58] Thus, any restraints in the hydroxylation process reduce the production of *N*-Pp-Hle and thereby enhance the formation of *N*-Pp-Leu. Disruptions in the interface of FrsH and FrsA₁A_{main} should reduce their affinity towards each other, resulting in a slower convergence and less frequent hydroxylation. Consequently, the percentage of *N*-Pp-Hle on the overall product formation may be used to quantify the effect of mutations on the interaction.

In a first step, the side chain assembly assay was conducted to determine the ratio of hydroxylated to non-hydroxylated product formation. Thus, FrsH and FrsA₁CAT/FrsB were heterologously expressed, purified and incubated with all necessary substrates as described before (see chapter 3.3.4). Products were detected after liberation by alkaline thioester cleavage via LC-MS as single ion record (SIR) at $m/z = 202.1$ for *N*-Pp-Hle and at $m/z = 186.1$ for *N*-Pp-Leu in negative mode. The chromatograms were analysed with the automated peak detection of the software (Empower® 3) and the determined AUC was used to calculate the amount of hydroxylated product as a ratio of the overall product formation with the following formula:

$$[\%] \text{ hydroxylated product} = \frac{AUC_{N-Pp-Hle}}{AUC_{N-Pp-Hle} + AUC_{N-Pp-Leu}} * 100$$

Remarkably, the amount of *N*-Pp-Leu accounted for approximately 10% of the overall product formation, which is in line with the results of Kaniusaite *et al* on the teicoplanin biosynthesis. In their set up, a surprisingly high amount of shunt product was accumulated, which blocked the assembly line and thus further reduced the formation of the main product (Figure 3.29 A).^[5]

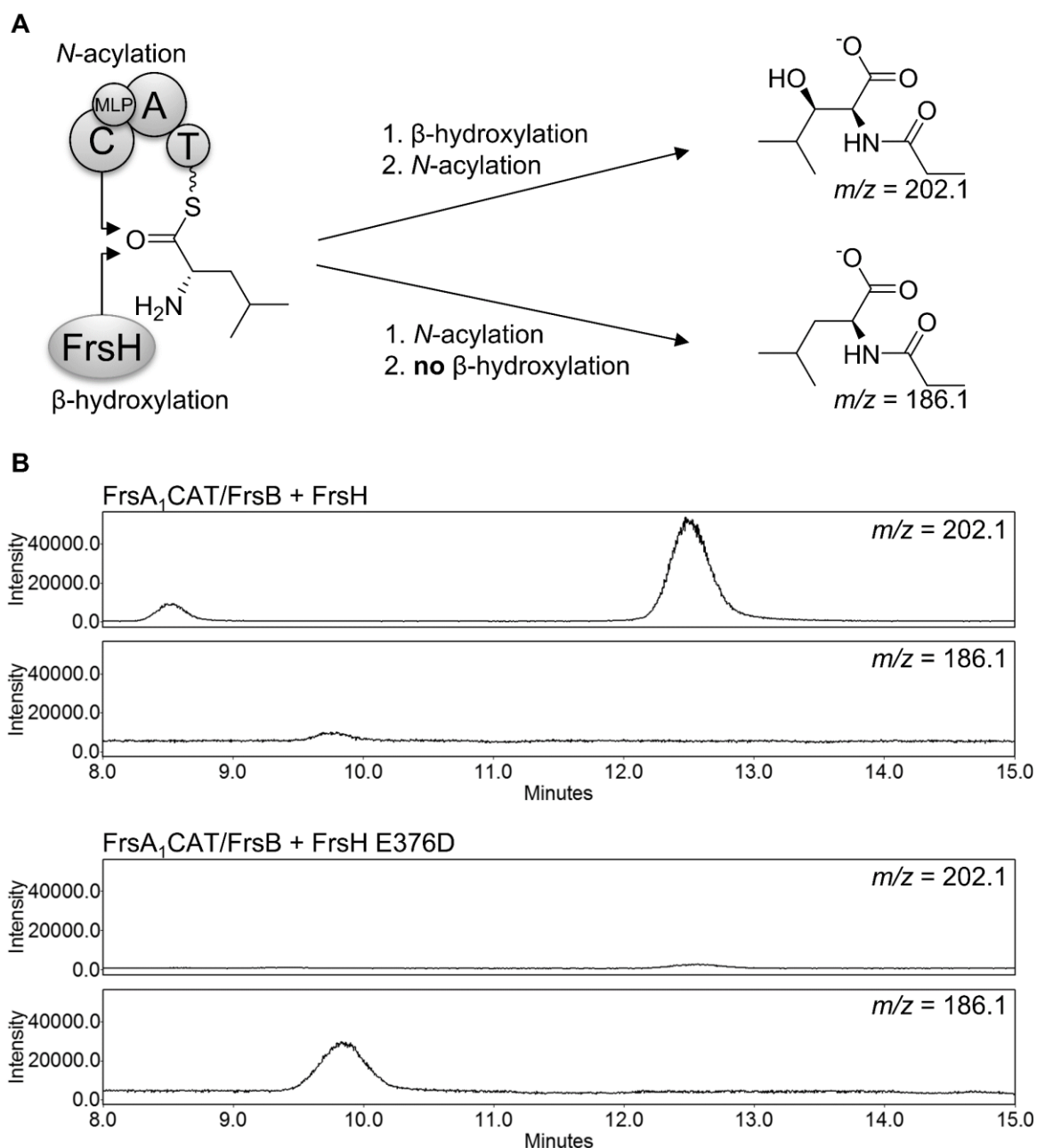


Figure 3.29: Formation of hydroxylated vs. non-hydroxylated product in the side chain assembly assay. A Visualization of the reaction paths, starting from T domain-bound L-Leu. After hydroxylation by FrsH, the C domain conducts *N*-acylation, leading to the formation of *N*-Pp-Hle ($m/z = 202.1$). If *N*-acylation occurs first, no subsequent hydroxylation is possible, resulting in *N*-Pp-Leu ($m/z = 186.1$). **B** Representative SIR chromatograms of the extracted side chain assembly assays, at the respective m/z to quantify *N*-Pp-Hle and *N*-Pp-Leu. The assay of the upper chromatograms was conducted with the native enzymes to serve as positive control. For the negative control, the inactive mutant FrsH E376D was employed (lower chromatograms). Without hydroxylating activity, *N*-Pp-Hle production abolishes almost completely, whereas the amount of the shunt-product *N*-Pp-Leu is clearly enhanced.

To test if the ratio of hydroxylated to non-hydroxylated product is a convenient approach to evaluate the association of FrsH and FrsA₁CAT, FrsH was substituted with the inactive mutant FrsH E376D, which was already established in chapter 3.4.2. In this set-up, it was used to mimic the complete depletion of the interaction between FrsH and FrsA₁CAT. Hence, the ratio of *N*-Pp-Hle to the overall product formation should converge against zero, and the absolute formation of non-

hydroxylated *N*-Pp-Leu should increase by several magnitudes. Indeed, the AUC of *N*-Pp-Leu is increased by ten-fold, whereas the amount of *N*-Pp-Hle is merely detectable and makes up roughly 3 % of the overall product formation (Figure 3.29 B, Table 3.8). These results demonstrate the successful implementation of a workflow to assess and compare the influence of mutations on the hydroxylation process during side chain assembly.

In the next step, single residues of the proposed interface of FrsH and FrsA₁A_{main} (see sections 3.4.8 and 3.4.9) were chosen to be mutated. If the hydroxylation rate is decelerated by a mutation, this would count as a proof for the proposed interaction within the hydroxylation process. In a first approach, the residues were substituted with amino acids with opposed characteristics to induce a strong effect on the interface. The mutants FrsA₁CAT M555R, FrsA₁CAT V509D, FrsH R515L, and FrsH L531R were created with three-fragment Gibson assembly of the respective expression plasmid. Heterologous expression of the mutated proteins and subsequent SDS-PAGE analysis gave no evidence for apparent alteration of the expression yield or solubility (see appendix Figure 8.4). Interestingly, the mutants of FrsA₁CAT were extracted without visible impurities, in contrast to the unaltered construct. Subsequently, the purified enzymes were tested for their hydroxylating ability with the established method. Table 3.8 displays the results of the assays, all chromatograms are depicted in appendix section 8.1.4. Unfortunately, neither the single mutants, nor a combination of mutated enzymes led to a significant decrease of hydroxylating activity.

The unaltered hydroxylating activity of the mutated enzymes might have different reasons. The proposed hypothesis that the hydroxylation of T domain-bound L-Leu by FrsH depends on an interaction with the adjacent A domain is based on bioinformatic analysis and causal conclusions which arise from published studies. The emerging data gave good hints, yet no evidence for the interaction of the enzymes. Hence, the mechanism for the recruitment of FrsH to the NRPS module might still depend exclusively on the structure of the T domain, similar to P450_{sky} in skyllyamycin BGC.^[42]

The established method is essentially based on the differing reaction kinetics of the C domain and FrsH. In the standard reaction, where the tethered amino acid is given as substrate, the reaction velocity (V_{\max}) of FrsHs hydroxylation is probably several magnitudes higher than this of FrsA₁C *N*-acylation, which would explain, why hydroxylated product is primarily formed. The reaction rates of C domains have never been quantified due to their dependency on covalently bound substrate, allowing only single-turnover experiments. Nevertheless, *in vitro* assays with the gatekeeping C domain of teicoplanin biosynthesis yielded only trace amounts of product, with the non-hydroxylated amino acid as substrate after 25 min, whereas almost quantitative turnover was measured after five min when the hydroxylated amino acid was supplemented.^[5]

Table 3.8: Ratio of hydroxylated product on the overall product formation in the side chain assembly assay, performed with mutants of FrsA₁CAT and FrsH to detect the influence of single mutations on the interaction of the enzymes. The assay with FrsH E376D serves as negative control.

FrsA ₁ CAT + FrsB	FrsH	AUC		Mean <i>N</i> -Pp-Hle formation [%] (SD)
		<i>N</i> -Pp-Hle	<i>N</i> -Pp-Leu	
native	native	221583	18117	92.0 (4.1)
		740892	37644	
		832852	41163	
		181948	31842	
native	E376D	45417	643898	3.3 (2.3)
		5865	347677	
		6947	385030	
V509D	native	235554	115240	83.8 (11.8)
		1106138	87641	
		5299799	480228	
M555R	native	447216	208146	82.2 (9.9)
		6215018	822243	
		2963004	323236	
native	R515L	929369	322143	88.7 (10.2)
		3390437	147174	
		2172651	89960	
native	L531R	91553	33458	88.3 (10.6)
		2547461	122490	
		2302315	92066	
M555R	L531R	2194741	276359	88.2 (0.62)
		658637	93516	

These results indicate that the velocity of the hydroxylation reaction has to be greatly diminished to significantly alter the product ratio. The approach to compare the ratio of hydroxylated with non-hydroxylated product might thus be too insensitive to detect the influence of single mutations on the interaction. The predicted interface has a size of at least 870 Å², which can hardly be disrupted by a single mutation except if the mutated residue was accountable for a major, noncovalent bond. With this rationale, the mutants were designed to not just lose the function of the replaced amino acid, but to oppose it. For example, FrsA₁CAT M555 was predicted to protrude into a small lipophilic groove of FrsH in several models. In a common approach, it would have been mutated to FrsA₁CAT M555A, reducing the overall affinity. Nevertheless, the aliphatic side chain of alanine might still allow its coordination above the lipophilic residues, whereas the introduced arginine

(FrsA₁CAT M555R) would repel the converging FrsH. On the other hand, the introduced arginine is much more prone to interact in a way that could not be anticipated with the results of the docking algorithm. In particular arginine might display an amphiphilic character due to the methylene groups, whereas the guanidine functionality is detracted by hydrogen bonding, salt bridges or cation-pi interactions with surrounding residues.^[176]

Consequently, for the next attempt, the strategy was changed to a screening approach, where more potentially relevant amino acids of the interface between FrsH and FrsA₁CAT were tested for their contribution to the interaction. Due to the ambivalent character of the substitution with functional residues, neutral alanine mutations were chosen for the screening. Moreover, the mutations were solely introduced into the A domain's interface, while FrsH was not rendered at all. The ratio of hydroxylated product formation is not influenced by the actual activity of the A domain because hydroxylation occurs post adenylation and thiolation. Thus, if an introduced mutation diminishes the A domain's turnover rate, the absolute product formation would also diminish. The ratio of hydroxylated product formation, however, is only influenced by the mutation's impact on the protein-protein interaction with FrsH. Thus, the assay results are not influenced by the activity of the A domain. On the other hand, if a certain mutation of FrsH reduces the ratio of hydroxylated product, it cannot be distinguished whether this effect arises from the disruption of the interface, or a general depletion of the enzyme's activity. The following constructs were chosen to be created and tested: FrsA₁CAT S502R, FrsA₁CAT D508A, FrsA₁CAT W573A, FrsA₁CAT D585A, FrsA₁CAT R597A, FrsA₁CAT Q613A, and FrsA₁CAT F635A. The latter one was located on the opposite site of the A domain, compared to the other ones, to take account of the two lone standing docking models.

4 Summary and Conclusion

4.1 Hypeptin

The potent lipid II binding antibiotic depsipeptide hypeptin was recently rediscovered in a *Lysobacter* strain. The aim of this project was to locate and investigate the putative hypeptin biosynthetic gene cluster (BGC) *hyn* with bioinformatic tools and to biochemically verify the proposed functions *in vitro*. Hypeptin was reported to contain the four β -hydroxylated amino acids (2*S*,3*R*)-3-hydroxyasparagine, (2*R*,3*R*)-3-hydroxyasparagine, (2*S*,3*S*)-3-hydroxytyrosine, and (2*S*,3*R*)-3-hydroxyisoleucine. The diverse structures and stereoconfiguration of these side chains posed questions about their biosynthesis, thus the primary aim of this project was to unveil their origin.

The primary analysis of the genome with AntiSMASH yielded a non-ribosomal peptide synthetase (NRPS) BGC, its predicted domain architecture perfectly fitted the structure of hypeptin by strictly following the collinearity rule (section 3.1.3). Apart from the octamodular NRPS the BGC also encoded for the *trans* acting non-heme diiron monooxygenase (NHDM) HynC and the *trans* acting iron- and α -ketoglutarate dependent hydroxylase (Fe/ α KG) HynE. Publications about the respective enzyme families in NRPS report that the hydroxylases only recognize their amino acid substrate once it is bound to the T domain of the respective NRPS module. Hence, for the investigation of the hydroxylases' activity, the NRPS modules had to be reconstituted *in vitro* first.

The respective modules four, five, six, and seven of the *hyn* BGC were cloned and heterologously expressed, each with varying domain architecture (section 3.1.4). To assess their functionality, each construct was tested for the adenylation activity of the embedded A domain with the $\gamma^{18}\text{O}_4$ -ATP exchange assay (section 3.1.5). The results clearly showed that larger constructs, which also contain at least parts of the adjacent C domains, are more reliable for *in vitro* reconstitution than smaller ones. Interestingly, the modules five and six also depended on the presence of the MbtH-like protein (MLP) HynMLP, which was not encoded within the *hyn* BGC. All investigated modules were shown to primarily activate the respective proteinogenic amino acids, so the obligatory β -hydroxylation and epimerization reactions occur after the adenylation and thiolation steps.

The heterologous expression of HynC was complicated by fast precipitation of the hydroxylase after purification. This could be avoided by coexpression of HynC with each of the NRPS modules to test the hydroxylating activity (section 3.1.6). To circumvent *in vivo* hydroxylation of the substrate during expression of the proteins in the negative control, the inactive mutant HynC E376D was generated. Finally, to determine the substrate specificity of the hydroxylases, each of the NRPS modules was separately incubated with HynC and HynE and all necessary substrates (sections 3.1.7 and 3.1.8). The product amino acids were detected via LC-MS and clearly showed that HynE is

responsible for the generation of the hydroxyasparagines in modules four and five, whereas HynC hydroxylates T domain-bound L-Tyr and L-Leu in modules six and seven, respectively.

The different C₂-stereoconfiguration of the 3-hydroxyasparagines in module four and five was in line with the presence of a bifunctional dual condensation/epimerization domain in module six. Interestingly, module five harbours an additional, noncanonical C domain. Phylogenetic analyses indicated a C₃ epimerization activity for this domain, which however could not be verified in *in vitro* experiments. Together with detailed analyses of its primary structure, it can be postulated that the additional C domain is an evolutionary scar, which was inactivated but still remained within the BGCs (section 3.1.10).

According to the primary publication about hypeptin, 3-hydroxytyrosine was (2*S*,3*S*) configured, whereas 3-hydroxyleucine is (2*S*,3*R*) configured. The opposed configuration of the hydroxyl functionalities directly contradicted the results of the *in vitro* assay, which demonstrated the hydroxylation of both amino acids by the same enzyme. Finally, in-depth NMR-analyses of hypeptin, combined with results of the bioinformatic and biochemical investigations gave evidence to reassign the absolute stereoconfiguration of 3-hydroxytyrosine from (2*S*,3*S*) to (2*S*,3*R*) (section 3.1.9).

The outcome of this project is consistent with prior studies about NRPS, regarding the individual characterization of NRPS domains and in *trans* acting β-hydroxylases. However, the results collectively provide guidance for forthcoming researchers to efficiently implement a basic set-up for the characterization of NRPS and NRPS-dependent enzymes *in silico* and *in vitro*. On this basis, detailed investigations to unveil the nature of rare and non-canonical features of NRPS might be more easily achievable.

The enzyme families of in *trans* acting NHDMS and Fe/αKGs are poorly investigated. For NHDMS, the prototype CmlA was profoundly characterized in spectroscopic and structural aspects, but no further member was yet reported despite their abundance in the BGCs of known metabolites.^[58,60] On the other hand, several members of in *trans* acting Fe/αKGs are reported, but none of them are characterized in detail.^[48,52,53] Thus, relatively little is known about the difference between in *trans* acting Fe/αKGs and the related family of NRPS-related precursor generating iron- and α-ketoglutarate dependent dioxygenases. The characterization of HynC and HynE reported in this thesis adds one member to each of the families of in *trans* acting NHDMS and Fe/αKGs and might thus serve as basis for further investigations, e.g., by direct comparison of the primary structures or by utilization of more sophisticated bioinformatic tools like BiG-FAM, BiG-Scape, and CORASON.^[177]

The structure refinement of hypeptin was crucial for further investigations of the antibiotic, which is nicely illustrated in the case of teixobactin. Here, the total synthesis, based on the published structure, facilitated the generation of derivatives to unveil crucial residues for the activity, leading to a simplified, economic structure for research proposes.^[83] For hypeptin, rational design might include the substitution or elongation of the N-terminus with lipophilic residues to promote membrane anchoring, comparable to studies with teixobactin.^[178] Additionally, chemical synthesis might generate simplified analogues by substitution of the β -hydroxylated amino acids, as most of them are not commercially available. Threonine might serve as an appropriate substitute, at least for hydroxyleucine. Additionally, 3-hydroxytyrosine is prone to dehydration and might thus be difficult to handle in chemical synthesis.

4.2 Lysobactin

The antibiotic depsipeptide lysobactin was a promising candidate to treat nosocomial infections of Gram-positive bacteria, but its development was terminated after preclinical studies reported putative toxic effects.^[90] From a biosynthetic point of view, the respective NRPS BGC *lyb* exhibits the very interesting feature of the single *trans* acting NHDM LybC, which is predicted to be responsible for the β -hydroxylation of L-Phe, L-Leu, and L-Asn in modules three, four, and ten. Apart from the ability to recognize structurally diverse substrates, LybC would also have to distinguish between several L-Leu recruiting NRPS modules. Thus, the aim of this project was to reconstitute the activity of LybC *in vitro*. Afterwards, continuing experiments to elucidate the underlying mechanism of substrate recognition would have been conducted.

Comparable to the investigations on HynC, the activity of LybC was predicted to rely on the cognate NRPS module, which carries the amino acid substrate tethered onto the T domain. Thus, different constructs of the modules three, four, and ten were generated and tested for their A domain's activity (sections 3.2.1 and 3.2.2). Again, the results showed that larger constructs should be favoured in case of heterologous construction of NRPS. Interestingly, for all modules, the AT didomain displayed the lowest adenylation activity, compared to larger constructs, but also to the respective single A domain.

In a first attempt, LybC was characterized by UV/Vis absorption spectroscopy, showing a small feature at $\lambda=340$ nm, which is common for diiron oxidases in ferrous state (section 3.2.3). After reduction to ferric state, the feature diminished, which gave evidence that LybC was expressed correctly. Nevertheless, any attempt to reconstitute the hydroxylase with the NRPS modules *in vitro* failed, as no hydroxylated product could be detected.

The absence of positive results, which finally led to suspension of this project illustrate the difficulty to investigate in *trans* acting enzymes of NRPS. In this case, the necessary NRPS modules showed adenylating activity, verifying the functionality of at least the A domain within the constructs. Together with the spectroscopic results of LybC, this was a promising prerequisite for further investigations of the hydroxylase's activity, but the interplay of the enzymes still seemed to be disturbed (section 3.2.4). An elaborate agenda to systematically approach this issue would certainly have resulted in the successful reconstitutions of the enzymes, alike in the hypeptin project. Unfortunately, the shortage of time did not allow for the performance of such laborious work, and promising results in the hypeptin project let to its prioritisation.

4.3 FR900359

The potent and specific Gαq inhibitor, FR900359 (FR) is a nonribosomally synthesised cyclic depsipeptide, isolated from the soil bacterium *Chromobacterium vaccinii*. Its potential as a pharmaceutical tool or even as clinical drug draws the attention to find or (bio-)synthesize new FR derivatives with advantageous activities. Apart from the obligatory NRPS-encoding genes, the *frs* BGC also encodes for the *trans* acting NHDH FrsH, which is responsible for the β-hydroxylation of L-Leu in modules one, two, and seven of the NRPS. The simple system of the monomodular NRPS FrsA, which is responsible for side chain biosynthesis, displays a good opportunity to investigate the interplay of the NRPS and the *trans* acting hydroxylase FrsH. The primary aim of this project was to verify the proposed activity of FrsH. Secondly, the interaction of FrsH and FrsA was described by a series of biochemical techniques, structure elucidation and *in silico* computations.

To assess the activity of FrsH *in vitro*, different constructs of the cognate NRPS FrsA were generated and heterologously expressed together with the MLP FrsB (section 3.3.1). The adenylation activity of the embedded A domain was determined with the $\gamma^{18}\text{O}_4$ -ATP exchange assay for each construct, which all exhibited a high turnover (section 3.3.2). The presence of a diiron cluster in FrsH was verified by UV/Vis spectroscopy, showing the typical feature at $\lambda=340$ nm, which diminished upon chemical reduction (section 3.3.3). The proposed activity of FrsH was confirmed by the detection of the FR side chain *N*-propionylhydroxyleucine (*N*-Pp-Hle) after incubation with the CAT tridomain of FrsA (FrsA₁CAT) and all necessary substrates (section 3.3.4).

At first, the heterodimerization of FrsA₁CAT and FrsH were sought to be qualitatively and quantitatively investigated with size-exclusion chromatography and isothermal titration calorimetry (sections 3.4.2 and 3.4.3). Unfortunately, the interaction was not reproducible throughout the course

of experiments, probably due to the transient character of the interaction and the high flexibility of FrsA₁CAT.

Subsequently, a crystal structure of FrsH could be obtained, giving valuable insight into the overall structure and the surface properties of FrsH (sections 3.4.5 and 3.4.6). The three-dimensional fold of FrsH was almost identical to its homologue CmlA, but the surface characteristics differed heavily. This seemed to be a general trait of this enzyme family that could be used to detect the surface patch for the interaction. Therefore, the genes encoding for the analogous NHDMs HynC and LybC were transformed into a *frsH*-deficient strain of *C. vaccinii*, leading to successful reestablishment of FR production (section 3.4.4). Thus, HynC and LybC were able to interact with FrsA by a functionally conserved surface patch. Homology-generated three-dimensional structures of HynC and LybC were compared with the crystal structure of FrsH to locate candidate residues for this patch.

In parallel, a protein-protein docking server was used to generate potential complex structures of FrsH and the A domain of FrsA *in silico* (sections 3.4.8 and 3.4.9). Strikingly, the server docked the A domain to the identified conserved surface patches of FrsH. The complex structures were thus examined for surface residues with high potential to contribute to the heterodimerization and four of these residues were finally picked and mutated in the heterologously expressed enzyme. The ability of the generated mutants of FrsH and FrsA₁CAT to generate *N*-Pp-Hle versus the non-hydroxylated product was assessed *in vitro* to determine the influence of the respective mutation on the interaction of the enzymes (section 3.4.10). Unfortunately, none of the mutants significantly rendered the ratio of hydroxylated to non-hydroxylated product. Hence, six further residues were suggested for mutational analysis for the continuation of this project.

The question for the mechanism of substrate recognition of *in trans* acting enzymes in NRPS was probably posed together with the first report of such an enzyme. The involvement of NRPS domains apart from the T domain has been postulated several times, but little evidence has arisen from experiments conducted previously.^[8] The results of this project added hints towards a participation of the A domain in the recruitment of *trans* acting NHDM. If the continuing *in vitro* experiments validate this hypothesis, it would greatly contribute to the general comprehension of NRPS. These results would be of major interest for artificial engineering of NRPS to create assembly lines which yield new peptide products which also contain non-proteinogenic amino acid building blocks.^[117,179] Detailed knowledge about the structural prerequisites to include β -hydroxylated amino acids would help to achieve this aim.

The results of the *in vivo* substitution of FrsH by the homologues HynC and LybC create further implications for artificial engineering of NRPS. In particular, LybC, which has not been biochemically characterized before, has great potential as a multifunctional hydroxylase with

remarkably low specificity towards the amino acid substrate. It is tempting to postulate that an artificially introduced specificity conferring surface patch on an A domain facilitates the recruitment of LybC with subsequent hydroxylation of the T domain bound amino acid. The true potential of this result however, cannot be estimated due to the coiled interactions and proofreading activities throughout the NRPS system, for example by the acceptor site of the adjacent C domain.^[5,22,27,31]

5 Material and Methods

5.1 Chemicals and reagents

All common chemicals and solvents were purchased from Merck KgaA (Darmstadt, Germany) or Carl Roth (Karlsruhe, Germany). The used enzymes were obtained from New England Biolabs (Frankfurt am Main, Germany), Promeg (Mannheim, Germany) or Fermentas (St. Leon Rot, Germany).

5.2 Bacterial strains

The *E. coli* strains α -select silver, NEB® turbo, and NEB® 10-beta were used for cloning and storage of plasmids. BL21(DE3) was used for heterologous expression of proteins. For proteins containing a T domain, the modified BL21(DE3) strain BAP1 was used to ensure *in vivo* phosphopantetheinylation of the conserved serine-residue. All strains are listed in Table 5.1.

Table 5.1: List of all bacterial organisms.

Strain	Genotype	Origin
<i>E. coli</i> α -select silver	F' <i>deoR endA1 recA1 relA1 gyrA96 hsdR17</i> (r _k ⁻ , m _k ⁺) <i>supE44 thi-1 phoA</i> Δ (<i>lacZYA argF</i>)U169 Φ 80 <i>lacZ</i> Δ M15 λ ⁻	Bioline
<i>E. coli</i> NEB® Turbo	F' <i>proA⁺B⁺ lacI^q ΔlacZM15</i> / <i>fhuA2 Δ(lac-proAB)</i> <i>glnV galK16 galE15 R(zgb-210::Tn10)Tet^S endA1 thi-</i> <i>1 Δ(hsdS-mcrB)5</i>	NEB® inc.
<i>E. coli</i> NEB® 10-beta	Δ (<i>ara-leu</i>) 7697 <i>araD139 fhuA ΔlacX74 galK16</i> <i>galE15 e14- ϕ80dlacZΔM15 recA1 relA1 endA1</i> <i>nupG rpsL (Str^R) rph spoT1 Δ(mrr-hsdRMS-mcrBC)</i>	NEB® inc.
<i>E. coli</i> BL21(DE3)	<i>fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS</i> λ DE3 = λ sBamHIo Δ EcoRI-B	NEB® inc.
<i>E. coli</i> BAP1(DE3)	BL21(DE3) Δ <i>prpRBCD::T7prom-sfp,T7prom-prpE</i> [180]	Prof. Bradley Moore DSMZ
<i>Chromobacterium</i> <i>vaccinii</i> MWU205 (DSM25250)	Wild type (type strain)	DSMZ
<i>Lysobacter</i> sp. ATCC® 53042™	Wild type (type strain)	ATCC
<i>Lysobacter</i> sp. K5869	Wild type	Prof. Kim Lewis

5.3 Plasmids

Different plasmids were used for heterologous expression (Table 5.2). For maps of the unaltered plasmids, please see chapter 8.3

Table 5.2: Plasmids for heterologous protein expression.

Plasmid	Specification	Resistance for selection (concentration)	Origin
pET28a(+)	For heterologous expression of genes C- and N-terminal 6x His-tag Promotor: T7 Induction: 400 μ M IPTG	Kanamycin (50 μ g/mL)	Dr. Max Crüsemann
pCDFDuet-1(Apra)	For heterologous coexpression MCS1: N-terminal 6x His-tag MCS2: N-terminal Strep-tag Promotor: T7 Induction: 400 μ M IPTG	Apramycin (50 μ g/mL)	Nils Böhringer
pHis8-TEV ^[181]	For heterologous expression N-terminal 8x His-tag Promotor: T7 Induction: 400 μ M IPTG	Kanamycin (50 μ g/mL)	Dr. René Richarz
pG-KJE8	ORF for chaperones: <i>groEL</i> , <i>groES</i> , <i>dnaK</i> , <i>dnaJ</i> , <i>grpE</i> Promoter: <i>araB</i> , <i>Pzt-1</i> Induction: 0.5 mg/mL L-Arabinose, 5 ng/mL Tetracyclin	Chloramphenicol (25 μ g/mL)	TaKaRa®
pCv1	For complementation of <i>C. vaccinii</i> Promotor: VioA Holds <i>mob</i> gene for mobilization	Tetracycline (10 μ g/mL)	Dr. René Richarz
pTA-Mob	Mobilization helper plasmid Holds <i>tra</i> genes for conjugation	Gentamicin (20 μ g/mL)	Dr. René Richarz

5.4 Media and buffer

5.4.1 Media

Depending on the intended use, bacteria were cultivated either in liquid media or on solid agar plates. If an *E. coli* strain contained one or more plasmids, the media was supplemented with the respective antibiotics. All media were prepared with demineralized water and autoclaved immediately, all non-autoclavable ingredients were sterile filtered.

Table 5.3: Liquid media used for cultivation of bacterial strains.

Medium	Ingredient	Concentration	
Luria-Broth (LB)	Tryptone	10 g/L	
	Yeast extract	5 g/L	
	NaCl	10 g/L	
		pH 7.5	
LB-agar	Tryptone	10 g/L	
	Yeast extract	5 g/L	
	NaCl	5 g/L	
	Agar-agar	15 g/L	
		pH 7.5	
Terrific Broth (TB)	Tryptone	12 g/L	
	Yeast extract	24 g/L	
	Glycerol	5 g/L	
	10x TB-salts	100 mL/L	add after autoclaving
10x TB-salts	KH ₂ PO ₄	0.17 mM	
	K ₂ HPO ₄	0.72 mM	
		pH 7.2	
SOC	Tryptone	20 g/L	
	Yeast extract	5 g/L	
	NaCl	0.58 g/L	
	KCl	0.19 g/L	
	MgCl ₂	10 mM	add after autoclaving
	MgSO ₄	10 mM	add after autoclaving
	D-glucose	0.36 % (w/v)	add after autoclaving
Tryptic Soy Broth (TSB)	Tryptone	17 g/L	
	K ₂ HPO ₄	2.5 g/L	
	D-glucose	2.5 g/L	
	NaCl	5 g/L	
	Soy peptone	3 g/L	
		pH 7.3	

5.4.2 Buffers

In the described methods, the following buffers and solutions were used. All buffers were prepared with demineralized water. Further buffers are listed with the respective method.

Buffer	Ingredient	Concentration
50x TAE	Tris	2 M
	Acetic acid (glacial)	1 M
	EDTA	50 mM pH 8.0
Hydroxylation assay buffer	Tris	50 mM
	MgCl ₂	10 mM
	NaCl	25 mM pH 7.5
$\gamma^{18}\text{O}_4$ -ATP exchange assay buffer	Tris	20 mM
	Glycerol	5 % (v/v)
		pH 7.5
Lysis buffer	NaH ₂ PO ₄	50 mM
	NaCl	300 mM
	Imidazole	10 mM pH 8.0
Wash buffer I	NaH ₂ PO ₄	50 mM
	NaCl	300 mM
	Imidazole	20 mM pH 8.0
Wash buffer II	NaH ₂ PO ₄	50 mM
	NaCl	300 mM
	Imidazole	35 mM pH 8.0
Elution buffer	NaH ₂ PO ₄	50 mM
	NaCl	300 mM
	Imidazole	250 mM pH 8.0
Urea buffer	Tris	10 mM
	NaH ₂ PO ₄	100 mM
	Urea	8 M pH 8.0
SEC interaction buffer	Tris	50 mM
	MgCl ₂	10 mM
	NaCl	300 mM pH 7.5
ITC interaction buffer	HEPES	50 mM
	MgCl ₂	10 mM
	NaCl	300 mM pH 7.5

5.5 Microbiological methods

5.5.1 Cultivation of bacterial strains

Table 5.4 shows the conditions used to cultivate the organisms. *E. coli* consistently reached stationary phase when cultivated overnight. *Lysobacter* and *Chromobacterium* were cultivated 48 h or longer. For *Chromobacterium vaccinii*, quorum-sensing dependent production of the pigment violacein was a reliable indicator for dense growth.

Table 5.4: Cultivation conditions of organisms.

Organism	Medium	Temperature	Shaking speed
<i>E. coli</i> (all strains)	LB, TB	37 °C	220 rpm
<i>E. coli</i> with tetracycline resistance	LB	30 °C	220 rpm
<i>Chromobacterium vaccinii</i> MWU205	LB	25 °C	200 rpm
<i>Lysobacter</i> sp. ATCC53042	TSB	30 °C	200 rpm
<i>Lysobacter</i> sp. K5869	TSB	30 °C	200 rpm

5.5.2 Cryopreservation of bacterial strains

For stable long-time storage of a bacterial strain, 600 mL of a high-density culture was mixed with 600 mL autoclaved glycerol (50 %) as a cryoprotectant in a Fisherbrand® cryogenic storage vial and subsequently stored at -80 °C. To inoculate medium from a cryogenic culture, the storage vial was kept frozen in a -20 °C vial-rack and immediately stored back to -80 °C after use.

5.5.3 Chemical transformation

Chemical transformation was used to transfer one or more plasmids into a dedicated *E. coli* strain. To make bacteria chemically competent, 100 mL LB-medium were inoculated with 1 % over night-culture of the strain. At $OD_{600}=0.4$, the cells were harvested by centrifugation (4000 \times g, 5 min, 4 °C) and resuspended in 3.5 mL 70 mM CaCl₂, 20 mM MgSO₄. After 30 min incubation on ice, the suspension was mixed with 875 μ l 100 % glycerol and subsequently aliquoted to 50 μ l in 1.5 mL reaction tubes. Aliquots were stored at -80 °C.

For transformation, 10 μ l of a ligation sample or up to 100 ng of plasmids were carefully mixed to the frozen aliquot and placed on ice for 30 minutes. After incubation, the cells were heat-shocked at 42 °C for 40 s in a heat-block and immediately placed back on ice for at least 2 minutes. Subsequently, 1 mL of SOC-medium was added to the aliquot and cells were allowed to grow at 37 °C, shaking vigorously. After 1 h, the cells were harvested and plated in different concentrations onto LB-agar plates containing the appropriate antibiotics for plasmid selection. After 16 h incubation at 37 °C, singly-grown, round-shaped colonies were picked and further processed.

5.5.4 Electroporation

Electroporation is much more efficient, compared to chemical transformation.^[182] Hence, this method was used to transform product plasmids from complicated ligation or Gibson assemblies, or to add plasmids into strains that already harboured other plasmids.

100 mL LB-medium was inoculated with 1 % (v/v) of an over-night culture of the strain. At $OD_{600}=0.4$, the cells were harvested via centrifugation ($4000 \times g$, 5 min, 4 °C) and resuspended in 5 mL chilled, sterile 10% (v/v) glycerol. After two subsequent washing steps, each with 5 mL 10% glycerol, the cells were finally resuspended in 1 mL 10% glycerol and aliquoted to 70 μ l into UV-sterilized electroporation-cuvettes.

For transformation, up to 10 μ l of plasmid solution was added into a cuvette and carefully mixed. The cells were exposed to a voltage of 2.5 kV for up to 7 ms and immediately mixed with 1 mL SOC-medium. Afterwards, the suspension was transferred into a 1.5 mL reaction tube and incubated at 37 °C for 1 h, shaking vigorously. After 1 h, the cells were harvested and plated in different concentrations onto LB-agar plates containing the appropriate antibiotics for plasmid selection. After 16 h incubation at 37 °C, singly-grown, round-shaped colonies were picked and further processed.

5.6 Molecular biological methods

5.6.1 Primer design

Primers were designed with CloneManager9 or SnapGene-software and ordered from Eurofins MWG Operon (Ebersberg, Germany). The lyophilized primers were diluted in autoclaved ddH₂O to get a 100 μ M stock solution. For PCRs, aliquots with 1+9-diluted stock solution were used.

5.6.1.1 *Primer for sequential cloning*

The annealing sequence of the primers was 18 bp long. Outside the gene-coding sequence, the restriction site for a specific endonuclease was added. Endonucleases often lose their efficiency, when cleaving close to the end of DNA fragments, so 3 bp (4 in case of *Xho*I) were added in front of the restriction site. Depending on the desired positioning of the hexahistidine tag after expression, the stop codon 5'-AAT-3' was placed between the restriction side and the coding annealing sequence in the reverse primer. All primers are listed in appendix section 8.2.

5.6.1.2 *Primer for site-directed mutagenesis via Gibson assembly*

The 3' part of the primer consisted of 18 bp annealing sequence, followed by at least 16 bp of overlap sequence, which included the mutated codon. The overlapping sequence was identical to that in the template DNA, except for the mutation. All primers are listed in appendix section 8.2.

5.6.2 Polymerase chain reaction (PCR)

5.6.2.1 *Q5-PCR*

PCR is a widely used method to amplify specific DNA fragments with a length of several thousand base pairs. For cloning and subsequent heterologous expression, Q5 High-Fidelity DNA-polymerase (NEB) was used. This *pfu* polymerase harbours a native proof-reading activity and is engineered for fast amplification with ultra-low error rates, even of suboptimal amplicons. As templates, the genomic DNA of *Chromobacterium vaccinii* MWU205, *Lysobacter* sp. ATCC53042 or *Lysobacter* sp. K5869 were used. Table 5.5 shows the standard composition for 25 µL of a Q5-PCR reaction.

Table 5.5: Composition of Q5-PCR reactions.

Volume	Compound
5 µL	5x Q5 reaction buffer
5 µL	Q5 high GC enhancer
0.5 µL	dNTPs (10 mM)
0.625 µL	Primer, forward (10 µM)
0.625 µL	Primer, reverse (10 µM)
1 µL	Template DNA (0.5 µg/µL)
0.125 µL	Q5-DNA polymerase
Ad 25 µL	ddH ₂ O

The enzyme was added in the final step. The PCR-tubes were then placed into a PCR-cycler and submitted to the temperature regiment (Table 5.6).

Table 5.6: Temperature regiment of Q5-PCR reactions.

Step	Temperature	Time	
Initial denaturation	98 °C	30 s	
Denaturation	98 °C	10 s	30 cycles
Annealing	Variable	20 s	
Polymerisation	72 °C	20- 30 s/kbp	
Final polymerisation	72 °C	120 s	
Chilling	10 °C		

5.6.2.2 *Purification of PCR products*

In order to process PCR-products, they were extracted using the FastGene Gel/PCR Extraction Kit (Nippon Genetics), following the producer's manual. PCR products were eluted in 20- 40 µL autoclaved ddH₂O.

5.6.2.3 Colony-PCR

Colony-PCR as an analytical method was used to screen grown colonies for the successful transformation of genes into a heterologous host. Therefore, several colonies were picked from an agar plate with a sterile toothpick and suspended in 10 µl ddH₂O, each. The cells were lysed at 98 °C for 10 min and then used as PCR template. As this method is error-prone, positive hits needed to be verified by analytical restriction digest and sequencing. Furthermore, positive controls with gDNA as template and negative controls without DNA-template were performed. Table 5.7 shows the standard composition for 20 µL of a Colony-PCR reaction. The temperature regiment for *Taq*-based Colony-PCR is described in Table 5.8.

Table 5.7: Composition of Colony-PCR reactions.

Volume	Compound
4 µL	5x GoTaq green reaction buffer
1 µL	MgCl ₂ solution (10 mM)
1 µL	DMSO
0.33 µL	dNTPs (10 mM)
0.33 µL	Primer, forward (10 µM)
0.33 µL	Primer, reverse (10 µM)
1 µL	Template DNA (0.5 µg/µL)
0.1 µL	GoTaq-DNA polymerase
Ad 20 µL	ddH ₂ O

Table 5.8: Temperature regiment of Colony-PCR reactions.

Step	Temperature	Time
Initial denaturation	95 °C	300 s
Denaturation	95 °C	45 s
Annealing	Variable	45 s
Polymerisation	72 °C	60 s/kbp
Final polymerisation	72 °C	300 s
Chilling	10 °C	

5.6.3 DNA isolation

5.6.3.1 Isolation of genomic DNA

Genomic DNA was isolated from stationary liquid cultures with the “Genomic DNA isolation kit” from Sigma Aldrich (Merck). As this column-based approach yielded highly fragmented products, the “Wizard® Genomic DNA Purification Kit” was preferred later on. Both kits were used, following the producers gDNA extraction protocol for gram-negative bacteria.

5.6.3.2 Isolation of plasmids

Plasmids were isolated from stationary liquid *E. coli* cultures using either the “FastGene Plasmid Mini Kit” (Nippon Genetics) or the “PureYield™ Plasmid Miniprep System” (Promega), following the manufacturers protocol. Plasmids were eluted in 20-40 µL autoclaved ddH₂O.

5.6.4 Agarose gel electrophoresis

Agarose gel electrophoresis was used as an analytical method to verify the presence of the desired PCR-products or to analyse plasmid restriction digests. 1% (w/v) Agarose was dissolved in 1x TAE buffer in a microwave and stored at 60°C until further use. The solution was poured into a cast and a comb was placed into the solution to form wells for the samples. To analyse fragments with high molecular mass, such as gDNA, 0.7% (w/v) agarose was used. Once cooled down, the cast was filled with 1x TAE-running buffer. Before applying, 5 µL sample was mixed with 1 µL 6x TriTrack DNA-loading dye (ThermoFisher) or 6x Purple gel loading dye (NEB). As reference, 2 µL of GeneRuler DNA Ladder Mix (ThermoFisher) was added into an empty well. After loading, an electric tension of 90-120 V was applied for 25-45 min. To analyse the gel, it was stained for 5 min in 1% (v/v) ethidium bromide and subsequently destained in demineralized water. DNA bands were visualized in an UV-cabinet.

5.6.5 DNA isolation from agarose gels

When Q5-PCRs yielded several products that might hamper further ligation processes, the desired DNA-fragment needed to be isolated. Therefore, agarose gel electrophoresis was conducted with 200-250 µL of the PCR reaction (section 5.6.4). To avoid intercalation and subsequent mutation of the DNA by ethidium bromide, the lanes containing PCR products must not be stained. Thus, small rims from the product lanes were cut off with a scalpel and stained. These parts were visualized with UV-light and served as reference for the rest of the product lanes to cut out the desired PCR fragment. The small agarose slices were transferred into a 2 mL reaction tube and purified using the FastGene Gel/PCR Extraction Kit (Nippon Genetics), following the producer's manual. PCR products were eluted in 20- 40 µL autoclaved ddH₂O.

5.6.6 Restriction hydrolysis

Linear DNA and circular plasmids were hydrolysed to create fragments with overlapping (sticky) ends, which allow subsequent ligation. All enzymes were provided by NEB and used in CutSmart-Buffer, allowing simultaneous digestion with several restriction enzymes. **Table 5.9** shows the composition of a common double digest. The enzymes were incubated at 37 °C for 90 mins. If both enzymes were “time saver”-qualified, the incubation time was reduced to 30 min. For subsequent ligation, the enzymes were heat inactivated or cleaned using the FastGene Gel/PCR Extraction Kit.

Table 5.9: Composition of restriction hydrolysis reactions.

Volume	Compound
8 µl	DNA (max. 1 µg)
1	10x CutSmart Buffer
0.5 µl	Enzyme 1
0.5 µl	Enzyme 2

5.6.7 Dephosphorylation of DNA fragments

Enzymatic restriction hydrolysis yields phosphorylated 5'-ends of the DNA-fragments. This promotes religation of plasmids, when attempting to insert heterologous DNA fragments into the plasmid, resulting in a high percentage of false-positive outcome. To prevent this, 1 μ l of FastAP (Promega) was added to the restriction hydrolysis reaction 30 min prior to the end of the incubation.

5.6.8 Ligation

Ligation of fragments with compatible overhangs was conducted with T4 DNA-ligase (NEB). 8 μ l of educts in a molar ratio of 1:3 (plasmid:insert) were mixed with 1 μ l of 10x T4-ligase buffer and 1 μ l of T4-ligase. The reactions were incubated overnight at 16 °C or for 2 d at 4 °C. Subsequent transformation was performed without prior purification steps (section 5.5.3).

5.6.9 Gibson assembly

Gibson assembly is a method for isothermal DNA amplification and cloning. In this thesis, it was used for site-directed mutagenesis of plasmid constructs. Primers were designed with an annealing sequence of 18 bases prior to the mutated codon, following 16 bases to overlap in the assembly. Educt fragments were generated *via* Q5-PCR (section 5.6.2.1) with the to be mutated plasmid as template. The template DNA was then eradicated from the solution either by restriction hydrolysis with *dam*⁺/*dcm*⁺-methylation selective *DpnI* (section 5.6.6), or by purification with agarose gel electrophoresis (section 5.6.4). 5 μ l of the PCR-products were then mixed with 15 μ l assembly mixture (see Table 5.10) and incubated for 60 min at 50 °C. After incubation, the whole mixture was used for chemical transformation. For electroporation, only 5 μ l was used, to prevent short-circuits.

This method required the amplification of the whole plasmid, which was sometimes troublesome due to the complex nature of the long Gibson primers. Alternatively, a multistep approach was developed. Here, the forward Gibson primer was paired with a primer annealing to the sequence pET-RP and the reverse Gibson primer was paired with the T7 promotor primer. These primer pairs were used in separate PCR reactions with the plasmid as template to amplify the insert up- and downstream of the mutated site, each with an overhang of the plasmid on the opposite side. The purified PCR fragments were used in a Gibson assembly, together with the *NcoI/HindIII* restricted pET28a as third fragment.

Table 5.10: Composition of Gibson assembly mixture.

Volume	Compound
320 μ L	5x ISO buffer
0.64 μ L	T5 exonuclease (10 U/ μ L)
20 μ L	Phusion polymerase (2 U/ μ L)
Ad 1.2 mL	ddH ₂ O

5.6.10 Sanger sequencing

Constructed plasmids were sent to Eurofins genomics (Luxembourg) for Sanger sequencing. The plasmid pET28a(+) was sequenced with the primers T7 and pET-RP. For pCDFDuet-1(Apra), either ACYCDuetUP1 and DuetDOWN1 or DuetUP2 and T7term were used, depending on the investigated MCS. Results were supplied as FASTA in .txt-format, or as chromatogram in .pd4-format and analysed with SnapGene 5.

5.6.11 Generation of specific *E. coli* expression strains

For heterologous expression in *E. coli*, the genes of the desired biosynthetic modules and enzymes were cloned into the plasmids pET28a, pCDFDuet-1(Apra) or pHis8-TEV. These plasmids harbour a N-terminal His-tag, resulting in 6-8 histidines attached to the proteins during translation. This tag enables the selective purification of the enzymes with Ni-NTA-affinity chromatography from the cell lysate. The following workflow was generally used for cloning: The genes of the desired domains/enzymes were amplified with specific primers and purified. The fragments and the plasmid were double digested with the same set of enzymes and the primer was additionally dephosphorylated. After inactivation of the restriction enzymes and ligation, the newly formed plasmids were transformed into *E. coli* α SS. The grown colonies were screened by PCR and analytical restriction hydrolysis, following sequencing. Positive clones were cryopreserved. The plasmids were further transformed into *E. coli* expression strains BL21(DE3) or BAP1(DE3).

5.7 Protein purification and analysis

5.7.1 Protein expression

Protein expression was conducted in 30- 500 mL TB medium, dependent on the intended use and the final yield of the purified protein. If possible, cultures with more than 500 mL were avoided and split into smaller volumes as the yield and the volume do not correspond linearly. *E. coli* expression strains were cultivated in baffled Erlenmeyer flasks that had at least twice the volume of the medium to ensure sufficient aeration. The culture was inoculated with 1% of an overnight culture and incubated in a rotary shaker at 220 rpm and 37 °C, until it reached a density of OD₆₀₀=1.0. After cooling down in iced water, the expression was induced with 400 μM IPTG (final concentration) and incubation proceeded at 16 °C for 16 h.

5.7.2 Protein purification from expression culture

Cells were harvested from expression cultures by centrifugation (10,000 *x g*, 4 °C, 2 min). The supernatant was discarded, and the pellet was resuspended in 5 mL chilled lysis buffer per g pellet. All further steps were conducted on ice to avoid degradation of the proteins. The cells were lysed with a sonicator in 10-second intervals and the lysate centrifugated (13,000 *x g*, 4 °C, 10 min). The clear supernatant, containing all soluble proteins, was transferred into a new vial and mixed with Ni-NTA-agarose (600 μL/100 mL culture), which was washed with lysis buffer beforehand. This suspension was then agitated moderately on ice for 1 h, to allow binding of the His-tagged protein to the agarose matrix.

The suspension was poured onto a polypropylene column (5 mL- 25 mL), washed with 4 mL- 8 mL wash I buffer, 0 mL- 4 mL wash II buffer, and finally eluted with 1.5- 5 mL elution buffer. The applied volumes depended on the amount of agarose matrix, the affinity of the protein to the matrix, and the intended use of the elution fraction. 30 μL of each fraction (flow-through, wash I, wash II, elution) were sampled for analysis with SDS-PAGE. To check if higher amounts of the desired protein remained insoluble after sonication, the cell debris was resuspended in 0.5 mL- 2 mL urea buffer and incubated at room temperature for 1 h. Subsequently, solubilized proteins were separated by centrifugation (13,000 *x g*, 10 min) and 30 μl were sampled for SDS-PAGE.

5.7.3 Desalting and concentration

If the enzymes were utilized for *in vitro* assays, the buffer system had to be changed as high concentrations of imidazole might accelerate protein degradation. Thus, PD-10 columns (GE Healthcare) were utilized, following the manufacturer's gravity protocol. Proteins were concentrated in VivaSpin 500 filters (Merck) with a maximum molecular weight cut off of half the proteins size. High amounts ($\gg 10$ mg) of protein were also desalted using spin filters by concentrating and diluting the solution at least 3 times with the desired buffer.

5.7.4 Concentration measurement

Protein concentration was measured by the method of Gill and Hippel.^[183] Here, the molar extinction coefficient at $\lambda=280$ nm is estimated by the number of tryptophans, tyrosines and disulfide bonds in the protein. Absorbance was measured with an Eppendorf Biophotometer *kinetic* in a 1 mm quartz cuvette. The photometer calculated the protein concentration in mg/mL after measurement from the provided molar extinction coefficient of the protein, calculated by the method of Gill and von Hippel.

5.7.5 Storage of proteins

Purified proteins were only stored if preceding experiments demonstrated that the catalytic activity was not affected by the procedure. Otherwise, proteins were only used for activity assays directly after purification.

After purification and buffer exchange, proteins were diluted to the desired concentration and aliquoted à 200 μ L into 500 μ L reaction tubes. The aliquots were flash frozen in liquid nitrogen and instantly stored at -80 °C. After thawing, the proteins were reactivated with up to 5 mM DTT from a 100 mM stock solution, which was further diluted in the final *in vitro* assay to less than 1 mM.

5.7.6 Fast protein liquid chromatography (FPLC)

If very high purity of proteins was desired, size exclusion chromatography (SEC) as a polishing purification step after Ni-NTA-affinity chromatography was performed. The used column (Superdex 200 increase 10/300 GL; Cytiva) was connected to the ÄKTA-FPLC-System (Cytiva) and equilibrated with 50 mL of degassed and filtered buffer. SEC columns contain a highly porous matrix whereby the pores vary in size. Depending on their apparent size, larger molecules such as proteins cannot enter all pores, while small molecules diffuse freely through the matrix. Very large proteins cannot enter the matrix pores at all. Consequently, the effective column volume for each protein depends on its apparent size. Larger proteins move faster through the length of the column than smaller ones, resulting in a shorter retention volume. Common FPLC SEC columns have a rather small capacity, compared to column that are based on principles like ion exchange (IEX) or

affinity chromatography. Nevertheless, their advantage is the outstanding high purity of the protein in the isolated elution fraction.

After Ni-NTA-affinity chromatography, the protein was concentrated with a Vivaspın 500 filter to a maximum of 50 mg/mL and centrifuged (15,000xg, 10 min, 4 °C) to eliminate any particles. The solution was soaked into a 1 mL syringe without producing any bubbles. 100 µL per run was injected through a 0.22 µm syringe filter. The flow was set to 0.5 mL/min and samples of 0.3 mL were collected. If several runs were necessary to purify the whole batch of protein, runs were interlaced to save time and buffer. A run was stopped after 0.75 column volumes (~18 mL) to already start the next run with the same protein.

Chromatograms were analysed and visualized with UniCorn 4.1 Software.

5.7.7 SDS-Polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was used to analyse the molecular weight and the purity of protein samples. Therefore, 30 µL of sample were mixed with 10 µL 4x NuPAGE Sample buffer and 4 µL 10x NuPAGE Reducing agent and denaturised at 70 °C for 10 min.

The proteins were separated in a discontinuous, vertical system, provided by Novex-systems. **Table 5.11** shows the composition of the running and the stacking gels. At first, the compounds for the stacking gel were mixed and poured into a cassette. The solution was covered with isopropanol to archive a plain margin between the separating and the stacking gel. After polymerization, the isopropanol was removed carefully, and the stacking gel was poured into the cassette. Finally, a comb was stucked into the unpolymerized stacking gel to form the sample wells. Finished gels were wrapped in wet towels and stored at 4 °C for up to 2 weeks.

Table 5.11: Composition of SDS-PAGE gels.

Compound	Running gel 12 %	Stacking gel 6 %	
Bis- Tris acrylamide 32.5:1 30 %	4 mL	0.51 mL	
H ₂ O	3.4 mL	2.34 mL	
Buffer	2.5 mL	0.375 mL	
SDS (10 %)	100 µL	30 µL	
APS (10 %)	100 µL	30 µL	Add to start polymerisation
TEMED	4 µL	3 µL	

For analysis, the adhesive tape from the bottom of the cassette was removed, the cassette was tightly mounted into the running chamber and covered with running buffer (25 mM Tris, 200 mM glycine, pH 8.8). For the analysis of small MbtH-like proteins, the Tris-Tricin buffer system was used. The lower pK_s of Tricin leads to a wider spread of proteins between 1- 50 kDa. Therefore, the upper part of the gel was covered with cathode buffer (100 mM Tris, 100 mM Tricin pH 8.25) and the lower part was filled with anode buffer (200 mM Tris pH 8.9).

Depending on the concentration of the desired protein, 3- 12 μL of protein sample was filled into the well after removing the comb. As a reference marker, 3 μl of blue pre-stained protein standard was used. To avoid horizontal osmotic diffusion of the samples during the run, the amount of running buffers in adjacent wells was equalised by adding 4x NuPage sample buffer. Chromatography was conducted by applying 90- 120 V power. After 3 h, proteins were analysed with Coomassie staining. The gel was removed from the cassette and carefully washed with dH_2O to remove excess SDS. To stain, the gel was covered with staining solution (10% (v/v) glacial acetic acid, 50% (v/v) methanol, 1% (w/v) Coomassie (R)), briefly boiled in a microwave and incubated on a shaking plate for 10 min. Afterwards, the staining solution was removed by washing with dH_2O , the gel covered with de-staining solution (10% (v/v) glacial acetic acid, 40% methanol (v/v)) and two cellulose papers were added to adsorb Coomassie until the protein bands of the gel were sharply visible. Finalized gels were documented with a customary scanner.

5.7.8 Thrombolytic removal of N-terminal His-tag

Most proteins, which were used in this thesis were heterologously expressed in pET28a with an *N*-terminal hexahistidine tag to facilitate efficient purification via NiNTA- based affinity chromatography. The small tag was scarcely reported to interfere with the enzyme but might still inhibit its catalytic function. Thus, the plasmid also encodes for a thrombin recognition site between the tag and the inserted protein, which enables the removal of the tag post protein purification.

To do so, the buffer of the purified enzyme was exchanged to $\gamma^{18}\text{O}_4\text{-ATP}$ exchange assay buffer as described in section 5.7.3. Subsequently, 200 μg of protein was mixed with 1.5 U thrombin in 0.1% bovine serum albumin (BSA) in a final volume of 1 mL. After incubation for 18 h at 4 $^\circ\text{C}$, thrombin was inhibited with 100x protease inhibitor cocktail set I (Sigma-Aldrich). Proteins with removed hexahistidine tag were separated by inverse NiNTA chromatography, where the flow-through and wash fractions contain the protein of interest. The fractions were pooled, and the buffer system switched depending on the intended use of the protein.

5.7.9 Peptide cleavage utilizing TEV-protease

Proteins, which were heterologously expressed from the pHis8-TEV vector harboured a TEV cleavage site between the hexahistidine tag and the protein sequence. This facilitates the specific removal of the tag, as the TEV protease exhibits a very high specificity towards its peptide sequence, which scarcely occurs in proteins.

To remove the tag from the protein, it was expressed, purified and desalted to hydroxylation assay buffer as described above. After concentration measurement, it was mixed with TEV protease in a ratio (w/w) of 100+1 (substrate + protease) and incubated at 4 $^\circ\text{C}$ for 20 h. After incubation, cleaved proteins were isolated by inverted NiNTA affinity chromatography. Thus, the reaction was

supplemented with 0.5- 1.0 mL Ni-NTA-agarose and shaken on ice for 1 h. The suspension was poured into a polypropylene column, washed with 2- 4 mL wash buffer I and eluted with 3 mL elution buffer. The flow through- and wash fractions contained the desired protein without hexahistidine tag. To further purify the proteins, a SEC was conducted (see section 5.7.6).

5.7.10 Analytical size-exclusion chromatography

The size of proteins is commonly estimated by SDS-PAGE analysis. This method however requires denaturation of the proteins with SDS, which disrupts noncovalent interactions. In the case of FrsH, experimental data suggested a homodimeric structure in solution, which could not be verified by SDS-PAGE. Hence, the apparent size of the protein was visualized by analytical size-exclusion chromatography (SEC).

The column (TSKgel G3000SW_{XL} 7.8 x 300 mm, Tosoh Bioscience) was connected to a HPLC system, coupled to a DAD detector. Proteins were detected at $\lambda=280$ nm. The SEC interaction buffer served as mobile phase with a flow of 1 mg/mL. The calibration mixture consisted of 0.5 mg/mL Thyroglobulin, 1.0 mg/mL γ -globulin, 1.0 mg/mL Ovalbumin, 1.5 mg/mL Ribonuclease A, and 0.01 mg/mL *p*-aminobenzoic acid. For calibration, 20 μ l of the mixture was injected into the system. *P*-aminobenzoic acid is a small molecule and its retention volume equivalent to the column volume. Thus, its apparent size was defined to be 3.0 kDa, which approximately resembles the minimum size for a protein to get influenced by the column matrix. The linear regression was calculated using GraphPad Prism.

To measure its apparent size, FrsH was purified via NiNTA-based chromatography, buffered into SEC interaction buffer and concentrated to 1 mg/mL. 10 μ l of the protein solution was injected into the chromatographic system. The apparent size of the protein was calculated with the function describing the linear regression of the calibration mixture.

5.8 *In vitro* assays

5.8.1 $\gamma^{18}\text{O}_4\text{-ATP}$ exchange assay

The $\gamma^{18}\text{O}_4\text{-ATP}$ exchange assay^[120] enables simple and fast mass spectrometry-based measurement of activity and substrate specificity of A domains. Three solutions were prepared:

-Solution 1: 3 mM $\gamma^{18}\text{O}_4\text{-ATP}$, 15 mM MgCl_2 in 20 mM Tris pH 7.5

-Solution 2: 3 mM amino acid, 15 mM Na-pyrophosphate in 20 mM Tris pH 7.5

-Solution 3: 5 μM of the enzyme in 5 % (v/v) glycerol, 20 mM Tris pH 7.5

2 μL of each solution was mixed and incubated at 22 °C for 90 min. For the negative control, solution 2 contained Na-pyrophosphate, but no amino acid. The reaction was stopped by adding 6 μL 9-aminoacridine in acetone (10 mg/mL). The assay was centrifuged (15,000 $\times g$, 5 min) to eliminate precipitated enzymes from the solution.

The $\gamma^{18}\text{O}_4\text{-ATP}$ exchange assay was analysed by MALDI-TOF mass spectrometry at the Mass Spectrometry Core Facility in the Institute of Biochemistry and Molecular Biology, University of Bonn. 1 μL of the clear supernatant was transferred to a ground steel sample carrier. All measurements were performed with an AutoFlex III Maldi-Tof/ToF (Bruker) in negative mode, ranging from $m/z= 400- 1200$. One spectrum was acquired by adding 3x 2000 impulses in random walk mode. Data were analysed with DataAnalysis 4.2 (Bruker). The absolute substrate conversion was calculated from the integrals of the peaks at $m/z= 506$ for $\gamma^{16}\text{O}_4\text{-ATP}$ and $m/z= 508, 510, 512, 514$ for $\gamma^{18}\text{O}_4\text{-ATP}$.

$$\text{absolute substrate conversion} = \frac{{}^{16}\text{O}}{({}^{16}\text{O} + {}^{18}\text{O})}$$

The equilibrium molar ratio of unlabelled pyrophosphate to $\gamma^{18}\text{O}_4\text{-ATP}$ in the assay is 5+1. Therefore, 83.33 % detected substrate conversion corresponds to 100% substrate exchange.

$$\% \text{ exchange} = \frac{\text{absolute substrate conversion}}{0.8333}$$

5.8.2 Hydroxylation assay

To prove the functionality of the NHDM HynC, the native reaction was reconstituted *in vitro*. The NHDM and the cognate NRPS-domains (as AT didomains or as CAT tridomains) were expressed heterologously and purified via NiNTA-based affinity chromatography as described in sections 5.7.1 to 5.7.4. After desalting to hydroxylation assay buffer (50 mM Tris pH 7.5, 10 mM MgCl₂, 25 mM NaCl) and concentration measurement, the assay was set up as a one-pot reaction in 500 μL. 20 μM of the proteins, calculated with the extinction coefficient of HynC, was mixed with 10 μM methyl viologen, 1 mM NADH, 1 mM ATP and 1 mM amino acid (final concentrations). The reaction was incubated for 3 h at 22 °C and subsequently at 4 °C for 12 h. The assay was simultaneously performed with HynC E376D as negative control.

In contrast to HynC, the NHDM LybC from the *lyb* BGC is stable in solution without coexpression of NRPS modules. Thus, to assess its activity, LybC and the NRPS modules were heterologously expressed and purified separately. After purification, 10 μM LybC was mixed with 10 μM of the respective NRPS module and 10 μM methyl viologen, 1 mM NADH, 1 mM ATP and 1 mM amino acid (final concentrations). The reaction was incubated for 3 h at 22 °C and subsequently at 4 °C for 12 h. Negative controls were performed by heat inactivating the NHDM for 10 min at 80 °C after activation with NADH, or with an inactivated mutant of the enzyme.

After incubation, the assay was stopped by adding 10% (v/v) TCA and incubated on ice for 30 minutes. The proteins were collected *via* centrifugation (15,000 *x g*, 5 min) and washed twice with assay buffer to remove TCA. For thioester cleavage, the protein pellet was resuspended in 100- 200 μl KOH (0,1 M) and incubated in a heat block at 70 °C for 20 min at 400 rpm, followed by lyophilisation. The soluble amino acid was dissolved in 200 μL ddH₂O, the solution centrifuged and transferred into a LC-MS vial.

The samples were analysed with HPLC, coupled to DAD and ESI-MS. HPLC was performed with an EC 250/4.6 NUCLEODUR 10-5 HILIC-column (Macherey-Nagel) as stationary phase and an isocratic gradient of 70 % MeCN in H₂O as mobile phase. After 30 min at a flow of 1 mL/min, a new sample could be injected.

5.8.3 Fe/αKG-dependent hydroxylase functionality assay with HynE

The functionality of Fe/αKG-dependent hydroxylases was also proven in a 500 μL one-pot reaction. 250 μM hydroxylase was mixed with 500 μM cognate NRPS-domains, 1 mM ATP, 1 mM amino acid, 0.5 mM (NH₄)₂Fe(SO₄), 0.5 mM α-ketoglutarate, 0.5 mM Na-ascorbate and 0.5 mM DTT. After incubation for 3 h at 20 °C, sample preparation was performed as described for the hydroxylation assay with HynC (see section 5.8.2).

5.8.4 Side chain assembly assay

The side chain assembly assay was performed with FrsA₁CAT and FrsH and yielded the FR side chain *N*-Pp-Hle. The enzymes were heterologously expressed and purified via NiNTA based-affinity chromatography. After desalting to hydroxylation assay buffer, the reaction was set-up in a 500 μ L one-pot assay. Thus, 10 μ M FrsA₁CAT was mixed with 10 μ M FrsH, 1 mM ATP, 1 mM L-Leu, 0.5 mM sodium propionyl-CoA (CoALA Biosciences, Elign, USA-Texas), 10 μ M methylviologen, 1 mM NADH, and 0.1 mM DTT (all final concentration) and incubated for 180 min at 22°C. Sample preparation was performed as described for the hydroxylation assay with HynC. Samples were analysed with LC-MS (Waters e2695 Separation Module, coupled to an ESI, Acquity QDa, and 2998 PDA detector). HPLC was performed with an EC 250/4.6 NUCLEODUR 10-5 HILIC-column (Macherey-Nagel) as stationary phase and an isocratic gradient of A:B 10:90 (A: 95% 5 mM ammonium acetate pH 7.4, 5 % MeCN; B: 5 % 5 mM ammonium acetate pH 7.4, 95 % MeCN). After 20 min at a flow rate of 0.7 mL/min, the next sample could be injected (5 μ L per injection).

5.8.5 Epimerisation activity assay

To investigate the C₃ epimerisation activity of HynA₅C₂, the hydroxylation assay with HynE was performed in a 500 μ l one pot reaction. 20 μ M HynA₅C₂, 25 μ M HynE, 0.5 μ M HynTE, 1 mM ATP, 1 mM L-Asn, 5 mM DTT, 0.5 mM (NH₄)₂Fe(SO₄), 0.5 mM α -ketoglutarate, and 0.5 mM Na-ascorbate were mixed and incubated at 20 °C, 200 rpm overnight. Afterwards, the proteins were precipitated at 65 °C for 10 min and pelleted by centrifugation. The supernatant was mixed with an equal volume of borate buffer (200 mM boric acid pH 9.0). The protein pellet was washed twice with H₂O, then alkaline thioester cleavage was performed by incubation with 200 μ l 0.1 M KOH at 70 °C for 20 min. After incubation, 400 μ l borate buffer was added. Derivatisation in both samples (supernatant and pellet) was performed by adding 1050 μ l Fmoc-Cl (30 mM in MeCN). After 120 s at RT, the reaction was quenched by addition of 1050 μ l 1-adamantylamine hydrochloride (ADAM) (120 mM). Precipitants were removed by centrifugation. Afterwards, the samples were transferred into pear shaped flasks to evaporate MeCN. The remaining H₂O was removed by lyophilization. Soluble parts of the remaining white powder were collected with 200 μ l mobile phase solvent (MeCN/0.1% FA 30/70) and 1 μ l thereof was injected into the HPLC system. Stationary phase: LuxCellulose-2 4.6x250 (Phenomex®); mobile phase: MeCN/0.1% FA 30/70, 1 mL/min, 60 min; Detection: DAD λ_1 =263 nm λ_2 =301 nm.

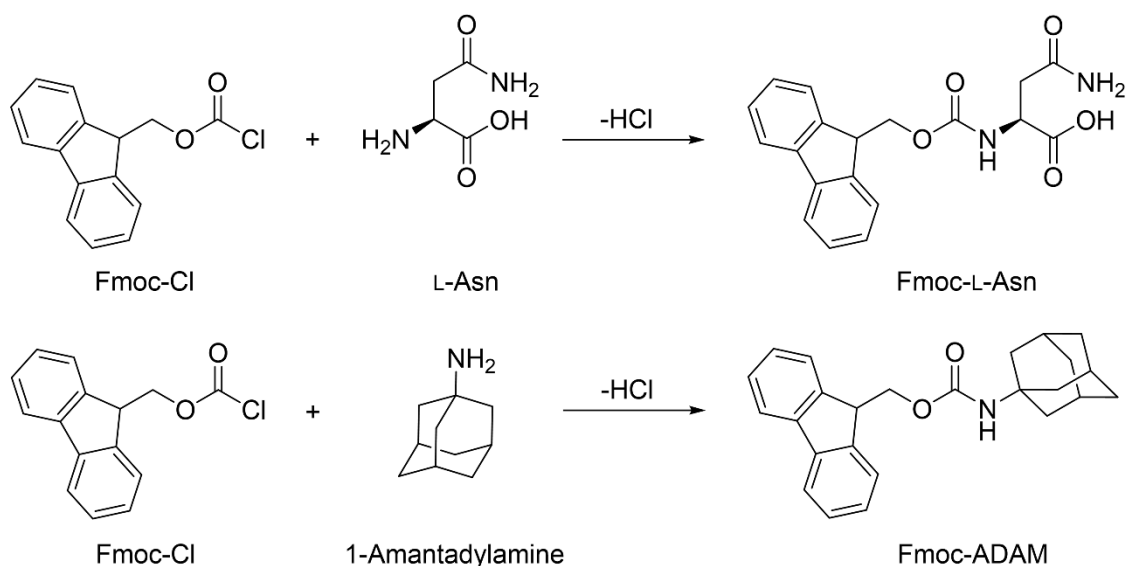


Figure 5.1: Derivatization of L-Asn and 1-adamantylamine (ADAM) with Fmoc-Cl.

5.9 Characterization of protein-protein interactions

5.9.1 Size exclusion chromatography

Size-exclusion chromatography (SEC) was used to analyse protein-protein interactions in a qualitative manner. The enzymes were heterologously expressed, purified via Ni-NTA affinity chromatography and buffered to assay buffer (50 mM Tris pH 7.5, 10 mM MgCl₂, 25 mM NaCl, 5 mM DTT). The NHDM was activated by incubation with equimolar amount of methyl viologen and twice the molar amount of Na₂S₂O₄. For each experiment, three samples were prepared:

-Sample 1: 500 μL NHDM (25 μM)

-Sample 2: 500 μL NRPS-module (25 μM)

-Sample 3: 500 μL of sample 1 mixed with 500 μL sample 2

All samples were supplemented with 1 mM ATP and 1 mM amino acid (final concentration), if not stated otherwise, and incubated at 20 °C for 90 min. After incubation, samples were concentrated with Vivaspın 500 membrane filters to roughly 100 μL and sequentially analysed via SEC, following the protocol described above (see 5.7.7). To potentiate non-polar interactions, the running buffer was enriched with NaCl (50 mM Tris pH 7.5, 10 mM MgCl₂, 300 mM NaCl). All collected fractions were analysed with SDS-PAGE.

5.9.2 Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry was used to measure quantitatively the dissociation constants of the protein-protein interaction between a NHDM and the cognate NRPS module. Both proteins were heterologously expressed and purified via SEC (see 5.7.6) to yield highly pure proteins in ITC-interaction buffer (50 mM HEPES pH 7.5, 10 mM MgCl₂, 300 mM NaCl). Measurements were performed with a PEAQ-ITC (Malvern). The concentrated NHDM FrsH (1.000 mM) was filled into the syringe and the NRPS module FrsA₁CAT (0.071 mM) into the reaction chamber. The temperature was adjusted to 20 °C and the stirring speed set to 790 rpm. After initial delay of 150 s and a first injection of 0.4 µL, 9 injections of 3 µL were performed every 200 s. Heat differences to the reference cell were detected in high feedback mode. Collected data were saved in .itc-format and evaluated using a triplet of analytical software. At first, raw data were loaded to NITPIC,^[154] to execute baseline correction of the thermogram, using default settings. Afterwards, the processed data were submitted to sedphat for analytical evaluation.^[155] The interaction model was set to “A + B <-> AB Hetero-Association” with titration of B into A. The following boxes were ticked to allow floating of the parameters during the fitting process: incompetent fraction B, log(Ka), dHAB, and baseline. The fitting was performed several times with different initial suggestions for the floating parameters, until it resulted in a decent curve fitting and low chi². Afterwards, it was verified that the floating parameters were still reasonable. Data were then submitted to Gussi for graphical illustration.^[156]

5.10 *In vivo* complementation of FrsH

5.10.1 Construction of complemented strains

For complementation of *C. vaccinii* knock-out strains, the gene of interest was cloned into the plasmid pCv1. This plasmid holds a *mob* gene, making it susceptible for mobilization. Thus, it was used for transformation by a helper strain into *C. vaccinii*. This method was established by Dr. René Richarz, as common chemical transformation or electroporation is not feasible with chromobacteria.^[13] The gene of the desired enzyme was amplified with specific primers and purified. The fragment and the plasmid were double digested with appropriate restriction enzymes to subsequently insert the PCR product into the *NcoI*-restriction site of the plasmid. After dephosphorylation of the plasmid, the enzymes were heat-inactivated and the DNA products ligated. The newly formed plasmids were transformed into *E. coli* NEB® 10-beta. The grown colonies were screened by PCR and analytical restriction hydrolysis, following sequencing. After verification, the plasmid was transformed into *E. coli* NEB® 10-beta, which already harboured the plasmid pTA-Mob by electroporation. After transformation, the cells were allowed to grow on agar

plates, containing half of the respective antibiotics, as the two plasmids harbour the same origin of amplification and thus need longer to adapt to the antibiotics. Grown colonies were screened for incorporation of pCv1 by PCR. A pre-preculture of a positive colony was grown with half of the concentration of the respective antibiotics to inoculate the preculture, supplemented with the common amount of antibiotics. Mobilization was conducted by growing 20 mL cultures of *C. vaccinii* and the helper strain each to $OD_{600}=0.4$. 2 mL of each culture was centrifuged and the cell pellet washed two times with LB. After washing, 0.2 mL of *C. vaccinii* was mixed with 0.6 mL of the helper strain and then plated onto LB-agar without antibiotics. After 20 h incubation at 30 °C, the resulting cell layer was scraped off and resuspended in 1 mL LB. 100 µL of an 1:10,000 and an 1:100,000 dilution were plated on selective agar (Amp200 + Tet10) and incubated over night at 30 °C until first colonies became visible. Colonies were screened by PCR for uptake of pCv1 without integrating it into the genome. Positively complemented knock-out strains were further subtracted to fermentation and extraction.

5.10.2 Fermentation of *C. vaccinii* and extraction of FR

10 mL of LB media were inoculated with *C. vaccinii* from an agar plate. All strains were supplemented with ampicillin and all strains harbouring the plasmid pCv1 were also supplemented with tetracycline. The next day, 40 mL of LB was inoculated with the preculture in a ratio of 1:100 and incubated for 36 h. Extraction was conducted with 40 mL butanol for 12 h. After incubation, the organic phase was separated by centrifugation (4000 rpm, 15 min) and evaporated to dryness. The extract was diluted in methanol to 1 mg/mL prior to LCMS analysis. For LC-MS analysis, a Waters e2695 Separation Module, coupled to an ESI, Acquity QDa, 2998 PDA detector was used. HPLC was performed with a XBridge® Shield RP18 3.5 µm 2.1 x 100 mm column as stationary phase. Elution started with a flow of 0.3 mL/min at 80 % A (95% 5 mM ammonium acetate pH 7.4, 5% MeCN) and 20 % B (5% 5 mM ammonium acetate pH 7.4, 95% MeCN) in a linear gradient to 80% B over 20 minutes, following an isocratic step with 80% B for 10 min. 5 µL of sample was injected per run.

5.11 Bioinformatic methods

5.11.1 Determination of the reading frame

The domains of NRPS are usually part of a mega-enzyme. Thus, they do not contain a start- or stop-codon, which would indicate the reading frame of the gene. To ensure in-frame cloning of the desired NRPS-domains, their genes were automatically aligned against known homologues domains by AntiSmash.^[112] After the primer design, the complete process of cloning was performed *in silico* with SnapGene software to eliminate any errors and ensure that the cloned gene was in-frame with the plasmids start codon and the histidine tag.

5.11.2 AntiSmash

AntiSmash is a free online tool that allows rapid identification, annotation and analysis of biosynthetic gene clusters from bacteria.^[112] After upload of a DNA-sequence (small contigs to whole genomes) in .gbk, .embl or .fasta-format, the software utilises a large number of *in silico* tools, to perform an in-depth analysis of the genes. The results can be investigated in a browser-supported surface or downloaded in .gbk-format. AntiSMASH is powered by several open-source tools: NCBI BLAST+, HMMer 3, Muscle 3, FastTree, PySVG and JQuery SVG.

5.11.3 Construction of phylogenetic trees with MEGA X

MEGA X is a free software, designed to analyse DNA and protein sequences from an evolutionary perspective.^[184] The current version, MEGA 11, was used in graphical user interface mode. To construct a phylogenetic tree, all protein sequences in .fasta-format were imported and aligned, using the MUSCLE algorithm with default setting. The resulting alignment was revised for plausibility and further processed by deletion of non-overlapping sequence termini. Sequences with multiple non-aligning insertions (e.g., A domains with inserted MT domains) were excluded if the deletion of the insertion was not reasonable. The resulting .meg-file was then used for maximum likelihood tree construction, using default setting except for the substitution model, here, the Dayoff or Jones-Taylor-Thornton model were each tested. If the resulting tree was consistent overall, the calculation was repeated with 100 rounds of bootstrapping to test the validity of the branches. The resulting tree was exported in Newick format and finally visualized with “iTOL v6”.^[185]

5.11.4 Visualization of proteins with PyMOL

PyMOL is a commercially available software, dedicated to the molecular visualization of three-dimensional structures such as molecules and proteins. The basic version additionally provides elementary features to analyse the loaded molecules like measurement of distances, detection of binding partners and other nearby residues, visualization of surfaces and cavities, alignment of two or more structures, and many more. Due to its open-source character, many free add-ins facilitate the in-depth analyses upon implementation. Furthermore, PyMOL facilitates manipulation of the

loaded structure by addition or deletion of single atoms, residues, and whole chains, or by alternation of atom bonds in regards of torsion angle and length.

Within this thesis, all figures that illustrate a proteinogenic structure were created with PyMOL.

For the structural comparison of the NHDMs (chapter 3.4.6) the structures of LybC and HynC were generated by SWISS-MODEL (see chapter 5.11.5) and aligned to the obtained structure of FrsH with the “Alignment” plugin. All acetate ligands as well as the crystal water was removed, and the structures analysed with the ABPS Electrostatics algorithm in default settings. This algorithm provides a colour coded surface structure for each enzyme that indicates positive charges in blue and negative charges in red. Neutral spots are coloured in white. The strength of this algorithm is the detailed analysis, where each atom and not the whole amino acid is evaluated, which is important for the assessment of amphiphilic residues like arginine. By switching between the surface visualization, the electrostatic properties of each spot were compared between the individual enzymes, therefore, possible flexibility of the amino acid side chains or the protein backbone was taken into account.

The results of the *in silico* docking of FrsH or LybC and FrsA₁A_{main}/YbdZ, performed with SwarmDock and ClusPro were also analysed with PyMOL (sections 3.4.8 and 3.4.9). The top ten results of each docking experiment were loaded and aligned to the initial structure of FrsH, thus creating a single structure of FrsH, docked to FrsA₁A_{main} in every probable conformation and orientation. Each docking result was inspected individually for plausibility regarding its orientation and position and eventually deleted. For example, structures were dismissed, if the binding interface of the A domain was predicted to consist of the active site or if the binding site of the A domain was natively occupied by the upstream C domain. The remaining complex structures were further examined in detail to elucidate amino acid side chains that contribute to the dimerization. Here, the tools to measure distances between single atoms and to detect nearby residues of other molecules gave valuable information about the formation of the complex structure.

5.11.5 SWISS-MODEL

SWISS-MODEL is an automated server to predict the structure of proteins based on their primary structure and homology-modelling. Its application is based on a user-friendly web interface and free of charge. The modelling workflow is generally described in five steps. First, the user provides the primary structure of the desired protein in FASTA, clustal, or plain text format. Next, the algorithm uses the given sequence as template to search for related proteins in the dedicated SWISS-MODEL Template Library (SMTL), which actually mirrors the whole protein data bank (PDB) but contains further information about each structure and stores them in a special format. In the third step the resulting homologous structures from SMTL are evaluated for their likeliness to serve as a decent template for model building. Eventually, several templates are chosen to cover different

regions of the target protein. Fourth, the thus selected homologues serve as templates for 3D modelling of the given protein. Partial structures that are not covered by the template are calculated separately. Lastly, the algorithm uses several scoring functions to evaluate the quality of the resulting structure.^[172]

The straightforward approach of this tool, allows to generate structures of decent quality in less than 15 minutes for a 60 kDa protein.

5.11.6 I-TASSER

I-TASSER is currently the best on-line server for the prediction of proteins structure according to the current Critical Assessment of Techniques for Protein Structure Prediction.^[186] The underlying modelling workflow consists of three steps. First, the user submits the primary sequence of the protein, which is then used as query to identify proteins of similar fold from the PDB library by a dedicated algorithm. Secondly, the protein fragments with similar fold are retrieved from PDB and used to generate a homology model from the initial protein sequence. Depending on the number of similar structures, several template-based models are created. Missing parts are calculated *ab initio*. In the last step, the just generated model is used as a query to repeat the second step with additional algorithms to remove steric clashes and improve the overall topology of the structure. The final results are ranked according to different scoring functions and displayed to the user.^[139]

Due to the sophisticated algorithms that empower I-TASSER, the resulting models are of high quality, even if templates with high similarity are scarce. Nevertheless, the long computation time due to the complex calculations and a high server workload are a major drawback. The calculation of FrsA₁CAT took roughly 72 hours.

5.11.7 SwarmDock

SwarmDock is a webserver with simple browser interface, dedicated to the generation of three-dimensional structures of protein-protein complexes.^[170] The user has to submit the structures of the proteins in .pdb-format, which also facilitates the submission of homo- and heterodimeric protein. The larger protein should be given as receptor and the smaller as ligand. SwarmDock only accepts proteinogenic amino acids within the structures as input. Thus, the structures of FrsH and LybC were pre-processed with PyMOL to delete crystal water, as well as all ligands, including the iron cores of the active site.

In a first step, the server pre-processes the given structures for the actual modelling algorithm. This includes modelling of missing or disordered atoms and removal of posttranslational modifications (if present). Additionally, an energy minimizing tool called CHARMM is applied. If requested, the processed structures may be downloaded by the user. For the actual docking procedure, a set of roughly 120 starting positions is evenly distributed over the surface of the receptor protein. At each

starting point, the ligand structure is posed above the receptor and the overall energy minimized by variation of its orientation and position. If the energetic potential of a distinct position cannot be further diminished, it is stored and the next round docking round is started. The underlying algorithm for energy minimization includes particle swarm organization, hinting towards the server's name. For every starting position, the whole docking process is repeated four times. Afterwards, all complex structures are again energy minimized by the CHARMM package. The resulting structures are then evaluated according to CCharPII which integrates a plethora of initially stand-alone programs to account various parameters and finally lists the complexes dependent on the so called "democratic" score. Finally, all calculated complex structures are provided to the user in .pdb format.^[170] The top ten structures should be taken into account for further analyses.

Due to the sophisticated particle swarm optimization for energy optimization and the "democratic" scoring, SwarmDock belongs to the best servers for *in silico* protein docking. Nevertheless, without any initial information about the interaction, the chance that it is correctly postulated in the top ten hits is below 50 %.^[173]

5.11.8 ClusPro

According to the latest CAPRI CAS evaluation, ClusPro is the second-best server for *ab initio* modelling of protein-protein complexes with only slightly worse performance compared to SwarmDock.^[173] The user has to submit the structures in .pdb-format with the larger structure given as "receptor" to minimize the resources needed for the computation. ClusPro does not pre-process the given structures, so all non-conventional residues have to be altered or removed. Additionally, energy minimizing has to be performed with a separate tool. For this thesis, the pre-computed structures from SwarmDock were submitted to spare this step.

The docking of two proteins is achieved in three steps. Initially, the PIPER algorithm performs rigid body docking, whereby the receptor is in a fixed position and the ligand sampled on a sphere-based grid around it. On each pixel of the grid, which corresponds to a translation of 1 Å on one axis of the coordinate system, 70,000 rotations of the ligand are performed. The strength of the PIPER algorithm is a Fast Fourier Transform (FFT) correlation approach that facilitates efficient evaluation of the energy function of each of the billion sampling points. The program considers one major drawback of rigid body docking by allowing steric clashes of the proteins to simulate backbone flexibility. In the next step, the 1000 lowest energy docked structures are retained and among these, highly populated docking clusters, representing hot spots with high probability to represent the native interaction, are detected. The centres of these clusters are used to model the ligand to the receptor with preferably low overall energy. The resulting complex structures are then energy minimized.^[187] Finally, the top ten results are chosen as output, whereby the receptor and the ligands are given in separate files.

One major advantage of ClusPro is that the PIPER algorithm generates four sets of models by weighting different terms in the calculation of overall energy. Thus, the outcome is presented with the scoring schemes called balanced, electrostatic-favoured, hydrophobic-favoured, and van der Waals + electrostatics. For some kinds of protein-protein interactions, the properties can be confidently assumed, so the respective scoring model would yield more accurate complex predictions. For this thesis, the scoring scheme “balanced” was chosen for further evaluation.

6 References

- [1] M. H. Medema, T. de Rond, B. S. Moore, *Nat. Rev. Genet.* **2021**, *22*, 553.
- [2] K. N. Raymond, E. A. Dertz, S. S. Kim, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3584.
- [3] J. E. Spraker, L. M. Sanchez, T. M. Lowe, P. C. Dorrestein, N. P. Keller, *ISME J.* **2016**, *10*, 2317.
- [4] M. J. Cryle, A. Meinhart, I. Schlichting, *J. Biol. Chem.* **2010**, *285*, 24562.
- [5] M. Kaniusaite, J. Tailhades, E. A. Marschall, R. J. A. Goode, R. B. Schittenhelm, M. J. Cryle, *Chem. Sci.* **2019**, *10*, 9466.
- [6] a) B. Shen, L. Du, C. Sanchez, D. J. Edwards, M. Chen, J. M. Murrell, *J. Nat. Prod.* **2002**, *65*, 422; b) H. Ueda, T. Manda, S. Matsumoto, S. Mukumoto, F. Nishigaki, I. Kawamura, K. Shimomura, *J. Antibiot.* **1994**, *47*, 315.
- [7] M. Alanjary, C. Cano-Prieto, H. Gross, M. H. Medema, *Nat. Prod. Rep.* **2019**, *36*, 1249.
- [8] R. D. Süßmuth, A. Mainz, *Angew. Chem. Int. Ed.* **2017**, *56*, 3770.
- [9] A. M. Gehring, I. Mori, C. T. Walsh, *Biochemistry.* **1998**, *37*, 2648.
- [10] T. Stachelhaus, H. D. Mootz, M. A. Marahiel, *Chem. Biol.* **1999**, *6*, 493.
- [11] J. Beld, E. C. Sonnenschein, C. R. Vickery, J. P. Noel, M. D. Burkart, *Nat. Prod. Rep.* **2014**, *31*, 61.
- [12] A. Chhabra, A. S. Haque, R. K. Pal, A. Goyal, R. Rai, S. Joshi, S. Panjekar, S. Pasha, R. Sankaranarayanan, R. S. Gokhale, *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 5681.
- [13] C. Hermes, R. Richarz, D. A. Wirtz, J. Patt, W. Hanke, S. Kehraus, J. H. Voß, J. Küppers, T. Ohbayashi, V. Namasivayam et al., *Nat. Commun.* **2021**, *12*, 144.
- [14] X. Gao, S. W. Haynes, B. D. Ames, P. Wang, L. P. Vien, C. T. Walsh, Y. Tang, *Nat. Chem. Biol.* **2012**, *8*, 823.
- [15] a) S. D. Bruner, T. Weber, R. M. Kohli, D. Schwarzer, M. A. Marahiel, C. T. Walsh, M. T. Stubbs, *Structure.* **2002**, *10*, 301; b) J. J. May, N. Kessler, M. A. Marahiel, M. T. Stubbs, *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12120.
- [16] E. J. Drake, B. R. Miller, C. Shi, J. T. Tarrasch, J. A. Sundlov, C. L. Allen, G. Skiniotis, C. C. Aldrich, A. M. Gulick, *Nature.* **2016**, *529*, 235.
- [17] Y. Katsuyama, K. Sone, A. Harada, S. Kawai, N. Urano, N. Adachi, T. Moriya, M. Kawasaki, K. Shin-Ya, T. Senda et al., *Angew. Chem. Int. Ed.* **2021**.
- [18] J. M. Reimer, M. Eivaskhani, I. Harb, A. Guarné, M. Weigt, T. M. Schmeing, *Science.* **2019**, 366.
- [19] B. R. Miller, E. J. Drake, C. Shi, C. C. Aldrich, A. M. Gulick, *J. Biol. Chem.* **2016**, *291*, 22559.
- [20] A. M. Gulick, *ACS Chem. Biol.* **2009**, *4*, 811.
- [21] K. J. Esquilín-Lebrón, T. O. Boynton, L. J. Shimkets, M. G. Thomas, *J. Bacteriol.* **2018**, *200*.

- [22] T. Izoré, M. J. Cryle, *Nat. Prod. Rep.* **2018**, *35*, 1120.
- [23] A. Chen, R. N. Re, M. D. Burkart, *Nat. Prod. Rep.* **2018**, *35*, 1029.
- [24] K. Bloudoff, T. M. Schmeing, *Biochim. Biophys. Acta, Proteins Proteomics* **2017**, *1865*, 1587.
- [25] V. Bergendahl, U. Linne, M. A. Marahiel, *Eur. J. Biochem.* **2002**, *269*, 620.
- [26] S. A. Samel, G. Schoenafinger, T. A. Knappe, M. A. Marahiel, L.-O. Essen, *Structure.* **2007**, *15*, 781.
- [27] T. Izoré, Y. T. Candace Ho, J. A. Kaczmarek, A. Gavriilidou, K. H. Chow, D. L. Steer, R. J. A. Goode, R. B. Schittenhelm, J. Tailhades, M. Tosin et al., *Nat. Commun.* **2021**, *12*, 2511.
- [28] T. A. Lundy, S. Mori, N. Thamban Chandrika, S. Garneau-Tsodikova, *ACS Chem. Biol.* **2020**, *15*, 282.
- [29] J. M. Reimer, M. N. Aloise, P. M. Harrison, T. M. Schmeing, *Nature.* **2016**, *529*, 239.
- [30] K. J. Labby, S. G. Watsula, S. Garneau-Tsodikova, *Nat. Prod. Rep.* **2015**, *32*, 641.
- [31] S. Dekimpe, J. Masschelein, *Nat. Prod. Rep.* **2021**.
- [32] a) T. Kittilä, C. Kittel, J. Tailhades, D. Butz, M. Schoppet, A. Büttner, R. J. A. Goode, R. B. Schittenhelm, K.-H. van Pee, R. D. Süßmuth et al., *Chem. Sci.* **2017**, *8*, 5992; b) M. Bernhardt, S. Berman, D. Zechel, A. Bechthold, *ChemBioChem.* **2020**, *21*, 2659.
- [33] M. C. White, J. Zhao, *J. Am. Chem. Soc.* **2018**, *140*, 13988.
- [34] D. A. Wirtz, K. C. Ludwig, M. Arts, C. E. Marx, S. Krannich, P. Barac, S. Kehraus, M. Josten, B. Henrichfreise, A. Müller et al., *Angew. Chem. Int. Ed.* **2021**.
- [35] J. Hou, L. Robbel, M. A. Marahiel, *Chem. Biol.* **2011**, *18*, 655.
- [36] A. A. Yunis, *Annu. Rev. Pharmacol. Toxicol.* **1988**, *28*, 83.
- [37] F. P. Guengerich, M. R. Waterman, M. Egli, *Trends Pharmacol. Sci.* **2016**, *37*, 625.
- [38] A. Greule, J. E. Stok, J. J. de Voss, M. J. Cryle, *Nat. Prod. Rep.* **2018**, *35*, 757.
- [39] D. J. Cook, J. D. Finnigan, K. Cook, G. W. Black, S. J. Charnock, *Adv. Protein Chem. Struct. Biol.* **2016**, *105*, 105.
- [40] C. E. Wise, T. M. Makris, *ACS Chem. Biol.* **2017**, *12*, 1316.
- [41] H. Chen, C. T. Walsh, *Chem. Biol.* **2001**, *8*, 301.
- [42] S. Uhlmann, R. D. Süßmuth, M. J. Cryle, *ACS Chem. Biol.* **2013**, *8*, 2586.
- [43] K. Haslinger, C. Brieke, S. Uhlmann, L. Sieverling, R. D. Süßmuth, M. J. Cryle, *Angew. Chem. Int. Ed.* **2014**, *53*, 8518.
- [44] C. Q. Herr, R. P. Hausinger, *Trends Biochem. Sci.* **2018**, *43*, 517.
- [45] M. S. Islam, T. M. Leissing, R. Chowdhury, R. J. Hopkinson, C. J. Schofield, *Annu. Rev. Biochem.* **2018**, *87*, 585.
- [46] S. P. de Visser, *Chem. Rec.* **2018**, *18*, 1501.
- [47] C. R. Zwick, H. Renata, *Nat. Prod. Rep.* **2020**, *37*, 1065.
- [48] J. M. Neary, A. Powell, L. Gordon, C. Milne, F. Flett, B. Wilkinson, C. P. Smith, J. Micklefield, *Microbiology.* **2007**, *153*, 768.

- [49] M. Strieker, E. M. Nolan, C. T. Walsh, M. A. Marahiel, *J. Am. Chem. Soc.* **2009**, *131*, 13523.
- [50] S. Groß, B. Schnell, P. A. Haack, D. Auerbach, R. Müller, *Nat. Commun.* **2021**, *12*, 1696.
- [51] Y. Li, L. Liu, G. Zhang, N. He, W. Guo, B. Hong, Y. Xie, *Front. Chem.* **2020**, *8*, 197.
- [52] G. M. Singh, P. D. Fortin, A. Koglin, C. T. Walsh, *Biochemistry.* **2008**, *47*, 11310.
- [53] Z. L. Reitz, C. D. Hardy, J. Suk, J. Bouvet, A. Butler, *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 19805.
- [54] T. M. Makris, M. Chakrabarti, E. Münck, J. D. Lipscomb, *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 15391.
- [55] C. Bebrone, *Biochem. Pharmacol.* **2007**, *74*, 1686.
- [56] F. Baier, N. Tokuriki, *J. Mol. Biol.* **2014**, *426*, 2442.
- [57] A. J. Jasniewski, L. Que, *Chem. Rev.* **2018**, *118*, 2554.
- [58] A. J. Jasniewski, C. J. Knoot, J. D. Lipscomb, L. Que, *Biochemistry.* **2016**, *55*, 5818.
- [59] T. M. Makris, C. J. Knoot, C. M. Wilmot, J. D. Lipscomb, *Biochemistry.* **2013**, *52*, 6662.
- [60] M. Kaniusaite, R. J. A. Goode, R. B. Schittenhelm, T. M. Makris, M. J. Cryle, *ACS Chem. Biol.* **2019**, *14*, 2932.
- [61] W. E. Stites, *Chem. Rev.* **1997**, *97*, 1233.
- [62] M. Rohde in *Gram-Positive Pathogens* (Eds.: V. A. Fischetti, R. P. Novick, J. J. Ferretti, D. A. Portnoy, M. Braunstein, J. I. Rood), ASM Press, Washington, DC, USA, **2019**, pp. 3–18.
- [63] W. Vollmer, D. Blanot, M. A. de Pedro, *FEMS Microbiol. Rev.* **2008**, *32*, 149.
- [64] H. J. Rogers, *Bacteriol. Rev.* **1970**, *34*, 194.
- [65] W. Vollmer, *FEMS Microbiol. Rev.* **2008**, *32*, 287.
- [66] K. Schleifer, R. Plapp, O. Kandler, *Biochim. Biophys. Acta.* **1968**, *154*, 573.
- [67] T. Schneider, M. M. Senn, B. Berger-Bächi, A. Tossi, H.-G. Sahl, I. Wiedemann, *Mol. Microbiol.* **2004**, *53*, 675.
- [68] H. Barreteau, A. Kovac, A. Boniface, M. Sova, S. Gobec, D. Blanot, *FEMS Microbiol. Rev.* **2008**, *32*, 168.
- [69] A. Bouhss, M. Crouvoisier, D. Blanot, D. Mengin-Lecreulx, *J. Biol. Chem.* **2004**, *279*, 29974.
- [70] S. Ha, D. Walker, Y. Shi, S. Walker, *Protein Sci.* **2000**, *9*, 1045.
- [71] N. Ruiz, *Lipid. Insights.* **2015**, *8*, 21.
- [72] S. D. Workman, N. C. J. Strynadka, *J. Mol. Biol.* **2020**, *432*, 4964.
- [73] S. D. Liston, L. M. Willis, *Curr. Opin. Struct. Biol.* **2021**, *68*, 55.
- [74] R. Biswas, R. E. Martinez, N. Göhring, M. Schlag, M. Josten, G. Xia, F. Hegler, C. Gekeler, A.-K. Gleske, F. Götz et al., *PloS One.* **2012**, *7*, e41415.
- [75] G. S. Pettis, A. S. Mukerji, *Int. J. Mol. Sci.* **2020**, *21*.
- [76] H. I. Hussain, A. I. Aqib, M. N. Seleem, M. A. Shabbir, H. Hao, Z. Iqbal, M. F.-E.-A. Kulyar, T. Zaheer, K. Li, *Microb. Pathogen.* **2021**, *158*, 105040.
- [77] J. J. Malin, E. de Leeuw, *Infect. Drug. Resist.* **2019**, *12*, 2613.

- [78] V. M. D'Costa, C. E. King, L. Kalan, M. Morar, W. W. L. Sung, C. Schwarz, D. Froese, G. Zazula, F. Calmels, R. Debruyne et al., *Nature*. **2011**, *477*, 457.
- [79] P. J. Stogios, A. Savchenko, *Protein Sci.* **2020**, *29*, 654.
- [80] R. Leclercq, E. Derlot, J. Duval, P. Courvalin, *N. Engl. J. Med.* **1988**, *319*, 157.
- [81] F. Grein, T. Schneider, H.-G. Sahl, *J. Mol Biol.* **2019**, *431*, 3520.
- [82] L. L. Ling, T. Schneider, A. J. Peoples, A. L. Spoering, I. Engels, B. P. Conlon, A. Mueller, T. F. Schäberle, D. E. Hughes, S. Epstein et al., *Nature*. **2015**, *517*, 455.
- [83] R. Shukla, J. Medeiros-Silva, A. Parmar, B. J. A. Vermeulen, S. Das, A. L. Paioni, S. Jekhmane, J. Lorent, A. M. J. J. Bonvin, M. Baldus et al., *Nat. Commun.* **2020**, *11*, 2848.
- [84] P.-C. Wen, J. M. Vanegas, S. B. Rempe, E. Tajkhorshid, *Chem. Sci.* **2018**, *9*, 6997.
- [85] J. Shoji, H. Hino, T. Hattori, K. Hirooka, Y. Kimura, T. Yoshida, *J. Antibiot.* **1989**, *42*, 1460.
- [86] F. von Nussbaum, R. D. Süßmuth, *Angew. Chem. Int. Ed.* **2015**, *54*, 6684.
- [87] J. O'Sullivan, J. E. McCullough, A. A. Tymiak, D. R. Kirsch, W. H. Trejo, P. A. Principe, *J. Antibiot.* **1988**, *41*, 1740.
- [88] J. Shoji, H. Hino, K. Matsumoto, T. Hattori, T. Yoshida, S. Matsuura, E. Kondo, *J. Antibiot.* **1988**, *41*, 713.
- [89] W. Lee, K. Schaefer, Y. Qiao, V. Srisuknimit, H. Steinmetz, R. Müller, D. Kahne, S. Walker, *J. Am. Chem. Soc.* **2016**, *138*, 100.
- [90] A. Hunt, D. R. Kirsch, *Int. J. Pharm.* **2020**, *581*, 119251.
- [91] M. Fujioka, S. Koda, Y. Morimoto, K. Biemann, *J. Org. Chem.* **1988**, *53*, 2820.
- [92] L. Ma, W. Li, H. Wang, X. Kuang, Q. Li, Y. Wang, P. Xie, K. Koike, *J. Pharm. Biomed. Anal.* **2015**, *102*, 400.
- [93] M. Crüsemann, R. Reher, I. Schamari, A. O. Brachmann, T. Ohbayashi, M. Kuschak, D. Malfacini, A. Seidinger, M. Pinto-Carbó, R. Richarz et al., *Angew. Chem. Int. Ed.* **2018**, *57*, 836.
- [94] A. Carlier, L. Fehr, M. Pinto-Carbó, T. Schäberle, R. Reher, S. Dessein, G. König, L. Eberl, *Environ. Microbiol.* **2016**, *18*, 2507.
- [95] M. Taniguchi, K. Nagai, N. Arao, T. Kawasaki, T. Saito, Y. Moritani, J. Takasaki, K. Hayashi, S. Fujita, K. Suzuki et al., *J. Antibiot.* **2003**, *56*, 358.
- [96] W. Hanke, J. Patt, J. Alenfelder, J. H. Voss, M. M. Zdouc, S. Kehraus, J. B. Kim, G. V. Grujičić, V. Namasivayam, R. Reher et al., *J. Nat. Prod.* **2021**.
- [97] R. Schrage, A.-L. Schmitz, E. Gaffal, S. Annala, S. Kehraus, D. Wenzel, K. M. Büllsbach, T. Bald, A. Inoue, Y. Shinjo et al., *Nat. Commun.* **2015**, *6*, 10156.
- [98] J. Li, Y. Ge, J.-X. Huang, K. Strømgaard, X. Zhang, X.-F. Xiong, *J. Med. Chem.* **2020**, *63*, 5013.

- [99] G.-W. Chang, C.-C. Hsiao, Y.-M. Peng, F. A. Vieira Braga, N. A. M. Kragten, E. B. M. Remmerswaal, M. D. B. van de Garde, R. Straussberg, G. M. König, E. Kostenis et al., *Cell Rep.* **2016**, *15*, 1757.
- [100] J. K. Ebner, G. M. König, E. Kostenis, P. Siegert, K. Aktories, J. H. C. Orth, *Bone.* **2019**, *127*, 592.
- [101] M. Grundmann, N. Merten, D. Malfacini, A. Inoue, P. Preis, K. Simon, N. Rüttiger, N. Ziegler, T. Benkel, N. K. Schmitt et al., *Nat. Commun.* **2018**, *9*, 341.
- [102] T. Kawasaki, M. Taniguchi, Y. Moritani, T. Uemura, T. Shigenaga, H. Takamatsu, K. Hayashi, J. Takasaki, T. Saito, K. Nagai, *Thromb. Haemost.* **2005**, *94*, 184.
- [103] M. Matthey, R. Roberts, A. Seidinger, A. Simon, R. Schröder, M. Kuschak, S. Annala, G. M. König, C. E. Müller, I. P. Hall et al., *Sci. Transl. Med.* **2017**, *9*.
- [104] S. Annala, X. Feng, N. Shridhar, F. Eryilmaz, J. Patt, J. Yang, E. M. Pfeil, R. D. Cervantes-Villagrana, A. Inoue, F. Häberlein et al., *Sci. Signaling.* **2019**, *12*.
- [105] K. Klepac, A. Kilić, T. Gnad, L. M. Brown, B. Herrmann, A. Wilderman, A. Balkow, A. Glöde, K. Simon, M. E. Lidell et al., *Nat. Commun.* **2016**, *7*, 10895.
- [106] K. Machida, D. Arai, R. Katsumata, S. Otsuka, J. K. Yamashita, T. Ye, S. Tang, N. Fusetani, Y. Nakao, *Bioorg. Med. Chem.* **2018**, *26*, 3852.
- [107] S. Wang, Q. Fang, Z. Lu, Y. Gao, L. Trembleau, R. Ebel, J. H. Andersen, C. Philips, S. Law, H. Deng, *Angew. Chem. Int. Ed.* **2021**, *60*, 3229.
- [108] J. T. Staley, A. Konopka, *Annu. Rev. Microbiol.* **1985**, *39*, 321.
- [109] a) D. Nichols, N. Cahoon, E. M. Trakhtenberg, L. Pham, A. Mehta, A. Belanger, T. Kanigan, K. Lewis, S. S. Epstein, *Appl. Environ. Microbiol.* **2010**, *76*, 2445; b) K. Lewis, *Nature.* **2012**, *485*, 439.
- [110] S.-H. Yoon, S.-M. Ha, S. Kwon, J. Lim, Y. Kim, H. Seo, J. Chun, *Int. J. Syst. Evol. Microbiol.* **2017**, *67*, 1613.
- [111] a) D. H. Parks, M. Chuvochina, D. W. Waite, C. Rinke, A. Skarszewski, P.-A. Chaumeil, P. Hugenholtz, *Nat. Biotechnol.* **2018**, *36*, 996; b) K. Dixit, D. Davray, D. Chaudhari, P. Kadam, R. Kshirsagar, Y. Shouche, D. Dhotre, S. D. Saroj, *Bioinformation.* **2021**, *17*, 377.
- [112] K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema, T. Weber, *Nucleic Acids Res.* **2019**, *47*, W81-W87.
- [113] S. Panthee, H. Hamamoto, A. Paudel, K. Sekimizu, *Arch. Microbiol.* **2016**, *198*, 839.
- [114] B. O. Bachmann, J. Ravel, *Methods Enzymol.* **2009**, *458*, 181.
- [115] Y. Luo, H. Dong, M. Zhou, Y. Huang, H. Zhang, W. He, H. Sheng, L. An, *Int. J. Syst. Evol. Microbiol.* **2019**, *69*, 926.
- [116] K. A. J. Bozhüyük, F. Fleischhacker, A. Linck, F. Wesche, A. Tietze, C.-P. Niesert, H. B. Bode, *Nat. Chem.* **2018**, *10*, 275.

- [117] K. A. J. Bozhüyük, A. Linck, A. Tietze, J. Kranz, F. Wesche, S. Nowak, F. Fleischhacker, Y.-N. Shi, P. Grün, H. B. Bode, *Nat. Chem.* **2019**, *11*, 653.
- [118] D. S. Waugh, *Trends Biotechnol.* **2005**, *23*, 316.
- [119] A. Kato, K. Maki, T. Ebina, K. Kuwajima, K. Soda, Y. Kuroda, *Biopolymers.* **2007**, *85*, 12.
- [120] V. V. Phelan, Y. Du, J. A. McLean, B. O. Bachmann, *Chem. Biol.* **2009**, *16*, 473.
- [121] D. Kalb, G. Lackner, M. Rappe, D. Hoffmeister, *ChemBioChem* **2015**, *16*, 1426.
- [122] M. Crüsemann, *Studien zur Biosynthese des Hormaomycins*, Dissertationsschrift, Rheinische Friedrich-Wilhelms-Universität Bonn, **2012**.
- [123] A. Greule, T. Izoré, D. Iftime, J. Tailhades, M. Schoppet, Y. Zhao, M. Peschke, I. Ahmed, A. Kulik, M. Adamek et al., *Nat. Commun.* **2019**, *10*, 2613.
- [124] a) M. Lebendiker, T. Danieli, *FEBS Lett.* **2014**, *588*, 236; b) D. J. Leibly, T. N. Nguyen, L. T. Kao, S. N. Hewitt, L. K. Barrett, W. C. van Voorhis, *PloS One.* **2012**, *7*, e52482.
- [125] E. Eichhorn, J. R. van der Ploeg, M. A. Kertesz, T. Leisinger, *J. Biol. Chem.* **1997**, *272*, 23031.
- [126] S.-S. Gao, N. Naowarajna, R. Cheng, X. Liu, P. Liu, *Nat. Prod. Rep.* **2018**, *35*, 792.
- [127] J. P. Danaceau, E. E. Chambers, K. J. Fountain, *Bioanalysis.* **2012**, *4*, 783.
- [128] D. A. Kopp, G. T. Gassner, J. L. Blazyk, S. J. Lippard, *Biochemistry.* **2001**, *40*, 14932.
- [129] N. A. Bowden, J. P. M. Sanders, M. E. Bruins, *J. Chem. Eng. Data* **2018**, *63*, 488.
- [130] Z. Lin, J. O. Falkinham, K. A. Tawfik, P. Jeffs, B. Bray, G. Dubay, J. E. Cox, E. W. Schmidt, *J. Nat. Prod.* **2012**, *75*, 1518.
- [131] N. S. Punekar, *ENZYMES: Catalysis, Kinetics and Mechanisms*, Springer Singapore, Singapore, **2018**.
- [132] K.-B. Cho, H. Hirao, S. Shaik, W. Nam, *Chem. Soc. Rev.* **2016**, *45*, 1197.
- [133] E. Breitmaier, *Vom NMR-Spektrum zur Strukturformel organischer Verbindungen*, Wiley-VCH, s.l., **2012**.
- [134] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, G. R. Hutchison, *J. Cheminf.* **2012**, *4*, 17.
- [135] C. Rausch, I. Hoof, T. Weber, W. Wohlleben, D. H. Huson, *BMC Evol. Biol.* **2007**, *7*, 78.
- [136] K. Haslinger, M. Peschke, C. Brieke, E. Maximowitsch, M. J. Cryle, *Nature.* **2015**, *521*, 105.
- [137] K. Graupner, K. Scherlach, T. Bretschneider, G. Lackner, M. Roth, H. Gross, C. Hertweck, *Angew. Chem. Int. Ed.* **2012**, *51*, 13173.
- [138] T. Mori, J. K. B. Cahn, M. C. Wilson, R. A. Meoded, V. Wiebach, A. F. C. Martinez, E. J. N. Helfrich, A. Albersmeier, D. Wibberg, S. Dätwyler et al., *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 1718.
- [139] J. Yang, Y. Zhang, *Nucleic Acids Res.* **2015**, *43*, W174-81.

- [140] a) W. E. Kim, A. Patel, G. H. Hur, P. Tufar, M. G. Wuo, J. A. McCammon, M. D. Burkart, *ChemBioChem*. **2019**, *20*, 147; b) W.-H. Chen, K. Li, N. S. Guntaka, S. D. Bruner, *ACS Chem. Biol.* **2016**, *11*, 2293.
- [141] D. J. Wilson, C. C. Aldrich, *Anal. Biochem.* **2010**, *404*, 56.
- [142] T. Kittilä, A. Mollo, L. K. Charkoudian, M. J. Cryle, *Angew. Chem. Int. Ed.* **2016**, *55*, 9834.
- [143] B. G. Fox, J. Shanklin, C. Somerville, E. Münck, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 2486.
- [144] C. Hermes, G. M. König, M. Crüsemann, *Nat. Prod. Rep.* **2021**.
- [145] F. I. Kraas, V. Helmetag, M. Wittmann, M. Strieker, M. A. Marahiel, *Chem. Biol.* **2010**, *17*, 872.
- [146] F. Baldeweg, H. Kage, S. Schieferdecker, C. Allen, D. Hoffmeister, M. Nett, *Org. Lett.* **2017**, *19*, 4868.
- [147] M. Adamek, M. Alanjary, N. Ziemert, *Nat. Prod. Rep.* **2019**, *36*, 1295.
- [148] C. L. Meyerkord, H. Fu (Eds.) *Methods in Molecular Biology*, Vol. 1278, Springer New York, New York, NY, **2015**.
- [149] M. Taha, M.-J. Lee, *Phys. Chem. Chem. Phys.* **2010**, *12*, 12840.
- [150] A. Velazquez-Campoy, S. A. Leavitt, E. Freire, *Methods Mol. Biol.* **2004**, *261*, 35.
- [151] K. P. Murphy, S. J. Gill, *Thermochim. Acta* **1990**, *172*, 11.
- [152] E. Muñoz, J. Sabín, J. Rial, D. Pérez, E. Ennifar, P. Dumas, Á. Piñeiro, *Methods Mol. Biol.* **2019**, *1964*, 225.
- [153] M. W. Freyer, E. A. Lewis in *Methods in Cell Biology*, v. 84 (Eds.: H. W. Detrich, J. J. Correia), Elsevier, Amsterdam, s.l., **2008**, pp. 79–113.
- [154] T. H. Scheuermann, C. A. Brautigam, *Methods*. **2015**, *76*, 87.
- [155] J. C. D. Houtman, P. H. Brown, B. Bowden, H. Yamaguchi, E. Appella, L. E. Samelson, P. Schuck, *Protein Sci.* **2007**, *16*, 30.
- [156] C. A. Brautigam, *Methods Enzymol.* **2015**, *562*, 109.
- [157] B. Kokona, E. S. Winesett, A. N. von Krusenstiern, M. J. Cryle, R. Fairman, L. K. Charkoudian, *Anal. Biochem.* **2016**, *495*, 42.
- [158] M. D. Lefebre, M. A. Valvano, *Appl. Environ. Microbiol.* **2002**, *68*, 5956.
- [159] S. D. Soby, S. R. Gadagkar, C. Contreras, F. L. Caruso, *Int. J. Syst. Evol. Microbiol.* **2013**, *63*, 1840.
- [160] T. A. Strand, R. Lale, K. F. Degnes, M. Lando, S. Valla, *PloS One*. **2014**, *9*, e90372.
- [161] K. B. Handing, E. Niedzialkowska, I. G. Shabalin, M. L. Kuhn, H. Zheng, W. Minor, *Nat. Protoc.* **2018**, *13*, 1062.
- [162] A. McPherson, *Methods Mol. Biol.* **2017**, *1607*, 17.
- [163] F. Cesaratto, O. R. Burrone, G. Petris, *J. Biotechnol.* **2016**, *231*, 239.
- [164] S. Fekete, A. Beck, J.-L. Veuthey, D. Guillarme, *J. Pharm. Biomed. Anal.* **2014**, *101*, 161.

- [165] L. Holm, *Protein Sci.* **2020**, *29*, 128.
- [166] A. F. Neuwald, J. S. Liu, D. J. Lipman, C. E. Lawrence, *Nucleic Acids Res.* **1997**, *25*, 1665.
- [167] G. Garau, I. García-Sáez, C. Bebrone, C. Anne, P. Mercuri, M. Galleni, J.-M. Frère, O. Dideberg, *Antimicrob. Agents Chemother.* **2004**, *48*, 2347.
- [168] F. Krause, H. Seelert, *Curr. Protoc. Protein. Sci.* **2008**, *Chapter 14*, Unit 14.11.
- [169] H. Nakagawa, T. Saio, S. Ajito, M. Sugiyama, R. Inoue, M. Nagao, T. Tominaga, Y. Kawakita, *Biophys. J.* **2021**.
- [170] M. Torchala, I. H. Moal, R. A. G. Chaleil, J. Fernandez-Recio, P. A. Bates, *Bioinformatics.* **2013**, *29*, 807.
- [171] R. Nussinov, J. A. Papin, I. Vakser, *PLoS Comput. Biol.* **2017**, *13*, e1005290.
- [172] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli et al., *Nucleic Acids Res.* **2018**, *46*, W296-W303.
- [173] K. A. Porter, I. Desta, D. Kozakov, S. Vajda, *Curr. Opin. Struct. Biol.* **2019**, *55*, 1.
- [174] S. Vajda, C. Yueh, D. Beglov, T. Bohnuud, S. E. Mottarella, B. Xia, D. R. Hall, D. Kozakov, *Proteins.* **2017**, *85*, 435.
- [175] P. Chakrabarti, J. Janin, *Proteins.* **2002**, *47*, 334.
- [176] P. B. Crowley, A. Golovin, *Proteins.* **2005**, *59*, 231.
- [177] a) S. A. Kautsar, K. Blin, S. Shaw, T. Weber, M. H. Medema, *Nucleic Acids Res.* **2021**, *49*, D490-D497; b) J. C. Navarro-Muñoz, N. Selem-Mojica, M. W. MULLowney, S. A. Kautsar, J. H. Tryon, E. I. Parkinson, E. L. C. de Los Santos, M. Yeong, P. Cruz-Morales, S. Abubucker et al., *Nat. Chem. Biol.* **2020**, *16*, 60.
- [178] Y. Zong, F. Fang, K. J. Meyer, L. Wang, Z. Ni, H. Gao, K. Lewis, J. Zhang, Y. Rao, *Nat. Commun.* **2019**, *10*, 3268.
- [179] a) K. A. J. Bozhueyuek, J. Watzel, N. Abbood, H. B. Bode, *Angew. Chem. Int. Ed.* **2021**, *60*, 17531; b) H.-M. Huang, P. Stephan, H. Kries, *Cell Chem. Biol.* **2021**, *28*, 221-227.e7.
- [180] B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane, C. Khosla, *Science.* **2001**, *291*, 1790.
- [181] A. Hülya, S. Shuangjie, S. Francoise, H. Hanna, R. René, H. Sabrina, E. Manuel, M. Tobias M., S. Sabine, G. Tobias A. M., *Angew. Chem. Int. Ed. (submitted)*.
- [182] M. R. Green, J. Sambrook, *Cold. Spring. Harb. Protoc.* **2020**, *2020*, 101220.
- [183] S. C. Gill, P. H. von Hippel, *Anal. Biochem.* **1989**, *182*, 319.
- [184] K. Tamura, G. Stecher, S. Kumar, *Mol. Biol. Evol.* **2021**, *38*, 3022.
- [185] I. Letunic, P. Bork, *Nucleic Acids Res.* **2021**, *49*, W293-W296.
- [186] B. Ozden, A. Kryshafovich, E. Karaca, *Proteins.* **2021**.
- [187] D. Kozakov, D. R. Hall, B. Xia, K. A. Porter, D. Padhorny, C. Yueh, D. Beglov, S. Vajda, *Nat. Protoc.* **2017**, *12*, 255.
- [188] B. K. Scholz-Schroeder, J. D. Soule, D. C. Gross, *Mol. Plant-Microbe Interact.* **2003**, *16*, 271.

7 List of abbreviations

Abbreviation	Word/Phrase
A domain	Adenylation domain
ADAM	1-adamantylamin hydrochloride
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
BGC	Biosynthetic gene cluster
BLAST	Basic local alignment search tool
bp	Base pairs
C domain	Condensation domain
<i>C. vaccinii</i>	<i>Chromobacterium vaccinii</i> MWU205
C/E domain	Dual condensation/ epimerization domain
C ₅₅ P	Undecaprenyl monophosphate
C ₅₅ PP	Undecaprenyl pyrophosphate
CoA	Coenzyme A
C _{starter} domain	Starter condensation domain
CV	Column volume
Cy domain	Cyclization domain
CYP450	NRPS-related, <i>trans</i> acting cytochrome P450 hydroxylase
Da	Dalton
Dha	Dehydroalanine
d-iGlu	D-isoglutamic acid
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E domain	Epimerization domain
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Ethylendiaminetetraacetic acid
ESI	Electrospray ionization
EtOH	Ethanol
FA	Formic acid
Fe/ α KG	NRPS-related, <i>trans</i> acting iron- and α -ketoglutarate dependent hydroxylase
Fmoc-Cl	Fluorenylmethyloxycarbonyl chloride
FPLC	Fast protein liquid chromatography
FR	FR900359
GDP	Guanosin diphosphate
GlcNAc	<i>N</i> -acetylglucosamine
GPCR	G protein-coupled receptor
GTP	Guanosine triphosphate
Has	β -hydroxy-L-asparagine
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	Histidine
Hle	β -hydroxy-L-leucine
Hph	Phenylserine
Hty	β -hydroxy-L-tyrosine
IPTG	Isopropyl- β -D-thiogalactopyranoside
ITC	Isothermal titration calorimetry
KOH	Potassium hydroxide
<i>L. sp.</i>	<i>Lysobacter species</i>
LC-MS	Liquid chromatography coupled to mass spectrometry
L-PAPA	L- <i>p</i> -aminophenylalanine

List of abbreviations

LTA	Lipoteichonic acid
MALDI-TOF	Matrix assisted laser desorption ionization-time of flight
MBL	Metallo-beta-lactamase
<i>m</i> DAP	<i>Meso</i> -diaminopimelic acid
MeCN	Acetone
MeOH	Methanol
MLP	MbtH-like protein
MRSA	Methicillin resistant <i>Staphylococcus aureus</i>
MurNAc	<i>N</i> -acetylmuramic acid
MWCO	Molecular weight cut-off
NADH	Nicotinamide adenine dinucleotide
NHDM	NRPS-related, <i>trans</i> acting non-heme diiron monooxygenase
NiNTA	Nickel nitrilotriacetic acid
NMR	Nuclear magnetic resonance spectroscopy
NP	Natural product
<i>N</i> -Pp-Hle	(2 <i>S</i> ,3 <i>R</i>)-(N-propionyl)-3-hydroxyleucine
<i>N</i> -Pp-Leu	(2 <i>S</i>)-(N-propionyl)-leucine
NRP	Non-ribosomal peptide
NRPS	Non-ribosomal peptide synthetase
OD ₆₀₀	Optical density at $\lambda=600$ nm
PCR	Polymerase chain reaction
PGN	Peptidoglycan
PPant	Phosphopantetheinyl
PP _i	Inorganic pyrophosphate
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SIR	Single ion record
SOC	Super outgrowth broth + glucose (medium)
ssNMR	Solid-state nuclear magnetic resonance spectroscopy
T domain	Thiolation domain
TA	Teichonic acid
TAE	Tris-Acetate-EDTA buffer
TB	Terrific broth (medium)
TD domain	Thioreductase domain
TE domain	Thioesterase domain
UDP	Uridine diphosphate
UV/Vis	Ultraviolet/visible light
WTA	Wall teichonic acid

8 Appendix

8.1 Supplementary data

8.1.1 Sequence alignments

Module	Sequence Alignment	Identity to RmyA3 [%]
RmyA3	EQTLAGIWQTLLGVERVGRHDDFFALGGHSLQAVRLVAQVRTQLGAELGLTELFAQPSLSAVAQAIVR-	-
RmyA4	EQTLAGIWQTLLGVERVGRHDDFFALGGHSLQAVRLVAQVRTQLGAELGLTELFAQPSLSAVAQAIVR-	100.0
RmyB6	EQTLVGIWQTLLGVERVGRHDDFFALGGHSLQAVRLVAQVRTQLGAELGLTELFAQPSLSAVAQAIVR-	98.5
RmyB8	EVALAGIWQTLLGVERVGRHDDFFTLGGHSLQAVRLVAQVRTQLGAELGLTELFAQPSLSAVAQAIVR-	95.6
RmyB9	EVALAGIWQTLLGVERVGRHDDFFTLGGHSLQAVRLVAQVRTQLGAELGLTELFAQPSLSAVAQAIVR-	95.6
RmyA2	EQTLAGIWQTLLGVERVGRHDDFFALGGHSLQAVRLVTQVRVQLGAELGLTALFAQPSLSAVAQAIVR-	95.6
RmyB7	EQTLAGIWQTLLGVERVGRHDDFFALGGHSLQAVRLMSLV-EQAGWRADVSRLFLQPTLAGFSASIT--	67.2
RmyA5	ETMLAELWQDLLGVERVGRHDDFFELGGHSLAMSLMARM-DELGLSADVRVLFQPTLAGLAAEV---	54.6
RmyB10	ETRLASIWQALLGVETIGRHDDFFALGGNSLQAVRLIGLL-AKADCRVTLTQLLQHPNIASLAAVAERD	54.0
	* * . ** ***** .***** ***.** * * . *	

Figure 8.1: Multiple sequence alignment of T domains from the ralsolamycin BGC from *Ralstonia solanacearum* GMI1000 (Accession No. AL646053.1), performed with CLUSTALO by UniProt. Identities are indicated by (*) and similarities by (.). The T domains of modules three and four are exactly identical.

```

Cm1A ----MRYSLRQDI AVEPIAGWYGSYLLPPQTLARFVHNRFRNIVESYLDLDPQVHAAV 56
Tcp25 -MNEERLFLRPNTIIEPLVDRFYASMYATAPVTATMSLAFRWIPMLESYLQAPEWHYKGS 59
FrsH  MTVSDNVFLRSHTKIEPLIMRWYAWAHLVSPAQHALNIAFRHLPLMKSFVSPAVHEAAS 60
HynC  MTQKNFKSLAPGIRSIPLINSWFAMHYLVAPSTFGHYTR-YHLDLLESFVDDPSQHLESI 59
LycC  --MNTQSLSDAVRSVPLVNNWFAMHYLMSPLTLGLYTKHAHLAMLDSFIDDPQHLVSS 58
RmyC  -MSADELYLRGDI AVEPLVSGWYAWTHLISPATLAMNVVGRHLKIMASFVHAPKVHVAAV 59
      *      *:: :. : *      : : * : *
Cm1A  RQRRMHGGPWIHAHEHRD-AIEAWYRETAPRRERLDELFEAVRRLEEDILPRHHGCELDP 115
Tcp25  RDPKFRGGFVNIEDSRADEVRRALLAARRDRADMIRFAEIAEAEKIVREEATGYDLRP 119
FrsH  SNPMLGGPFLELKKSDAAVKALWQQTQQQAGRQIAFAEAL ELDRRLQQSETGLSLDH 120
HynC  KIPELIGGPFINYTGDPG-DMARFRDRTRERCGTQLRGADAINSMYQMVLAQAKGAGVPG 118
LycC  NTPPELLGGPFINYTGDPG-DMARFRDETMECALHLRCADAINDMYQMLFSQAKGSGVPA 117
RmyC  KNPKMLGGPFIDYAESRVAEVQGLIEQTRSEQARLIAFAEGVHQLNALLKRAATGAGLDD 119
      .: ** :. : : : : : : : : : : : * :
Cm1A  VYQELPAALAGRVEVYGRDNRTADYRFVEPLMYASEYDESQVRFVFPVTEADAREFAL 175
Tcp25  LYPKLAELAGLVEIAYDTSNA-ASLHFQEPLVYHSKAYTEQRQSVQLSVETGIERPFVM 178
FrsH  IYAELEPELQGLVEVSYDLHNN-PSLRLEIEELLYLEDWVDGAGQEIASFSLDKEEERAF 179
HynC  LYAQVDELIRHGVELSYDVCKQ-PVARIIEKVLVDGPLYDPSLQTCILEKATYEPTRFVL 177
LycC  LYSQVDELIRYGVDAIYDVSKQ-PGARFIESIFYNPLNAPLQTVVLEKASYEPRTFVL 176
RmyC  LYAQVDPDLRGVLEFYDARHQ-PTFRLEYSELLYRSAYYDRTAQSLLQHLVTVQNDHRPFVL 178
      * : : : : * : : * : : * : : * : : * :
Cm1A  TTPMLEYGPQELLVNVPLNSPLLDVAVFRGGLTGTEL-DDLAARFGLDGERAARFASYFEP 234
Tcp25  STPRLP-SPDVLELPIPFRRHPGLEQLFLSRIRPTTL-TALREALELGDSSQAAQLAGLLVP 236
FrsH  NTPRVD-APGRMVVPLPFADARFDL LSAASRLSSVSF-SQLADALEIPEDQRPAPFREYFTT 237
HynC  STPQIQRPASVELRLPFGDPLWNHSSGKHAIGE LLEMLRPHIDDPARDMPLLEGMFVD 237
LycC  STPQIKNDKTAVTLSLPGDPLWDEL YAGTRDPDGLAERLRPYMQEPEREIPLKGMLEP 236
RmyC  STPRLP-DATSVCVQLPFESVPLDRFFRAREYPTPV-SRLADELGTIAQESRALFRSFFTS 236
      ** : : : : * : : * : : * : : :
Cm1A  TPAASEAPAPASSEEDVLEYVGHACVFAHRGTTFLVDPVLSYSGYPGGAENRFTFADL 294
Tcp25  EPA-LAA--DRHIAAGARIRYWGHALMQTPDVAIMTDPFISADT---DATGRYTYNDL 290
FrsH  SAPQRNE--PEYEGDGVRRVYFGHACVLVQTAESVSLVDPFLTDWH--QPEQGRLTIFYDL 293
HynC  HPGEARA--DGLAPQVRRVYFGHACVLMECAGVSVLVDPLISYPG--ESTLDHFTFDDL 293
LycC  PKAAAAD--AEWTQDTPRIRYFGHACAMIQCAGVSIIDPLISYPN--EAEIPHFTFDDL 292
RmyC  EPP-AAR--QRYQGEQARIRFGHACLLETRAFSLITDPVVSFPY--RGANSRYTYEDL 291
      :. : **** : . :. : ** : : : : * : **
Cm1A  PERIDHLLITHNHQDHMLFETLLRIRHRVGRVLPKSTNASLVDPGLGGILRRLGFTDVV 354
Tcp25  PDRIDYVLI THGSHDLVPE TLLQLRGRVGTFFVPRTSRGNL CDP SLALYLKSF GFSAV- 349
FrsH  PDHIDYVFLTHNHQDHFSC EALLQLRGRIGHILVPRNNGN FADPSMKLTLKRLGFDNVI 353
HynC  PARIDYVLI THPHQDHVLE TLLRLRSQVEHV VGRSGGGHL PDVSLKMLNHC GF DKVI 353
LycC  PAHIDYVLI THPHQDHVVFETLLRLRHKVGVVGRAGGGGLADVSLKMLLESCGFENNV 352
RmyC  PPFIDYVLI THNHQDHVLE TLLQLRHMIGTVIVPRGGNGELQDPSLKLALQALGFKRVV 351
      * ** : : ** * ** * : ** : * : : * : : * : * ** *
Cm1A  EVDDLETLSCGSAEVALPFLGEHGLDIRSKTGWLIRFGERSVLFAADSTNISPTMYTK 414
Tcp25  EVDDFDEISFPGGKIAATPFFGEHADLDIRAKATYWINLGGKSIYVGADSSGLDPALYRG 409
FrsH  VMDEMAITLDPDGLVSLPSYGEHSDLSITSKHGLYSLKGRSFMFLADSDAKDRVLYRR 413
HynC  ELGEYETIEFAGGRIGAPFYGEHADLDIRSKLAFGVQMHDTHCLFFADSNPPMPEFYAP 413
LycC  ELSEYESVSFEGGRILGAPFYGEHADLDIRAKLVFAIEMNQSVNLFADSNPPAPEFYAP 412
RmyC  ELDELESLELPGGKLTGLPFLGEHADLGI RTKLCYHVAIGGWSTLFAADACIIDPQVYRH 411
      : : : : . . . : * ** * ** * * : : : : * : : * :
Cm1A  VAEVIGPVDTVFIMGESIGAAASWIYGPLYGEPLDRRTDQSRRLNGSNFPQAREIVDALE 474
Tcp25  VRRHLGRVDIAFLGMECDGAPLNWQYQAF LTKPIPNKMSDSRKMSGNAEQASAIIVTELG 469
FrsH  IIKQVGVKVDNLFIGMECDGAPL TWLYGYP LSNPIGRREDESRRLSGDCERAWRIVEECG 473
HynC  LKKLMRPVDCLFLGMECVGAPATWL YGPLLQKMLTRGEDQSRRLDGCDAKALEMHRFFD 473
LycC  LKAMFPKIDCLFLGMECVGAPATWL YGPLLQKMLTRGEDQSRRLDGCDSAKALAMHEYFN 472
RmyC  AHEIVGDIDVLYLGMCDGAPLSWLYGPLLSEPLTWEQDRSRTLSGSNFARARELVEVMFG 471
      . : * : : ** * ** * * : : : : * : : * : : * :
Cm1A  PDEVVYVYAMGLEPVMGVMVAVDYDESHPAIVSDLLVRRHVQDKGGTAERLHLRRTLRL-- 532
Tcp25  ADEVYIYAMGEESLGHVMATSYNEDSFQLRQIAEFETWCANNVKSGLLDRHEWHWSH 529
FrsH  CSQALVYAMGQESWFRFVVGLEYTPDKKQIVESDKFVDRCRQAGMAAQLHGCQTMLL-- 531
HynC  PERIFVYAMGAEPWLTHITSILYSEELPQFKEARVLESTVRSQGRHAEVLYGKRELLL-- 531
LycC  PERLFYVYAMGAEPWLTHITSILYSEETQFKEARVVEKALRERGREALLFGKMEPLL-- 530
RmyC  PSQVYVYAMGQEPWLNHIMAVKYTETSLPIVESNRLLDYCRERGIVAERLYCMKETFHAD 531
      .. : **** * * : . . * : : : . . * : *
Cm1A  -----
Tcp25  P----- 530
FrsH  -----
HynC  -----
LycC  -----
RmyC  AAPVEA 537

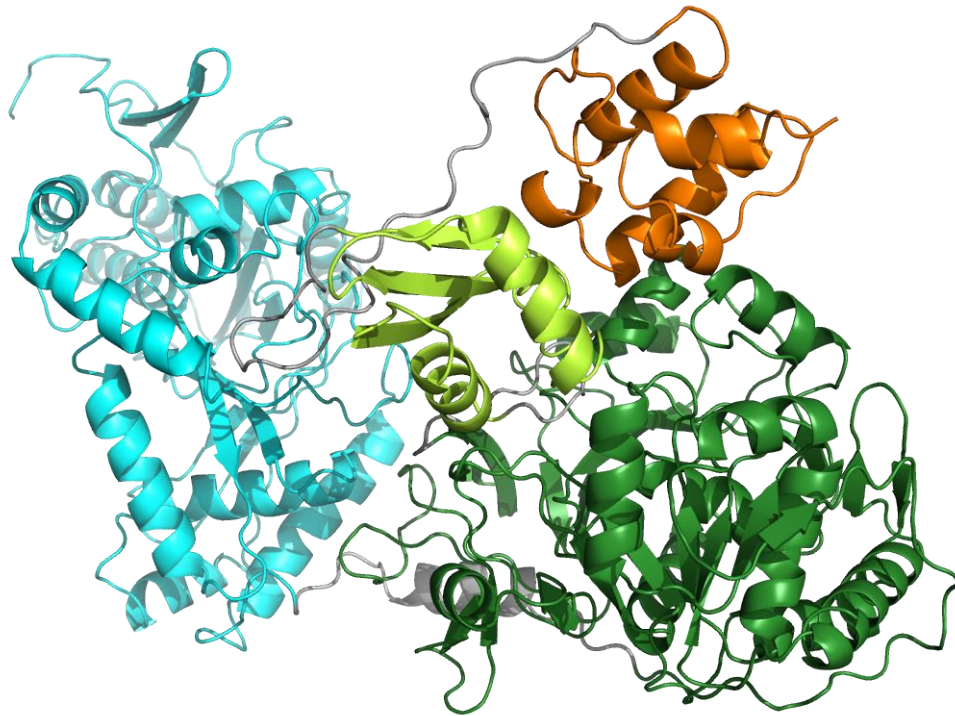
```

Figure 8.2: Multiple sequence alignment of *trans* acting NHD with relevance for this work. Alignment was generated with the CLUSTALO algorithm by the UniProt server. Identities make up 14.7% and are marked with (*), similarities make up additionally 22.8% and are marked with (: or (.) . Iron coordinating residues, which are crucial for MBL-fold metallo hydrolases are given in bold. Sequences were self assembled (FrsH: [*Chromobacterium vaccinii* MWU205] and HynC [*Lysobacter* sp. K5869]), or extracted from NCBI (Cm1A: 4J00_A [*Streptomyces venezuelae*]; Tcp25: CAE53366.1 [*Actinoplanes teichomyeticus*]; LycC: AEH59101.1 [*Lysobacter* sp. ATCC53042]; RmyC: CAD17790.1 [*Ralstonia solanacearum* GMI1000])

8.1.2 Sequence data for phylogenetic analysis of C domains

Table 8.1: Accession numbers and source organisms of protein sequences used for phylogenetic analysis of C domains (Figure 3.7). If possible, accession numbers from Rausch et al. and Reitz et al.^[53,135] were updated to NCBI nonredundant RefSeq (WP_).

Accession	Organism	Domain family	Ref.
WP_010924520.1	<i>Mycobacterium tuberculosis</i>	Cyclase	[135]
WP_010935625.1	<i>Corynebacterium diphtheriae</i>	Cyclase	[135]
WP_010949506.1	<i>Mycobacterium avium</i>	Cyclase	[135]
WP_002211388.1	<i>Yersinia pestis</i> CO92	Cyclase	[135]
WP_126987576.1	<i>Nostoc</i> sp. PCC 7120 = FACHB-418	Cyclase	[135]
CUV45544.1	<i>Ralstonia solanacearum</i>	Cyclase	[135]
WP_011030913.1	<i>Streptomyces coelicolor</i> A3(2)	Cyclase	[135]
AA80883.1	<i>Escherichia coli</i> CFT073	Cyclase	[135]
WP_011146562.1	<i>Photobacterium laumondii</i> subsp. <i>laumondii</i>	Cyclase	[135]
WP_011168195.1	<i>Pseudomonas savastanoi</i> pv. <i>phaseolicola</i>	Cyclase	[135]
WP_010895613.1	<i>Pseudomonas aeruginosa</i>	^D C _L	[135]
WP_010949130.1	<i>Mycobacterium avium</i> subsp. <i>paratuberculosis</i>	^D C _L	[135]
WP_011198677.1	<i>Bacillus cereus</i> E33L	^D C _L	[135]
WP_010949191.1	<i>Mycobacterium avium</i> subsp. <i>paratuberculosis</i>	^D C _L	[135]
WP_010954996.1	<i>Pseudomonas putida</i>	^D C _L	[135]
WP_011062374.1	<i>Pseudomonas protegens</i>	^D C _L	[135]
WP_011103865.1	<i>Pseudomonas syringae</i> pv. <i>tomato</i>	^D C _L	[135]
AA09419	<i>Bacillus cereus</i> ATCC 14579	^D C _L	[135]
AAZ34524.1	<i>Pseudomonas savastanoi</i> pv. <i>phaseolicola</i> 1448A	^D C _L	[135]
CAD17793.1	<i>Ralstonia solanacearum</i> GMI1000	Dual C/E	[135]
WP_011060446.1	<i>Pseudomonas protegens</i>	Dual C/E	[135]
AND87649.1	<i>Bradyrhizobium diazoefficiens</i> USDA 110	Dual C/E	[135]
WP_011093070.1	<i>Pectobacterium atrosepticum</i> SCRI1043	Dual C/E	[135]
WP_011105127.1	<i>Pseudomonas syringae</i> pv. <i>tomato</i>	Dual C/E	[135]
WP_011136349.1	<i>Chromobacterium violaceum</i>	Dual C/E	[135]
WP_011146892.1	<i>Photobacterium laumondii</i> subsp. <i>laumondii</i>	Dual C/E	[135]
WP_011204623.1	<i>Burkholderia mallei</i>	Dual C/E	[135]
WP_011205654.1	<i>Burkholderia pseudomallei</i> K96243	Dual C/E	[135]
WP_011104220.1	<i>Pseudomonas syringae</i> pv. <i>tomato</i>	Dual C/E	[135]
HynA ₃ C	<i>Lysobacter</i> sp. K5869	Dual C/E	[135]
HynA ₄ C	<i>Lysobacter</i> sp. K5869	Dual C/E	[135]
HynA ₆ C	<i>Lysobacter</i> sp. K5869	Dual C/E	[135]
WP_003916032.1	<i>Mycobacterium tuberculosis</i>	Epimerase	[135]
WP_010895613.1	<i>Pseudomonas aeruginosa</i>	Epimerase	[135]
WP_003412267.1	<i>Mycobacterium tuberculosis</i>	Epimerase	[135]
WP_010949130.1	<i>Mycobacterium avium</i> subsp. <i>paratuberculosis</i>	Epimerase	[135]
WP_010950704.1	<i>Mycobacterium tuberculosis</i> variant <i>bovis</i>	Epimerase	[135]
WP_010954975.1	<i>Pseudomonas putida</i>	Epimerase	[135]
NP_534179.1	<i>Agrobacterium tumefaciens</i> C58	Epimerase	[135]
WP_011062374.1	<i>Pseudomonas protegens</i>	Epimerase	[135]
AA09415.1	<i>Bacillus cereus</i> ATCC 14579	Epimerase	[135]
AAQ59905.1	<i>Chromobacterium violaceum</i> ATCC 12472	Epimerase	[135]
AA80537.1	<i>Streptomyces toyocaensis</i>	Glyco ^L C _L	[135]
Q93N88	<i>Streptomyces lavendulae</i>	Glyco ^L C _L	[135]
Q93N87	<i>Streptomyces lavendulae</i>	Glyco ^L C _L	[135]
Q8KLL4	<i>Streptomyces toyocaensis</i>	Glyco ^L C _L	[135]
AJF34464.1	<i>Eleutheria terrae</i>	C ₃ -epimerase	[82]
AEH59100.1	<i>Lysobacter</i> sp. ATCC 53042	C ₃ -epimerase	[35]
PRX87872.1	<i>Pseudomonas</i> sp. NFACC11-2	C ₃ -epimerase	[188]
CCJ67648.1	<i>Janthinobacterium agaricidamnosum</i>	C ₃ -epimerase	[137]
HynA ₅ C ₂	<i>Lysobacter</i> sp. K5869	C ₃ -epimerase	This work
CAQ71827.1	<i>Cupriavidus taiwanensis</i> LMG 19424	I domain	[53]
CAJ96471.1	<i>Cupriavidus necator</i> H16	I domain	[53]
WP_063365585.1	<i>Pseudoalteromonas luteoviolaceae</i>	I domain	[53]
WP_063365570.1	<i>Pseudoalteromonas luteoviolaceae</i>	I domain	[53]
WP_082236112.1	<i>Cupriavidus necator</i>	I domain	[53]
WP_052269209.1	<i>Alcanivorax pacificus</i>	I domain	[53]
WP_029293154.1	<i>Pseudomonas</i> sp. 06C 126	I domain	[53]
WP_053122086.1	<i>Pseudomonas thivervalensis</i>	I domain	[53]
WP_053122092.1	<i>Pseudomonas thivervalensis</i>	I domain	[53]
AJW67534.1	<i>Pseudomonas taiwanensis</i>	I domain	[53]
WP_011534377.1	<i>Pseudomonas entomophila</i>	I domain	[53]
WP_003400794.1	<i>Mycobacterium tuberculosis</i>	^L C _L	[135]
WP_010895611.1	<i>Pseudomonas aeruginosa</i>	^L C _L	[135]
WP_000605281.1	<i>Staphylococcus aureus</i> subsp. <i>aureus</i>	^L C _L	[135]
WP_010949130.1	<i>Mycobacterium avium</i> subsp. <i>paratuberculosis</i>	^L C _L	[135]
WP_010954974.1	<i>Pseudomonas putida</i>	^L C _L	[135]
WP_000605273.1	<i>Staphylococcus aureus</i> subsp. <i>aureus</i>	^L C _L	[135]
BAC70870.1	<i>Streptomyces avermitilis</i> MA-4680 = NBRC 14893	^L C _L	[135]
WP_010996798.1	<i>Nostoc</i> sp. PCC 7120 = FACHB-418	^L C _L	[135]
CAD17793.1	<i>Ralstonia solanacearum</i> GMI1000	^L C _L	[135]
WP_010973240.1	<i>Agrobacterium fabrum</i>	^L C _L	[135]
HynA ₂ C	<i>Lysobacter</i> sp. K5869	^L C _L	[135]
HynA ₅ C ₁	<i>Lysobacter</i> sp. K5869	^L C _L	[135]
HynB ₇ C	<i>Lysobacter</i> sp. K5869	^L C _L	[135]
HynB ₈ C	<i>Lysobacter</i> sp. K5869	^L C _L	[135]

8.1.3 *In silico* construction of FrsA₁A_{main}

C domain
 MKNSESP IHHFQASSAQLDVVISQEVSPNLPNNIAEYLNLAGSLDAGLFLQALSQVASESAELQYNFR

C domain
 HDGLQLTKFRRDDEGWEPDFIDVSTHGEPEHAALRAMRERVEKPFDLARDALFRWTLIRLADERHIFC

C domain
 HVYHHIAMDGAGYVMLLQRIAEVYGALREGQPAPACGFADADAIVREEERYRQSEQFAVDRAFVQARS

C domain
 AELATAEPPLPAADGPFLAFAQTAVI PEDACGGLRMTAERLGVSSRLLTAAIVAYFHRWGGQOEILF

C domain
 RLAVSARSDATRHAPGHLAHLPLLASLPPRASLADIARQLDGEVERMRPHTRYRAEDIVRDQAGAGL

C domain
 GRGAQGPVINLMPFAYRFEFGACRVESAHQLTVGVLDTLEVAVHDKNGDGLHLDLYASERGCPEPL

C domain | Linker C-A domain | A_{main}
 RRHALRLARFIVEAAAEPSQPVSDIELLDEAERRQLLVDWNRTGPDHGQATFPQLFETQAALTPHAVA

A_{main}
 LESP DARLSYAE LDARANRLARHLQSLGVGADVLVGICLERSIDMVAVL GALKSGAAYLPLSPEYPT

A_{main}
 ERLAYMLGDSMAPVLLTDSAQVERLPSYWGRVVELDRLLDALPDSAPERALRAEHLAYVIYTSGSTG

A_{main}
 QPKGVAVSHAGLAGLAGSQTERFALQGPTRVLQFASLSFDAAVMEMLMAFCSGGRLVLPAGPLLGEQ

A_{main}
 LLDTLNRHEISHALISPSALSTADAALAPVLRITLVVGGEACPGATVAAWSAGRRMVNAYGPTEATACV

A_{main}
 TMSEPLSGDGAPKLRPTHNARLYVLDGALQLAPVGVAGELYIAGAGLARGYLNRPGLTAEERFVANPY

A_{main} | Linker | A_{sub}
 GEGERLYRSGDLARWTEEGELEYLGRSDQVKVGRGFRIEPEGEIEAVLNRHPQVSQSVVVARQSQGGDS

A_{sub} | Linker A-T domain |
 QLVAYVAAVGGVEGSELRRRLAAGQLPEHMVPAAVVLES LPLPNGKLDKSLPAPEFGGSHYQRP RN

T domain
 AQEEMLCGLFAEVLDMEKVGRGDSFFDLGGHSLLATRLIRRIRETLDVELSIRDLEAPCVTELSRHI

AEGG

Figure 8.3: Upper: Structure of FrsA₁CAT, generated by iTASSER on basis of the tridomain NRPS EntF in thiolation state (PDB ID: 5T3D). The protein structure is shown in cartoon model. Cyan: C domain; dark and pale green: A_{main} and A_{sub}, domain; orange: T domain; grey: linker regions. The structure in dark green was extracted and used for in silico protein-protein docking to FrsH. **Lower:** Amino acid sequence of FrsA. The domain borders are indicated.

8.1.4 Side chain assembly with mutated enzymes

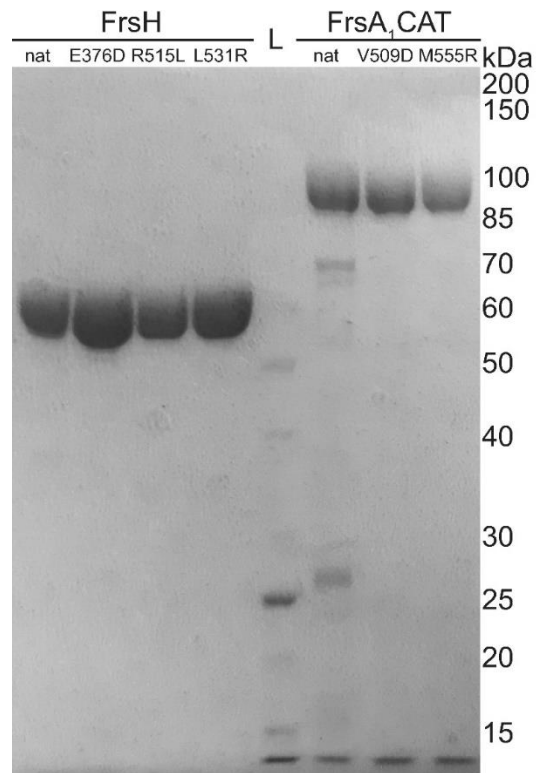


Figure 8.4: SDS-PAGE analysis of mutated and native FrsH and FrsA₁CAT, used to deplete the interaction of the enzymes in the side chain assembly assay.

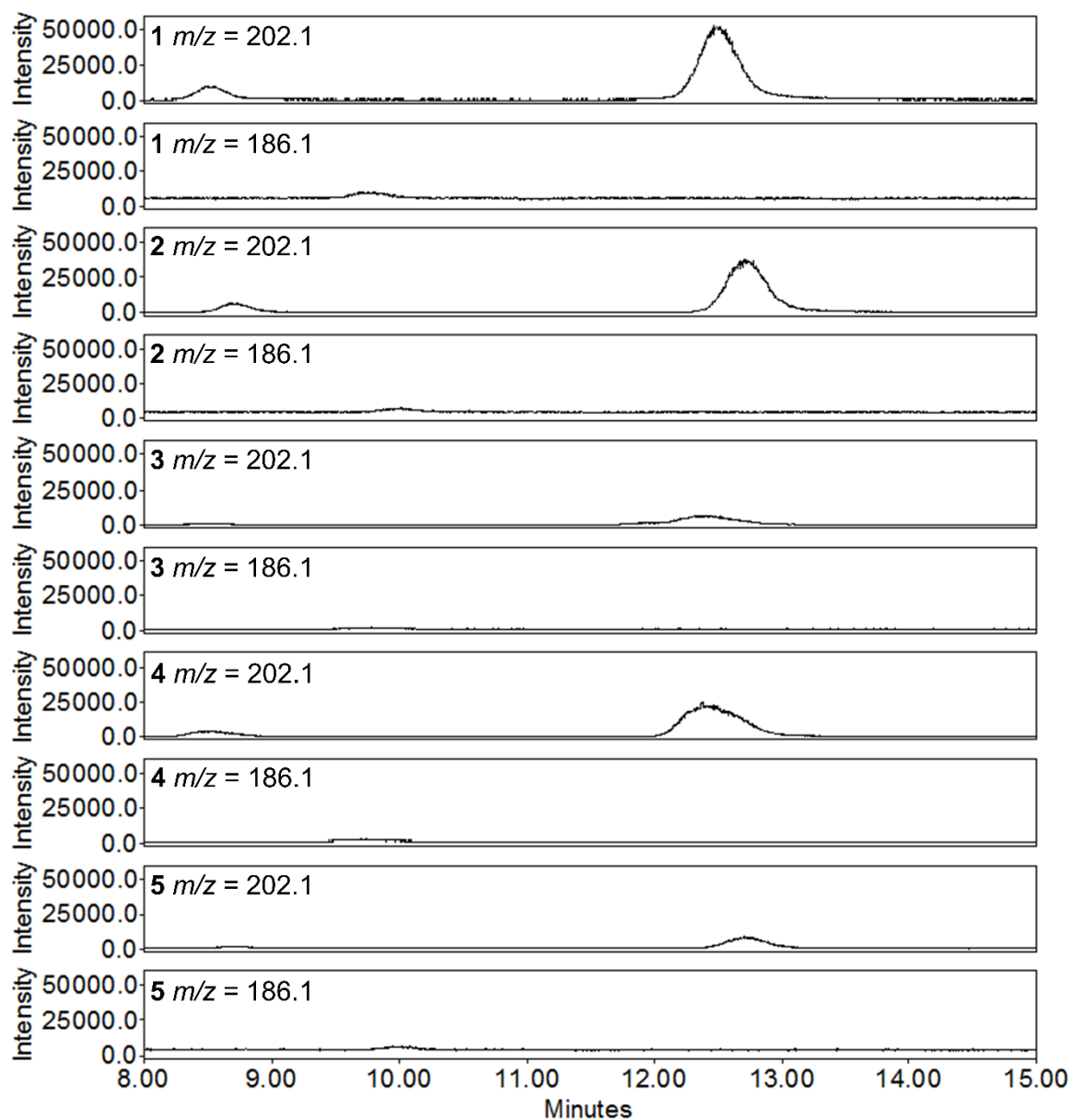


Figure 8.5: LC-MS SIR of the side chain assembly assay with non-mutated FrsA₁CAT and FrsH. The number of the respective replicate is given in bold. m/z (*N*-Pp-Hle) = 202.1, m/z (*N*-Pp-Leu) = 186.1. Minor changes in the retention time originate from non-exhaustive column equilibration.

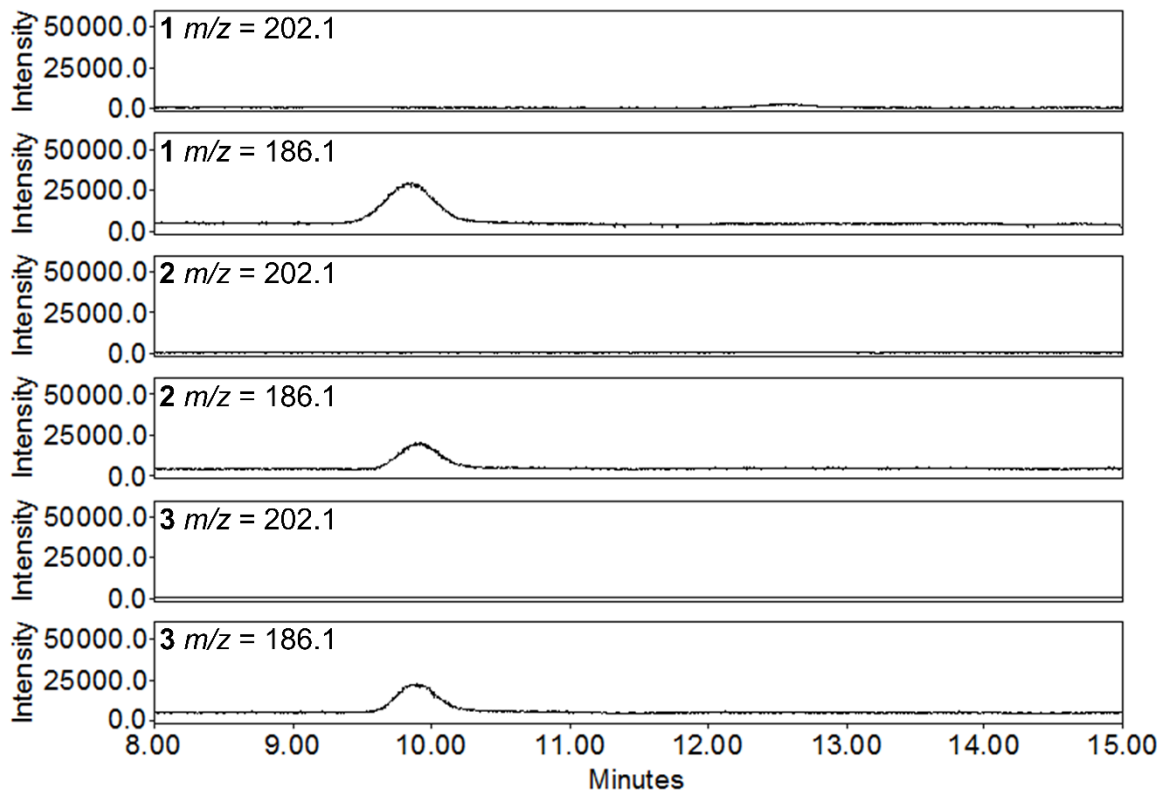


Figure 8.6: LC-MS SIR of the side chain assembly assay with FrsA₁CAT and FrsH E376D, which served as negative control. The number of the respective replicate is given in bold. m/z (*N*-Pp-Hle) = 202.1, m/z (*N*-Pp-Leu) = 186.1.

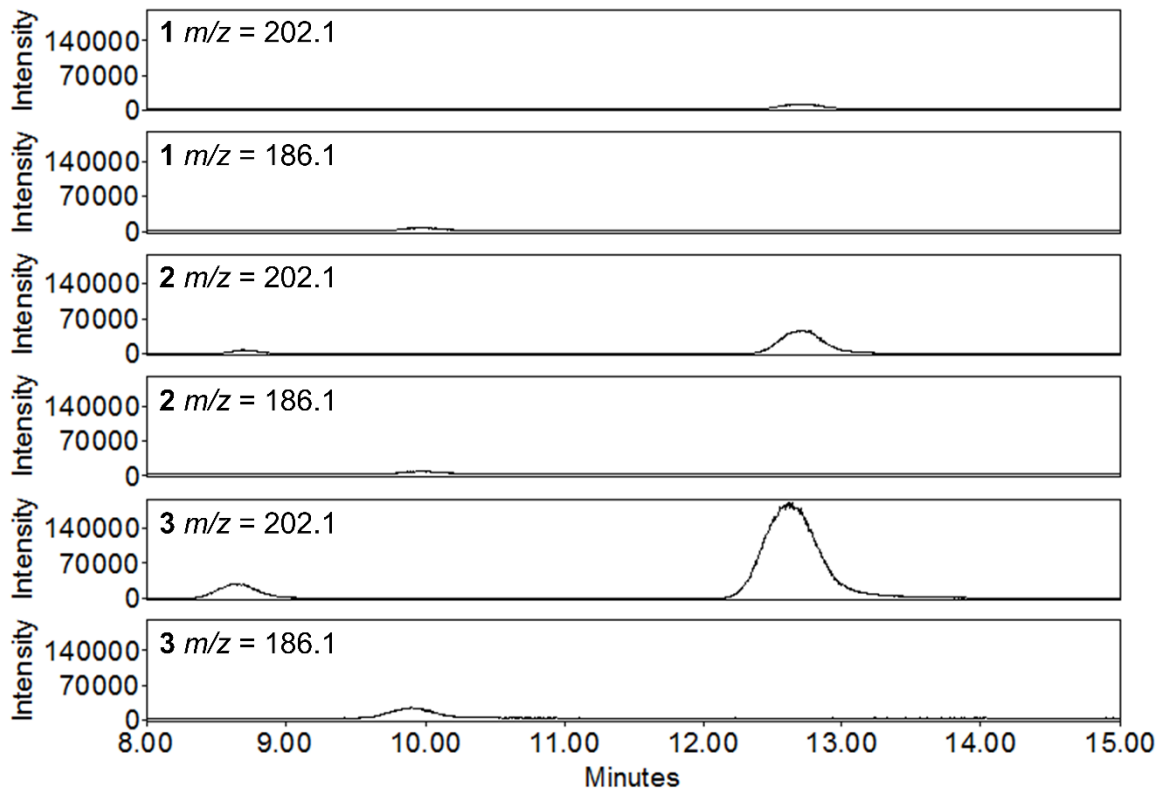


Figure 8.7: LC-MS SIR of the side chain assembly assay with FrsA₁CAT V509D and FrsH. The number of the respective replicate is given in bold. m/z (*N*-Pp-Hle) = 202.1, m/z (*N*-Pp-Leu) = 186.1.

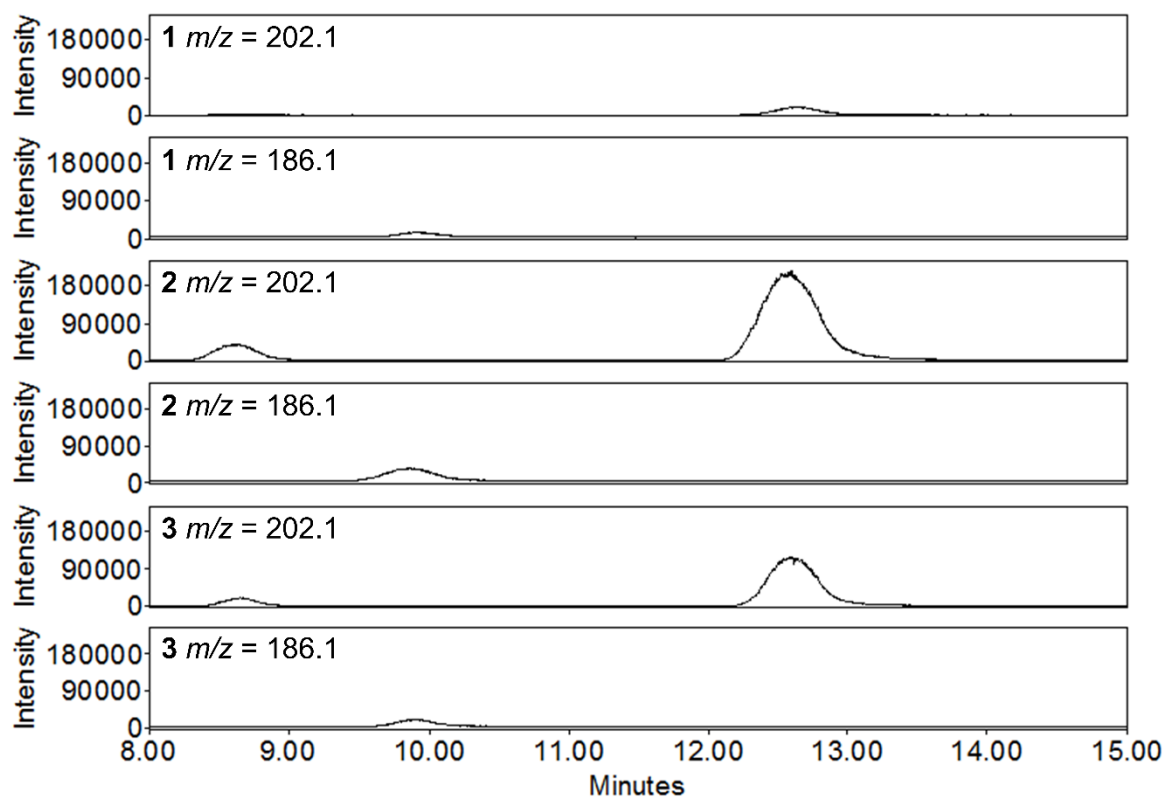


Figure 8.8: LC-MS SIR of the side chain assembly assay with FrsA₁CAT M555R and FrsH. The number of the respective replicate is given in bold. m/z (*N*-Pp-Hle) = 202.1, m/z (*N*-Pp-Leu) = 186.1.

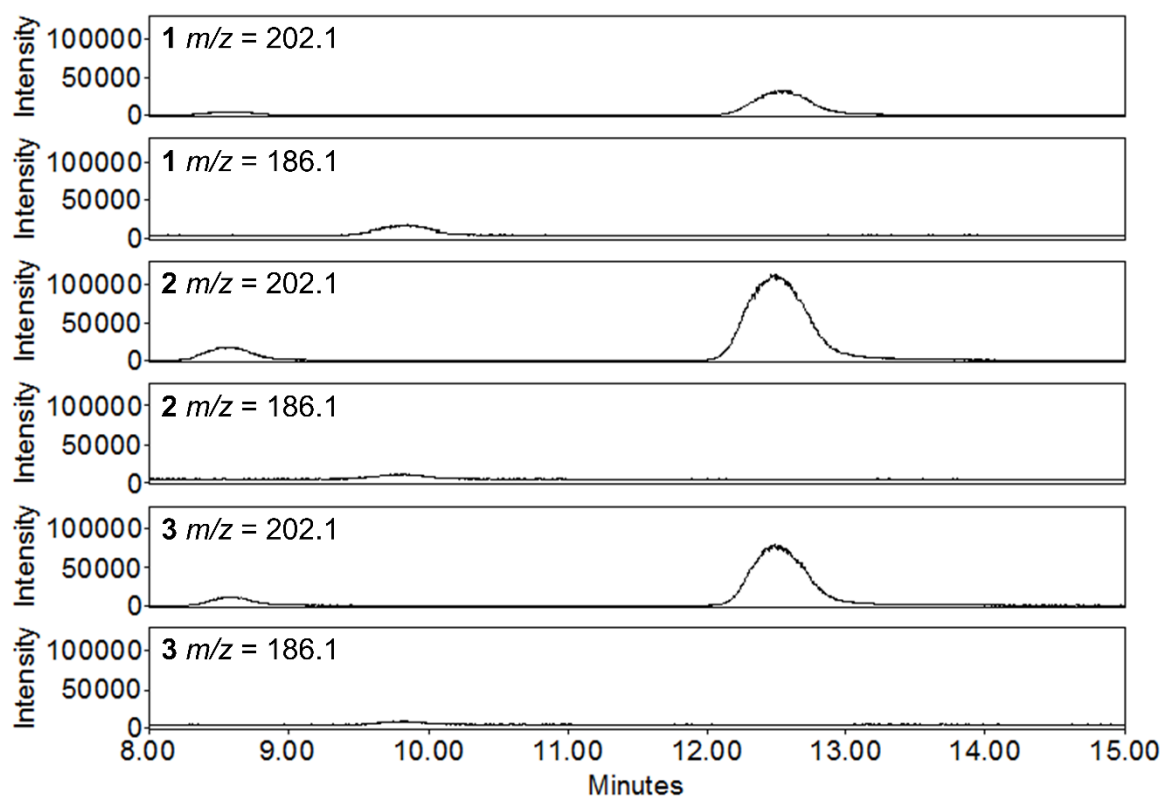


Figure 8.9: LC-MS SIR of the side chain assembly assay with FrsA₁CAT and FrsH R515L. The number of the respective replicate is given in bold. m/z (*N*-Pp-Hle) = 202.1, m/z (*N*-Pp-Leu) = 186.1.

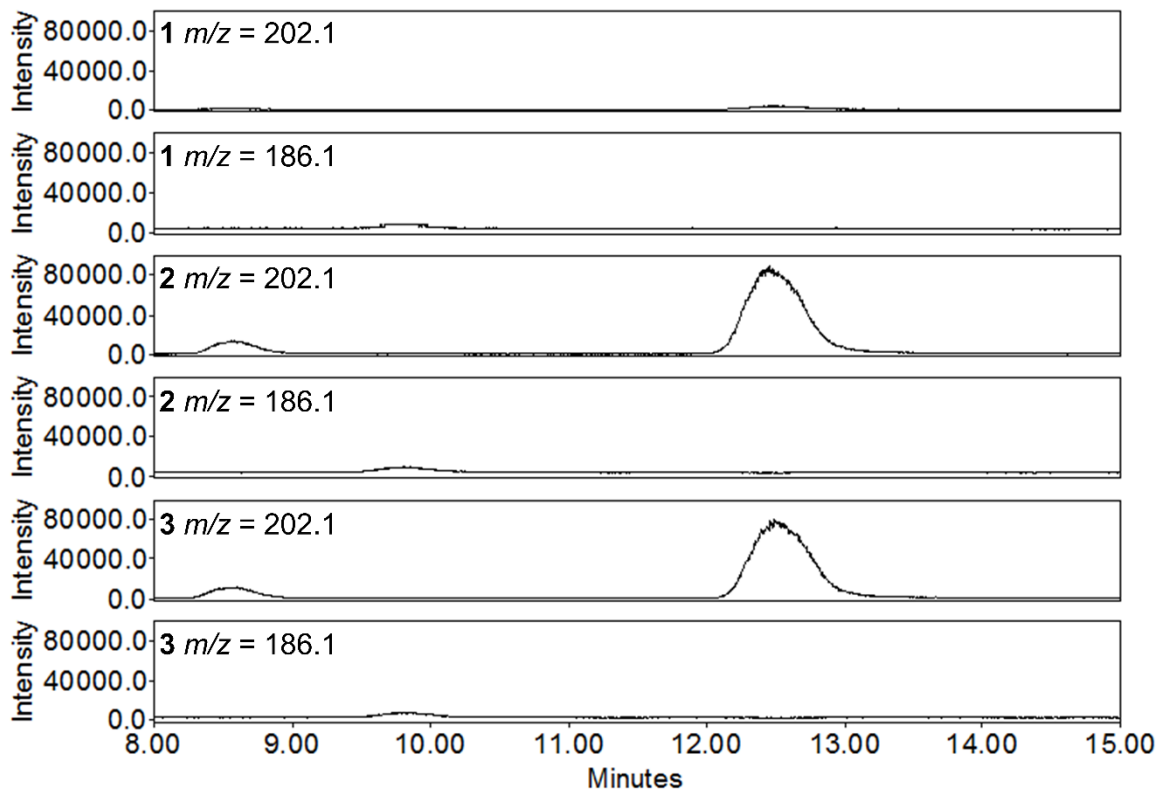


Figure 8.10: LC-MS SIR of the side chain assembly assay with FrsA₁CAT and FrsH L531R. The number of the respective replicate is given in bold. m/z (*N*-Pp-Hle) = 202.1, m/z (*N*-Pp-Leu) = 186.1.

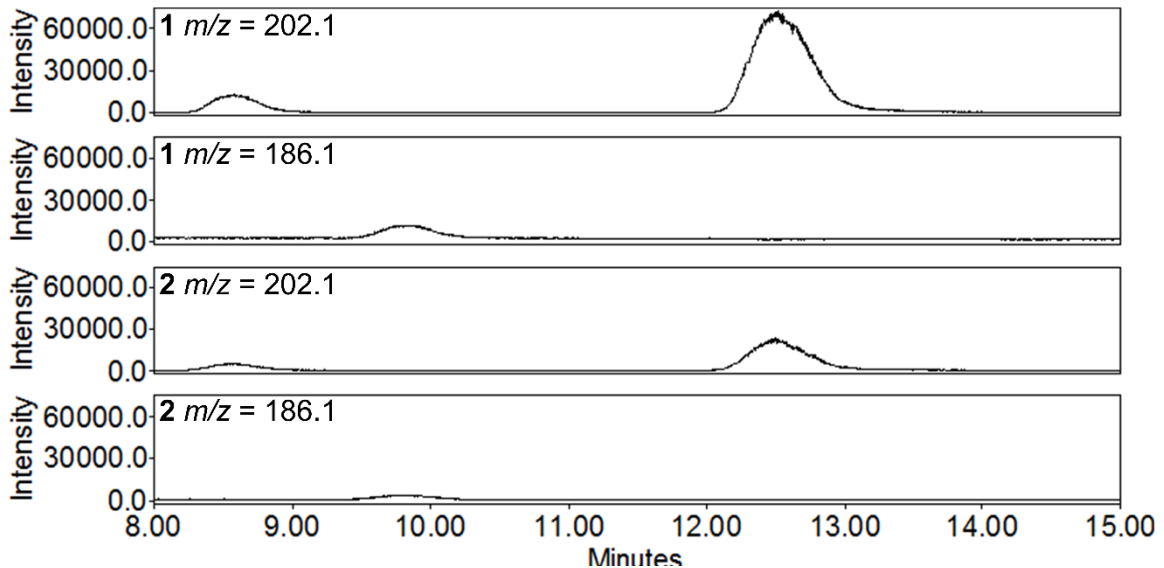


Figure 8.11: LC-MS SIR of the side chain assembly assay with FrsA₁CAT M555R and FrsH L531R. The number of the respective replicate is given in bold. m/z (*N*-Pp-Hle) = 202.1, m/z (*N*-Pp-Leu) = 186.1.

8.1.5 NRM data of hypeptin

All NMR data of hypeptin were measured and assigned by Dr. Stefan Kehraus.

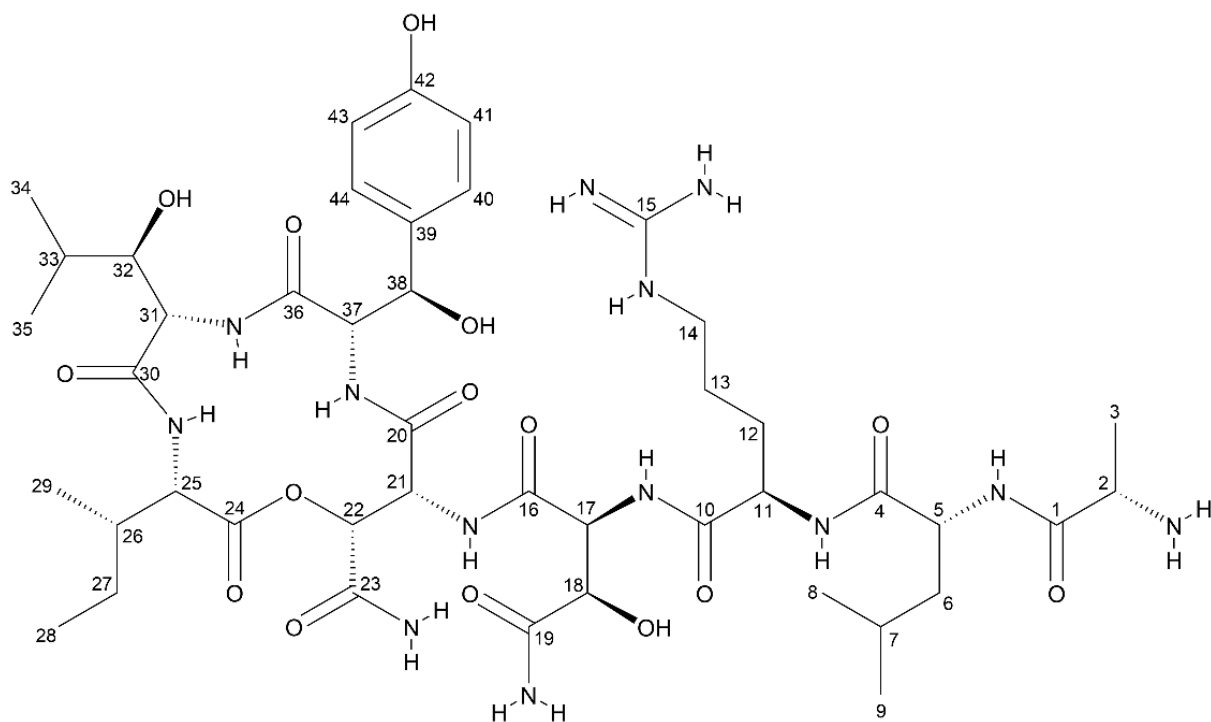


Figure 8.12: Chemical structure of hypeptin. Carbon atoms are numbered.

Appendix

Table 8.2: ^1H and ^{13}C NMR spectroscopic data of hypeptin in $\text{DMSO-}d_6$ at 40°C (^1H : 600 MHz; ^{13}C : 150 MHz). Atom count refers to Figure 8.12.

	C/H no.	δ_{H} (J in Hz) ^[a]	δ_{C} ^[a]	HMBC	ROESY
Ala	1		169.4, C		
	2	3.99, m	48.6, CH		3, NH ₂ -2, NH-5
	3	1.41, d (7.1)	16.9, CH ₃		2, NH ₂ -2, NH-5
	NH ₂ -2	8.10, m		1	2, 3
Leu	4		171.1, C		
	5	4.40, m	51.3, CH	1, 4	6a/b, 8, NH-5, NH-11
	6	a: 1.55, m b: 1.69, m	40.3, CH ₂		
	7	1.63, m	24.4, CH		
	8	0.89, d (6.4)	20.8, CH ₃		
	9	0.92, d (6.4)	23.0, CH ₃		
	NH-5	8.67, d (8.0)		1	2, 3, 5, NH-11, NH-17
	10	-	171.2, C		
	11	4.44, q (7.6)	55.2, CH	4, 10	12a/b, 13, 14, NH-11
Arg	12	a: 1.70, m b: 1.93, m	28.9, CH ₂		
	13	1.55, m	24.5, CH ₂		
	14	3.11, m	40.2, CH ₂		13, NH-15
	15		156.7, C		
	NH-11	8.11, m		4	5, 12a, 13, NH-5
	NH-15	7.61, brs			
	16	-	169.0, C		
	17	4.82, dd (3.5, 10.0)	56.3, CH	10, 16, 19	18, NH-17
	18	4.36, m	72.3, CH	16, 19	17
	19	-	173.5, C		
Has	OH-18	6.25, brs			
	NH-17	8.16, d (10.0)			11, 17
	NH ₂ -19	a: 7.37, brs b: 7.62, brs			NH-19b NH-19a
	20	-	167.6, C		
	21	5.13, dd (2.5, 10.0)	53.6, CH	16, 20	22, NH-21, NH ₂ -23, NH-37
	22	5.40, d (2.5)	72.7, CH	20, 23, 24	21, NH-21, NH ₂ -23
	23	-	168.6, C		
	NH-21	8.15, d (10.0)		16	21, NH-37
	NH ₂ -23	a: 7.18, brs b: 7.69, brs			NH-23b NH-23a
	24	-	168.7, C		
Ile	25	4.14, t (9.6)	56.6, CH	24, 30	26, 27a/b, 28, 29, NH-25
	26	1.95, m	36.0, CH		
	27	a: 1.18, m b: 1.55, m	24.4, CH ₂		
	28	0.80, t (7.5)	10.0, CH ₃		25, 26
	29	0.89, d (6.7)	14.9, CH ₃		25, 26
	NH-25	8.02, d (9.6)		30	25, 26, 32, NH-31
	30	-	169.6, C		
	31	4.34, dd (6.4, 10.4)	57.7, CH	30, 36	32,33,34,35, NH-25, NH-31
	32	3.58, m	75.1, CH		31, 33, 34, 35
Hle	33	1.60, m	30.4, CH		
	34	0.71, d (6.7)	19.3, CH ₃		31, 32, 33, 35
	35	0.92, d (6.4)	17.1, CH ₃		31, 32, 34
	OH-32	5.77, brs			
	NH-31	8.33, brs		36	31, 32, 33, 38, NH-25
	36	-	168.8, C		
	37	4.23, t (7.2)	63.3, CH	20, 36	38, 40, NH-37
	38	4.80, d (7.2)	71.6, CH	36	33, 35, 37, 40, 44
	39		130.8, C		
	40	7.12, d (8.2)	127.8, CH		33, 34, 35, 37, 38, 41
Hty	41	6.63, d (8.2)	114.4, CH		34, 40
	42		156.5, C		
	43	6.63, d (8.2)	114.4, CH		44
	44	7.12, d (8.2)	127.8, CH		43
	OH-38	6.25, brs			
	OH-42	9.20, brs			
	NH-37	8.32, d (7.2)		20	21, 37, 38, 44, NH-21

[a] assignments are based on extensive 1D and 2D NMR measurements (HMBC, HSQC, COSY) [b] OH resonance could not be assigned: 5.77, brs; 6.25, brs

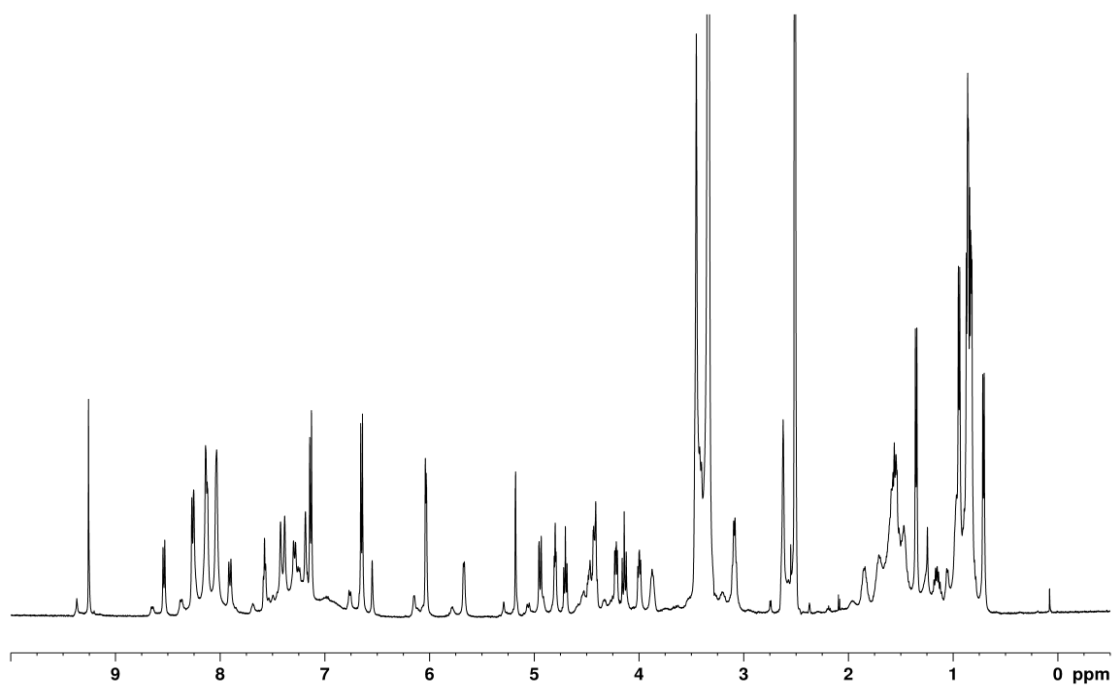


Figure 8.13: ^1H NMR spectrum of hypeptin in $\text{DMSO-}d_6$ (700 MHz).

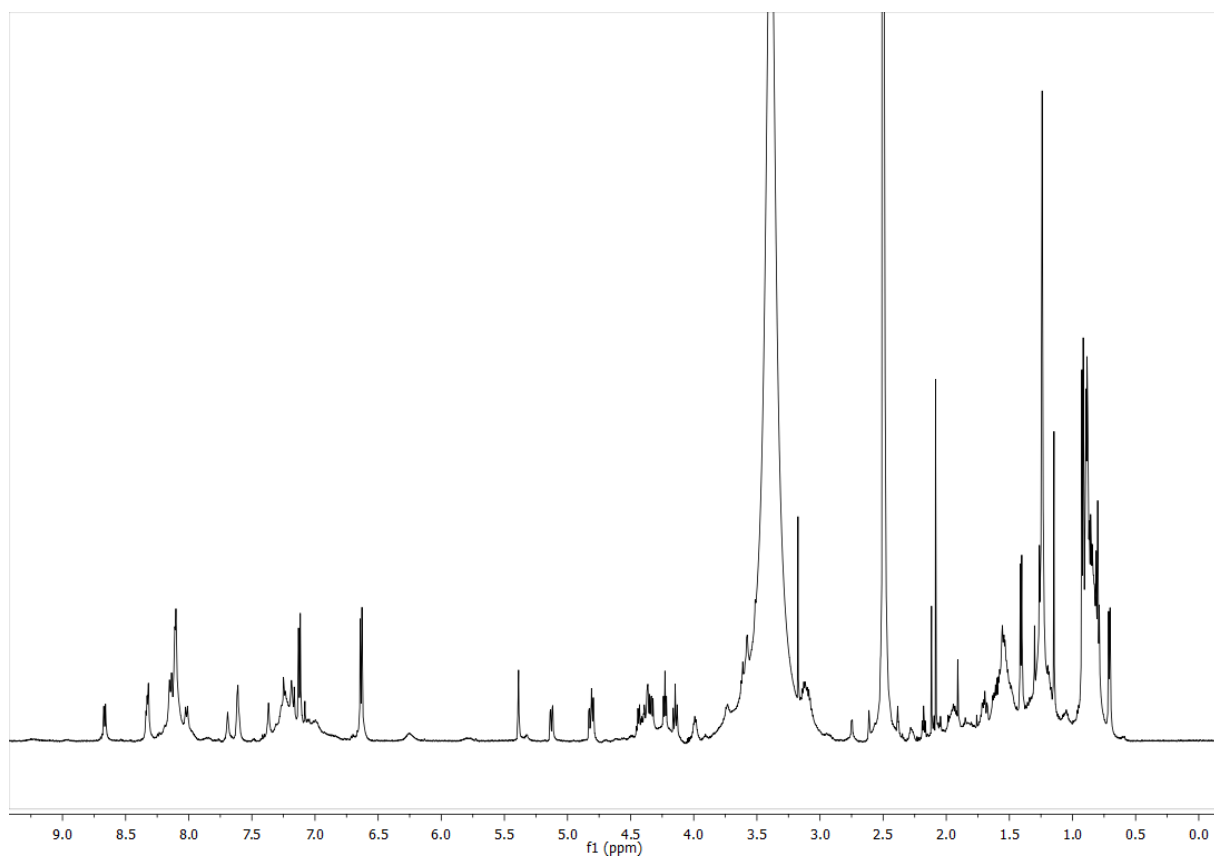


Figure 8.14: ^1H NMR spectrum of **1** in $\text{DMSO-}d_6$ (600 MHz, 40 °C).

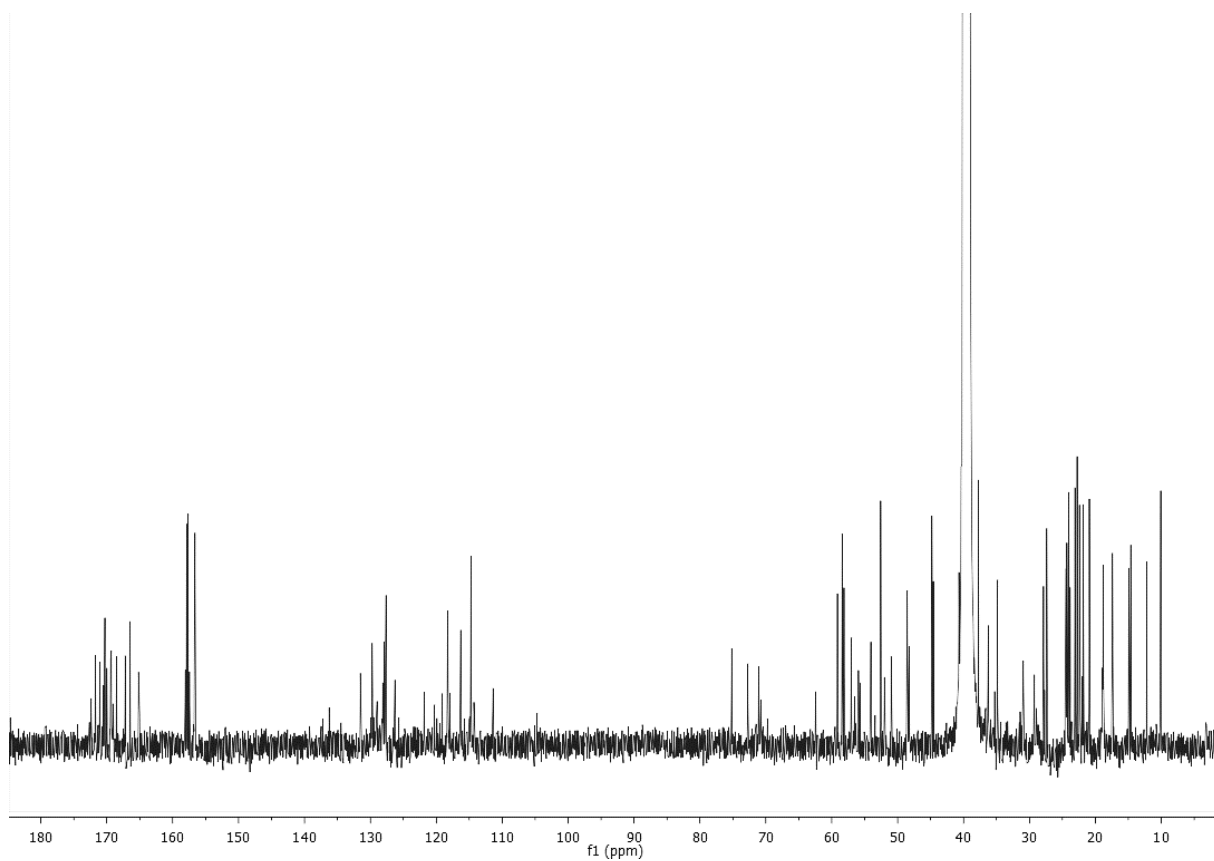


Figure 8.15: ^{13}C NMR spectrum of **1** in $\text{DMSO-}d_6$ (150 MHz).

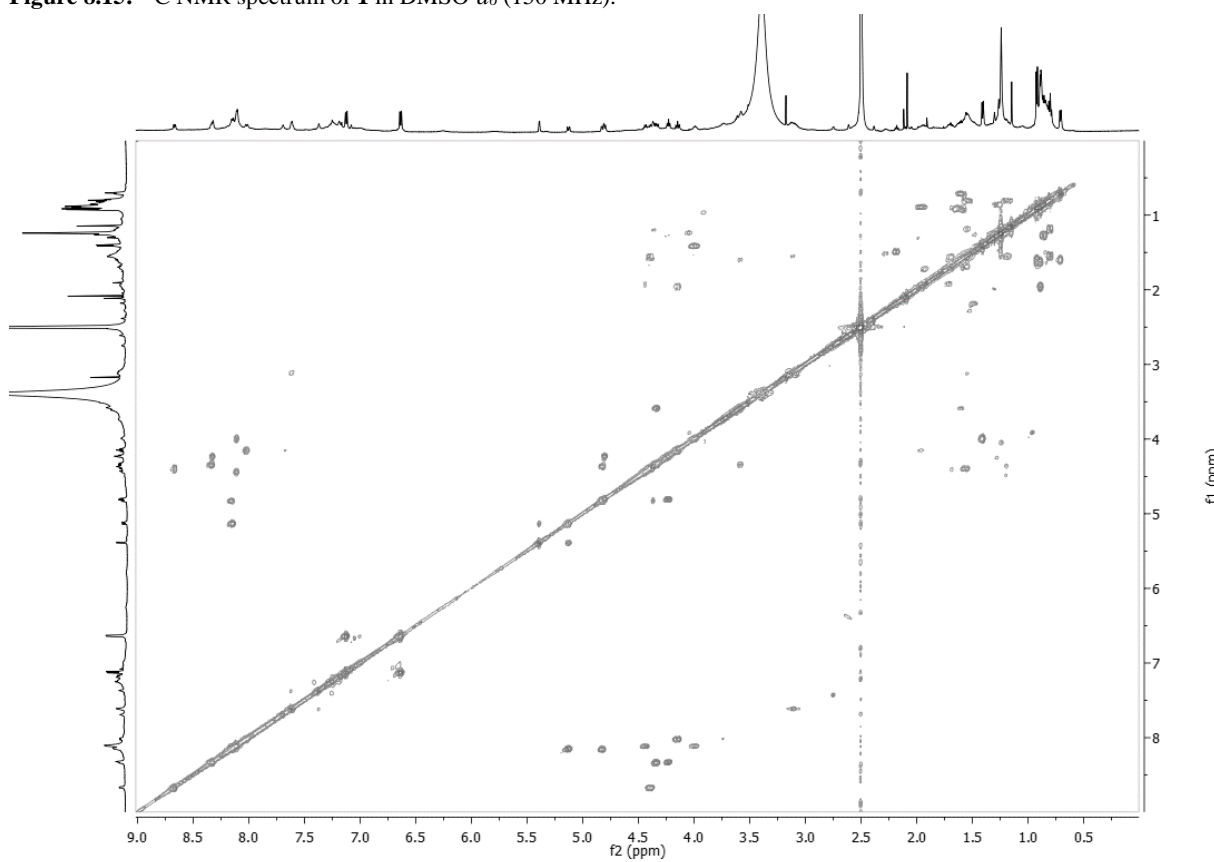


Figure 8.16: ^1H - ^1H COSY spectrum of **1** in $\text{DMSO-}d_6$ (600 MHz).

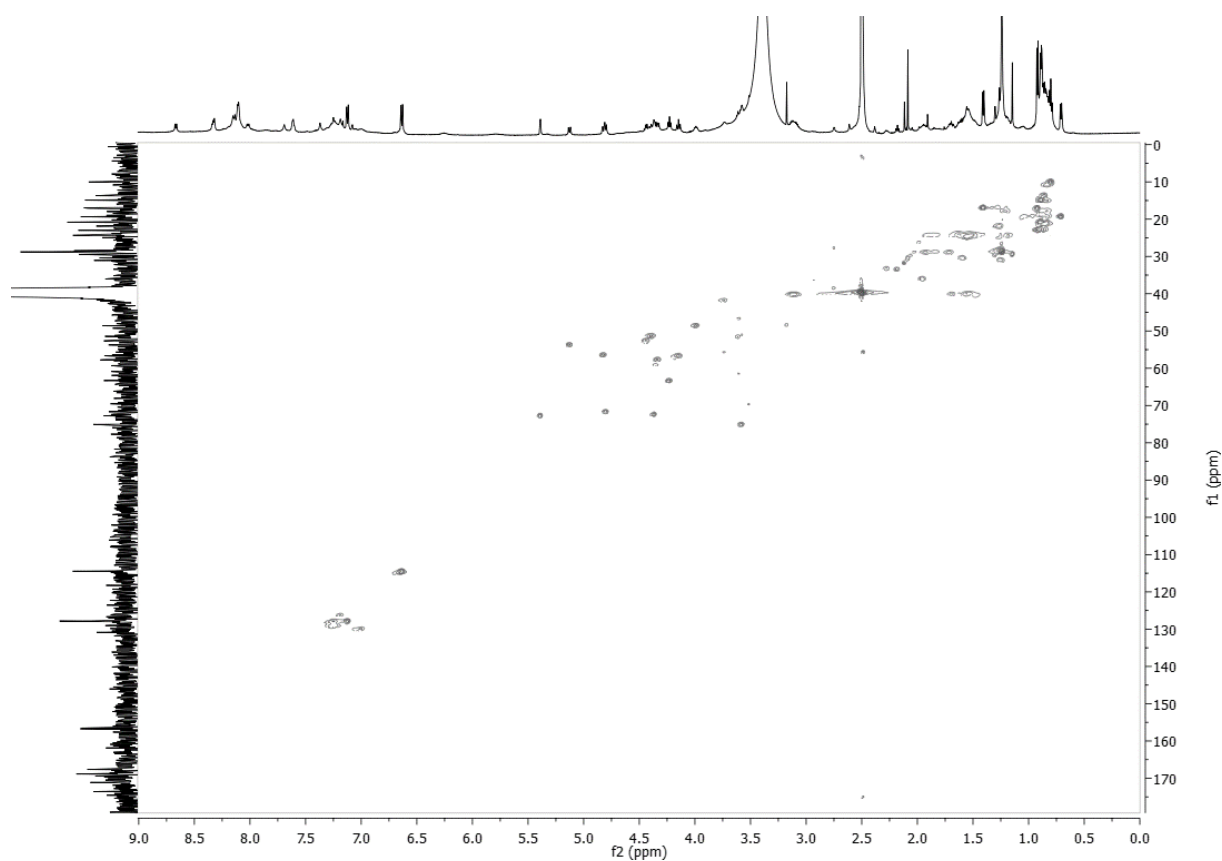


Figure 8.17: ^1H - ^{13}C HSQC spectrum of **1** in $\text{DMSO-}d_6$ (600 MHz).

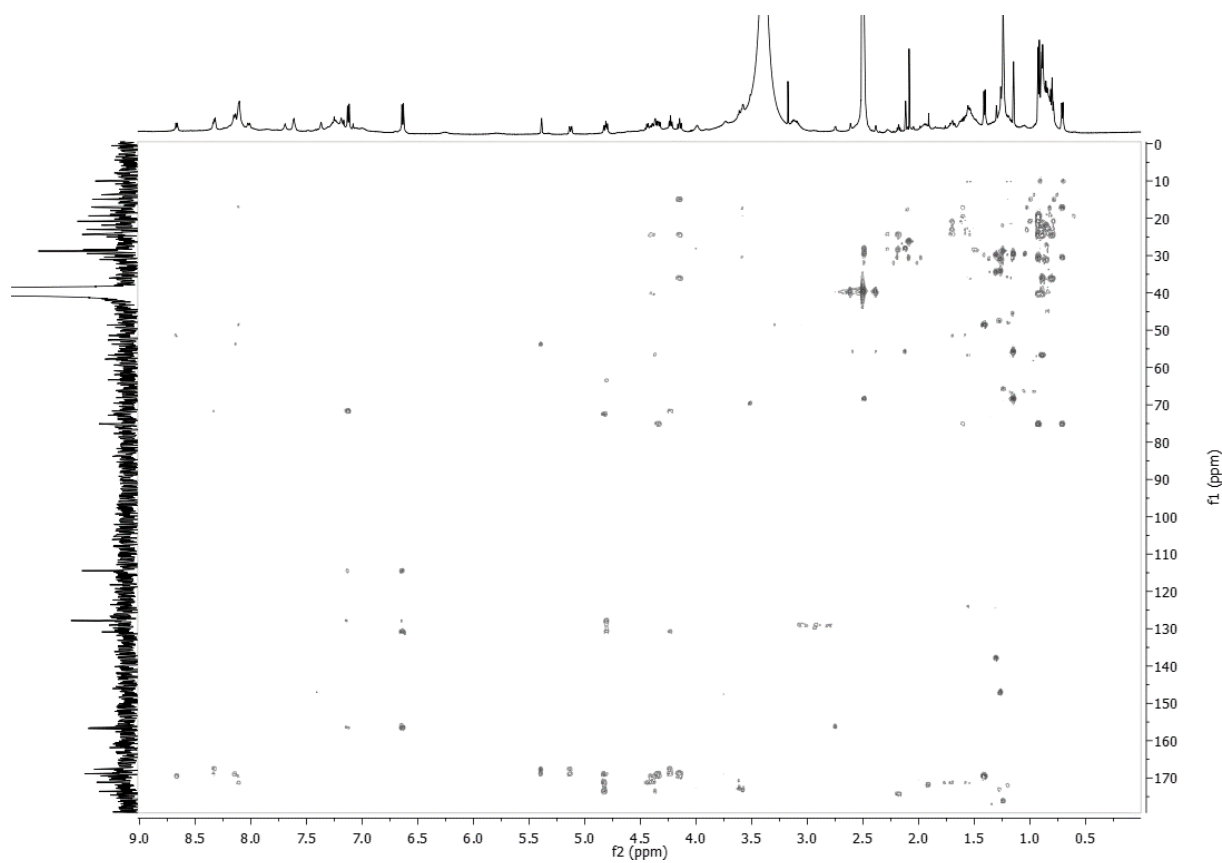


Figure 8.18: ^1H - ^{13}C HMBC spectrum of **1** in $\text{DMSO-}d_6$ (600 MHz).

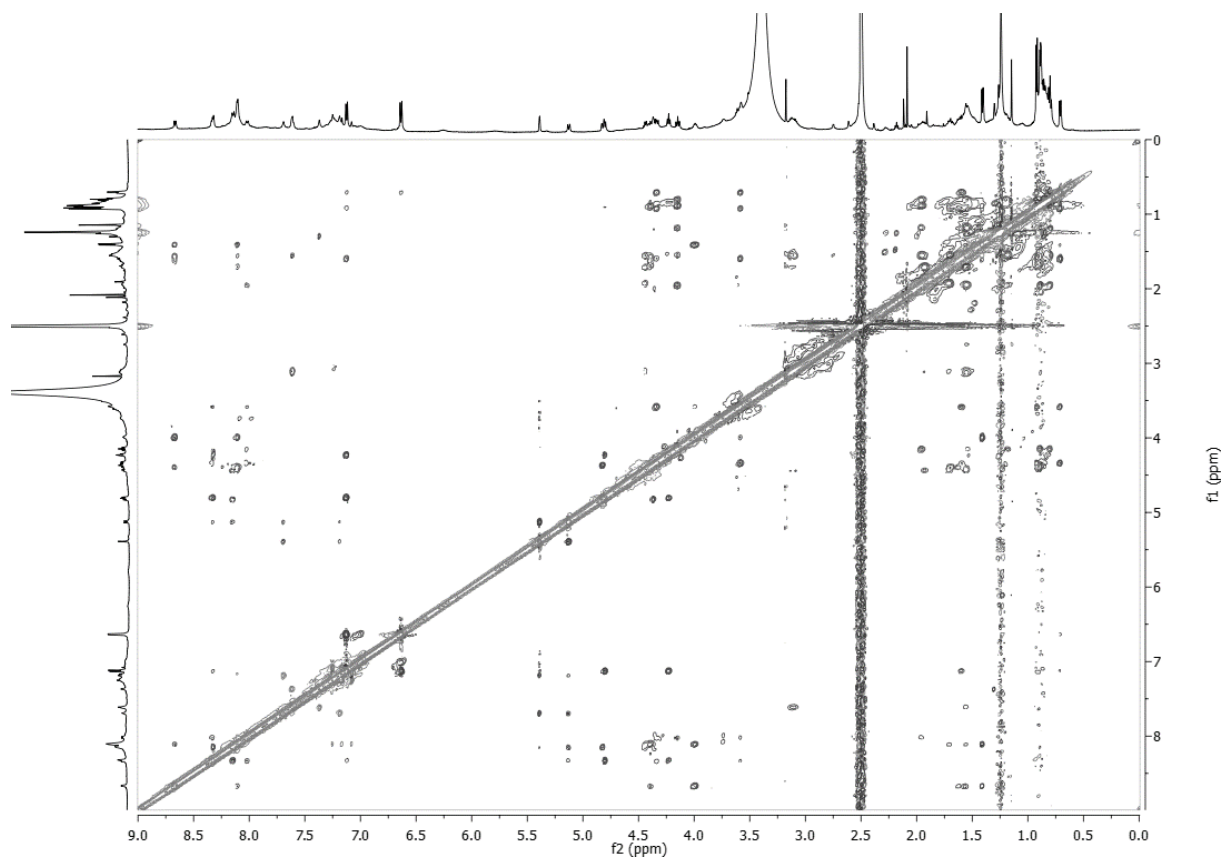


Figure 8.19: ^1H - ^1H ROESY spectrum of **1** in $\text{DMSO-}d_6$ (600 MHz).

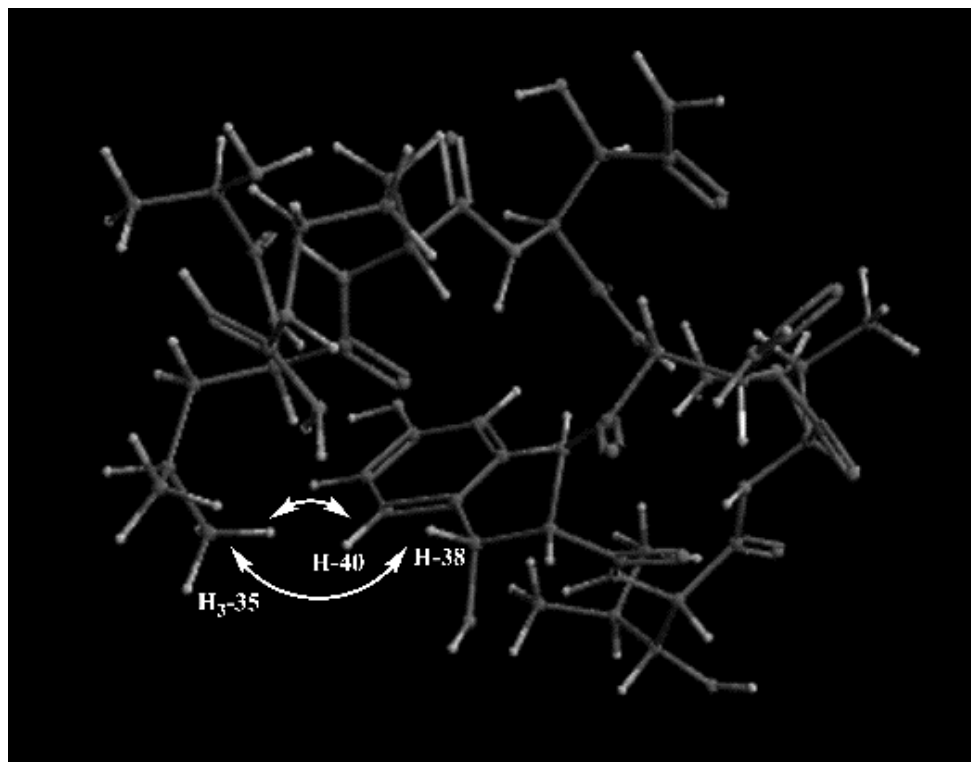


Figure 8.20: Conformation of hypeptin; white arrows indicate key ROESY correlations, supporting the (2*S*,3*R*) configuration of 3-Hydroxytyrosine, instead of (2*S*,3*S*). Figure provided by Dr. Stefan Kehraus.

8.2 Primer

8.2.1 Primers for the hypeptin project

Table 8.3: Primers for cloning of *hyn* genes. Restriction sites are underlined.

Name of the protein	Sequence (5'→3')	Template	Resulting Construct
HynA4AT	Fw: TATA <u>ACCATGG</u> AGCGCACGCGCCTGTGC Rv: TTATA <u>AAGCTTCG</u> CCGCGCCGAAGCTCGC	<i>Lysobacter</i> sp. K5869	pET28a: <i>hynA4AT</i>
HynA4CaccATCdon	Fw: TAT <u>CATATGG</u> TGAGTGCGCAAGAGCAC Rv: TAGA <u>AAGCTTAC</u> AGGCAGCGTCGTTGCCA	<i>Lysobacter</i> sp. K5869	pET28a: <i>hynA4CaccATCdon</i>
HynA5CAT	Fw: TGCC <u>CATATGG</u> CCTTGTTTCGAACCGGCG Rv: TATA <u>CTCGAGTTA</u> GCCGTTGGCCGGAACGAC	<i>Lysobacter</i> sp. K5869	pET28a: <i>hynA5CAT</i>
HynA6AT	Fw: TTAGGATCCGTGCTGTGCGCCTGGAAC Rv: TATA <u>AAGCTTTT</u> AGGCGCTGTGCGCCTGGAC	<i>Lysobacter</i> sp. K5869	pET28a: <i>hynA6AT</i>
HynA6CAT	Fw: TAT <u>CATATGC</u> AGATCGAACGCATCGTC Rv: TAGA <u>AAGCTTACT</u> GACAGTCGACGATGGC	<i>Lysobacter</i> sp. K5869	pET28a: <i>hynA6CAT</i>
HynB7AT	Fw: TGTCCATGGGACGCGAGCAAGTCCTG Rv: TATA <u>AAGCTTCT</u> GCCCGACCCGCGCTGCGAA	<i>Lysobacter</i> sp. K5869	pET28a: <i>hynB7AT</i>
HynC	Fw: AGCGGATCCATGACCCAGAAGAACTTCAAG Rv: TATA <u>AAGCTTTT</u> CACAGCAGCAGTTCCCT	<i>Lysobacter</i> sp. K5869	pET28a: <i>hynC</i>
HynC	Fw: CACGGGAATTCGACCCAGAAGAACTTCAA Rv: GAGA <u>AAGCTTTT</u> CACAGCAGCAGTTCCCT	<i>Lysobacter</i> sp. K5869	pCDFDuet-1: <i>hynC_hynMLP</i>
HynE	Fw: GCGCATATGATGAGCATGTTCAATCAGCTT Rv: TATA <u>AAGCTTTT</u> CATGCGCGCCCGTGCATCTG	<i>Lysobacter</i> sp. K5869	pET28a: <i>hynE</i>
HynMLP	Fw: GCGCATATGAGCAATCCCTTCGACGAC Rv: AATTTAATTAATTAGGCGCCGGCCGGGC	<i>Lysobacter</i> sp. K5869	pCDFDuet-1: <i>hynMLP</i>
HynTE	Fw: TAACATATGAACGCCAACGCCTCC Rv: TAGA <u>AAGCTTAG</u> ACATGGGCGTTTCCTTG	<i>Lysobacter</i> sp. K5869	pET28a: <i>hynTE</i>

Table 8.4: Primers for Gibson assembly of *hyn* genes. Mutated codons are underlined.

Name of the protein	Sequence (5'→3')	Template
HynA5CAT R201A	Fw: <u>GCG</u> CTCGTCGCCGACCGCCGTTTCGCTGCAATTG Rv: GTCGGCGACGAG <u>GCG</u> GTGGACGACGAGGCACAG	pET28a: <i>hynA5CAT</i>
HynA5CAT R201A D205A	Fw: <u>GCG</u> CTCGTCGCC <u>GAC</u> GCCGTTTCGCTGCAATTG Rv: <u>TGC</u> GGCGACGAG <u>GCG</u> GTGGACGACGAGGCACAG	pET28a: <i>hynA5CAT</i>
HynC E375D	Fw: TACGGCGAC <u>CA</u> CGCCGATCTCGACAT Rv: CGGCGTGGT <u>CG</u> CCGTAGAACGGC	pCDFDuet-1: <i>hynC_hynMLP</i>

8.2.2 Primers for the lysobactin project**Table 8.5:** Primers for cloning of *lyb* genes. Restriction sites are underlined.

Name of the protein	Sequence (5'→3')	Template	Resulting Construct
LybA ₁ AT	Fw: TCGCATATGATGAGAACGGACCACCAC	<i>Lysobacter</i> sp.	pET28a:
	Rv: TATA <u>AAGCTT</u> TTACGCCACTTCGCGCCAATC	ATCC53042	<i>lybA₁AT</i>
LybA ₃ A	Fw: TATCATATGGTGCGCGCCTGGAACCGC	<i>Lysobacter</i> sp.	pET28a: <i>lybA₃A</i>
	Rv: TGTA <u>AAGCTT</u> TTACTCGCGCTGCACGAAGGC	ATCC53042	
LybA ₃ AT	Fw: TATCATATGGTGCGCGCCTGGAACCGC	<i>Lysobacter</i> sp.	pET28a:
	Rv: TATA <u>AAGCTT</u> TTAGGCCTGGGCGACGCGCTC	ATCC53042	<i>lybA₃AT</i>
LybA ₃ ATC	Fw: TATCATATGGTGCGCGCCTGGAACCGC	<i>Lysobacter</i> sp.	pET28a:
	Rv: TGTA <u>AAGCTT</u> TTACTGGTACAGCGGCGTGTG	ATCC53042	<i>lybA₃ATC</i>
LybA ₃ C _{acc} ATC _{don}	Fw: TATCATATGGACGGCGACCTGCTGCGG	<i>Lysobacter</i> sp.	pET28a:
	Rv: TGTA <u>AAGCTT</u> TTAGAGCCACTGGCGCTGCCA	ATCC53042	<i>lybA₃C_{acc}ATC_{don}</i>
LybA ₄ A	Fw: CGCCATATGCTGGAACCTGGAACGCG	<i>Lysobacter</i> sp.	pET28a: <i>lybA₄A</i>
	Rv: TATA <u>AAGCTT</u> TTAGGCCTGCAGCGCTAGGC	ATCC53042	
LybA ₄ AT	Fw: CGCCATATGCTGGAACCTGGAACGCG	<i>Lysobacter</i> sp.	pET28a:
	Rv: CGCA <u>AAGCTT</u> TTAGTAGTCGACGATCACGCC	ATCC53042	<i>lybA₄AT</i>
LybA ₄ CAT	Fw: TATCATATGGCGCTGCCGCTGTCGTTT	<i>Lysobacter</i> sp.	pET28a:
	Rv: CGCA <u>AAGCTT</u> TTAGTAGTCGACGATCACGCC	ATCC53042	<i>lybA₄CAT</i>
LybB ₁₀ A	Fw: TGCCATATGCTGCACGAGTGAACGCC	<i>Lysobacter</i> sp.	pET28a:
	Rv: TATA <u>AAGCTT</u> TTAGCGCGGGCCGAAGGCATC	ATCC53042	<i>lybB₁₀A</i>
LybB ₁₀ AT	Fw: TGCCATATGCTGCACGAGTGAACGCC	<i>Lysobacter</i> sp.	pET28a:
	Rv: TATA <u>AAGCTT</u> TTACTGCCGCGCCGCGGCTTC	ATCC53042	<i>lybB₁₀AT</i>
LybC	Fw: GCGGATCCATGAATACTCACTCAGTCT	<i>Lysobacter</i> sp.	pET28a: <i>lybC</i>
	Rv: TAGCCTCGAGTCACAGAAGCAGTTCCAT	ATCC53042	

8.2.3 Primers for the FR900359 project

Table 8.6: Primers for cloning of *frs* genes. Restrictions sites are underlined.

Name of the ptotein	Sequence (5'→3')	Template	Resulting Construct
FrsA ₁ A	Fw: TAT <u>GGATCCT</u> CTGCTGGTCGACTGGAAC	<i>Chromobacterium</i>	pCDFDuet-1:
	Rv: GTC <u>AAGCTT</u> TTACCGCTGATAATGCGAGCC	<i>vaccinii</i> MWU205	<i>frsA₁A_frsB</i>
FrsA ₁ AT	Fw: TAT <u>GGATCCT</u> CTGCTGGTCGACTGGAAC	<i>Chromobacterium</i>	pET28a:
	Rv: TAT <u>AAGCTT</u> GCCTTCGGCGATATGCCGGGA	<i>vaccinii</i> MWU205	<i>frsA₁AT</i>
FrsA ₁ CAT	Fw: GAT <u>GGATCC</u> ATGAAAAACAGTGAATCGC	<i>Chromobacterium</i>	pET28a:
	Rv: TAT <u>AAGCTT</u> TCAGCTGTCGCCGCCTTCGGC	<i>vaccinii</i> MWU205	<i>frsA₁CAT</i>
FrsB	Fw: GCG <u>CATATG</u> AGCAATCCCTTTGATGAT	<i>Chromobacterium</i>	pCDFDuet-1:
	Rv: GCG <u>TTAATTA</u> ATTATTTATCATCGCACTCCAT	<i>vaccinii</i> MWU205	<i>frsB</i>
FrsH	Fw: GAT <u>GGATCC</u> ATGACCGTATCCGATAAC	<i>Chromobacterium</i>	pET28a: <i>frsH</i>
	Rv: GATA <u>CTCGAGT</u> TACAGCAGCATGGTTTG	<i>vaccinii</i> MWU205	
FrsH	Fw: GAT <u>GGATCC</u> ATGACCGTATCCGATAAC	<i>Chromobacterium</i>	pHis8-TEV:
	Rv: GATA <u>CTCGAGT</u> TACAGCAGCATGGTTTG	<i>vaccinii</i> MWU205	<i>frsH</i>

Table 8.7: Primers for Gibson assembly of *frs* genes. Mutated codons are underlined.

Name of the ptotein	Sequence (5'→3')	Template
FrsH E376D	Fw: TATGG <u>CGAT</u> CACTCCGACCTGAGCATC	pET28a:
	Rv: TCGGAGTG <u>ATC</u> GCCATAGGACGGCAGG	<i>frsH</i>
FrsH R515L	Fw: GTTTGTGCGACCGTTGCCTGCAAGCCGGCATGGCGGCG	pET28a:
	Rv: CCGCCATGCCGGCTTGCAGGCAACGGTCGACAACTT	<i>frsH</i>
FrsH L531R	Fw: GGCTGCCAAACCATG <u>CGC</u> CTGTAACCTCGAGCACCAC	pET28a:
	Rv: GGTGCTCGAGTTACAG <u>GCG</u> CATGGTTTGGCAGCCATG	<i>frsH</i>
FrsA ₁ CAT V509D	Fw: GGGCGTCGGCGCCGAC <u>GAT</u> CTGGTCGGCATCTGCCTG	pET28a:
	Rv: GGCAGATGCCGACCAGAT <u>TCG</u> TCGGCGCCGACGCCAG	<i>frsA₁CAT</i>
FrsA ₁ CAT M555R	Fw: CATGCTGGGCGACTCG <u>GCG</u> CCCCCGTGCTGCTGACC	pET28a:
	Rv: TCAGCAGCACGGGGG <u>GCG</u> CGAGTCGCCAGCATGTA	<i>frsA₁CAT</i>

Table 8.8: Primers for cloning of genes for complementation. Restrictions sites are underlined.

Name of the ptotein	Sequence (5'→3')	Template	Resulting Construct
HynC	Fw: TATTCATGACCCAGAAGA <u>ACTT</u> CAAG	<i>Lysobacter</i> sp.	pCv1: <i>hynC</i>
	Rv: TATA <u>AAGCTT</u> TCACAGCAGCAGTTCCCT	K5869	
LybC	Fw: AGATCATGAATACCACTCAGTCT	<i>Lysobacter</i> sp.	pCv1: <i>lybC</i>
	Rv: TATA <u>AAGCTT</u> TCACAGAAGCAGTTCCAT	ATCC53042	

8.3 Plasmid maps

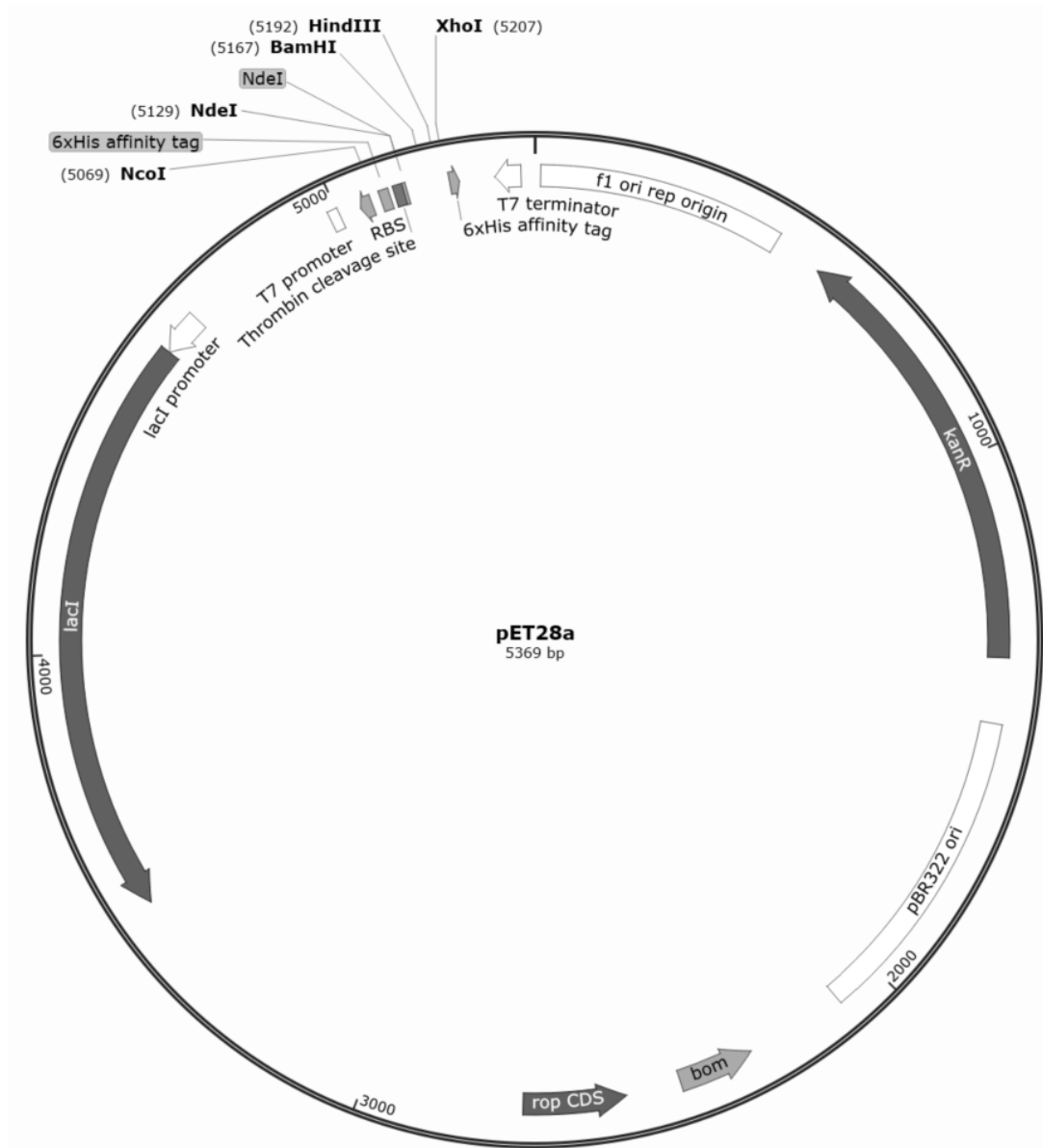


Figure 8.21: Plasmid map of pET28a. Utilized restriction sites are indicated.

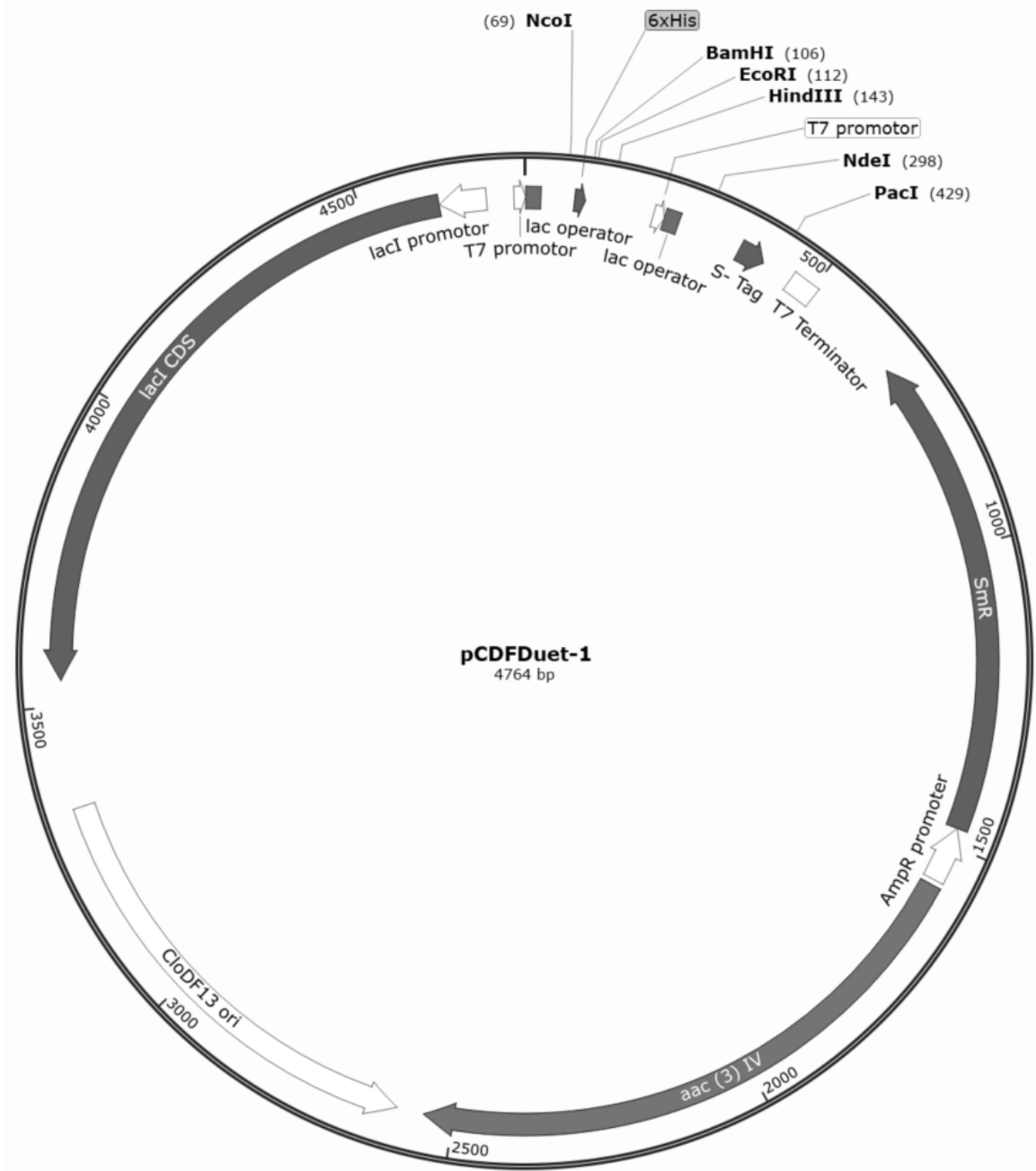


Figure 8.22: Plasmid map of pCDFDuet-1. Utilized restriction sites are indicated.

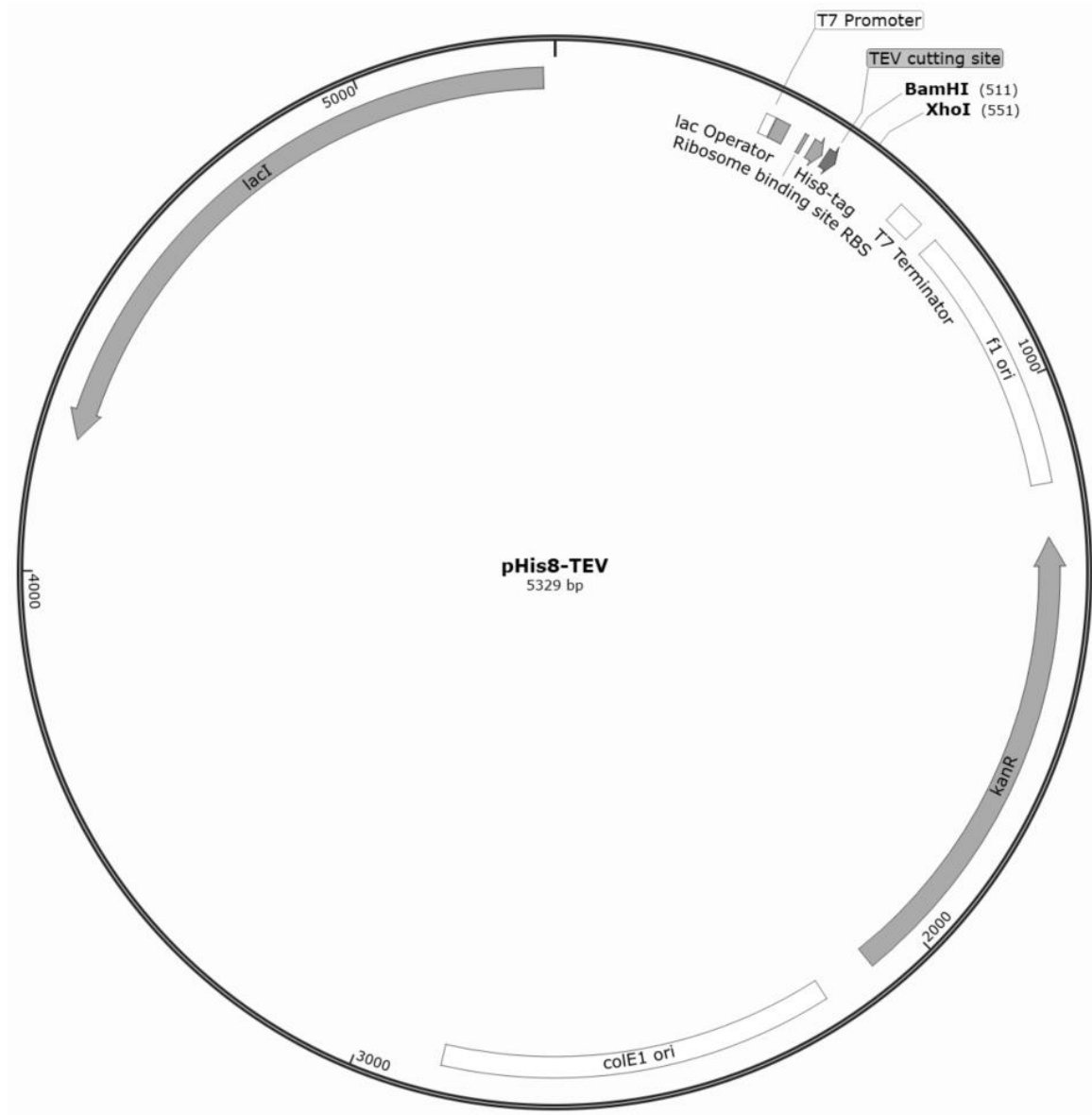


Figure 8.23: Plasmid map of pHis8-TEV. Utilized restriction sites are indicated.

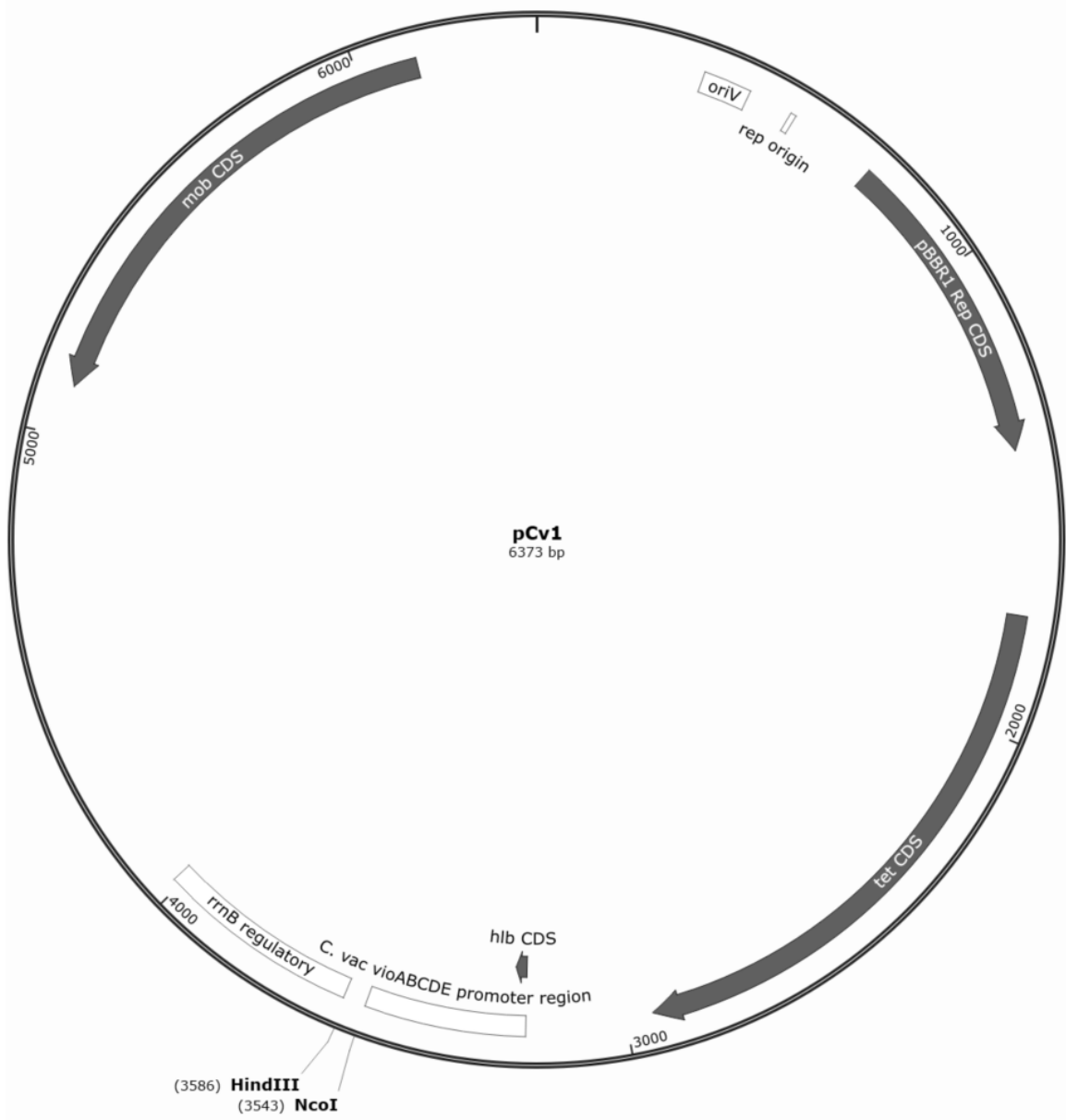


Figure 8.24: Plasmid map of pCv1. Utilized restriction sites are indicated.

9 Publications

Research publications:

C. Hermes, R. Richarz, D. A. Wirtz, J. Patt, W. Hanke, S. Kehraus, J. H. Voß, J. Küppers, T. Ohbayashi, V. Namasivayam, J. Alenfelder, A. Inoue, P. Mergaert, M. Gütschow, C. E. Müller, E. Kostenis, G. M. König, M. Crüsemann „Thioesterase-mediated side chain transesterification generates potent Gq signaling inhibitor FR900359“ *Nature communications* **2021**, *12*, 144.

DOI: [10.1038/s41467-020-20418-3](https://doi.org/10.1038/s41467-020-20418-3)

D. A. Wirtz, K. C. Ludwig, M. Arts, C. E. Marx, S. Krannich, P. Barac, S. Kehraus, M. Josten, B. Henrichfreise, A. Müller, G. M. König, A. J. Peoples, A. Nitti, A. L. Spoering, L. L. Ling, K. Lewis, M. Crüsemann, T. Schneider „Biosynthesis and Mechanism of Action of the Cell Wall Targeting Antibiotic Hypeptin“ *Angewandte Chemie (International ed. in English)* **2021**, *60*, 13579-86.

DOI: [10.1002/anie.202102224](https://doi.org/10.1002/anie.202102224)

Poster presentations:

D. A. Wirtz, I. Stritzinger, C. Hermes, G. M. König, M. Crüsemann „ β -Hydroxylases in NRPS - Functionality and Substrate Recognition“ Workshop of the International Association for General and Applied Microbiology (VAAM) 2019, Jena, Germany

D. A. Wirtz, S. Kehraus, G. M. König, K. Lewis, T. Schneider, M. Crüsemann „Characterization of stereospecific β -hydroxylases leads to structure refinement of the antibiotic hypeptin” Directing Biosynthesis 2021, online

10 Acknowledgements

Zu guter Letzt möchte ich mich bei allen Personen bedanken, die mich während der letzten vier Jahre unterstützten und somit zum guten Gelingen dieser Dissertation beigetragen haben.

Mein Dank gilt zunächst meiner Doktormutter, Frau Prof. Dr. Gabriele M. König, die mich spontan und offenherzig in ihrer Arbeitsgruppe aufgenommen hat. Sie hatte immer ein offenes Ohr und half dabei, den Überblick über die unterschiedlichen Projekte meiner Dissertation zu behalten. Durch ihre beherzten Entscheidungen habe ich die richtigen Projekte priorisiert und konnte diese zeitig und mit guten Ergebnissen abschließen. Vielen Dank für diese lehrreiche Lektion.

Genauso möchte ich Dr. Max Crüsemann für die Betreuung in den letzten vier Jahren danken. Er war immer motiviert und motivierend, selbst wenn es einmal nicht so gut lief. Max hat mir viele Möglichkeiten in den Projekten gelassen und auch ausgefalleneren Ideen Raum zur Verwirklichung gegeben. An seiner Bürotür wurde ich nie abgewiesen und er hatte immer Zeit für ein kleines Brainstorming.

Ich danke den weiteren Mitgliedern der Prüfungskommission für die Übernahme der Gutachten: Prof. Dr. Tanja Schneider, Prof. Dr. Michael Gütschow und Prof. Dr. Matthias Geyer. An dieser Stelle möchte ich mich auch für die Kooperationen mit ihren Arbeitsgruppen bedanken

Besonders möchte ich mich bei all meinen Mitarbeitern der AG König/Crüsemann und AG Kostenis bedanken. Anfangs bin ich im Institut für pharmazeutische Biologie offen empfangen worden und es waren stets alle hilfsbereit. Das herzliche Arbeitsumfeld hat mir durch so machen Tiefpunkt während der Promotion geholfen. Einzeln bedanken möchte ich mich bei Dr. Cornelia Hermes für die Einarbeitung im Labor, die erfolgreiche gemeinsame Etablierung diverser Enzymassays, das Korrekturlesen dieser Arbeit und die partielle Rettung der Praisoskirche. Bei Dr. René Richarz möchte ich mich bedanken für wertvolle Tipps im Bezug auf molekularbiologische Methoden, sowie bei meinen Bürokolleginnen Cora Hertzler, Wiebke Hanke und Sophie Klöppel für das Erhellern des Büroalltags und die Schublade des Glücks :-). Im Labor hätte nichts so gut funktioniert ohne die tatkräftige Unterstützung von Dr. Stefan Kehraus, Ekaterina Egereva und Emilie Mohr.

Auch wenn vorne auf dieser Dissertationsschrift alleinig mein Name steht, haben viele Kollegen und Studenten tatkräftig mitgeholfen, Daten zu sammeln. Stellvertretend für die vielen Studenten soll hier Melina Schadt genannt sein, die innerhalb von wenigen Monaten Dutzende Konstrukte des *hyn* BGC exprimiert hat. Mein Dank gilt außerdem Niels Schneberger für die viele Geduld beim Kristallisieren und an der ITC, sowie Kevin C. Ludwig und Melina Arts für die hervorragende Zusammenarbeit am Hypeptinprojekt (y wie ü). Es ist verrückt und sehr schade, dass wir publizieren konnten, ohne uns je gesehen zu haben.

Natürlich hätte ich es nie so weit geschafft ohne eine Familie, die mir dauerhaft den Rücken stärkt: Kirsten und Conny, es ist nicht zu Papier zu bringen, was ihr alles für mich getan habt. Allein dabei zuzuhören, wie ich von der Arbeit rede, ohne ein Wort zu verstehen war bestimmt nicht leicht.

Sarah, herzallerliebste Schwester, ab jetzt hast du den niedrigsten Bildungsabschluss von uns Dreien. Zum Glück gibt es im Land aber auch Arbeit für die einfachen Leute. Immerhin weiß die Ingenieurin, wie der Werkzeugkoffer aufgeht.

Christian, mein Bruder, danke für die Inspiration eine Promotion zu wagen.

Thank you Sive, for proofreading this thesis. The redundant commas were countless but you found them all.

Markus, mein Lieblingsschwager, danke für Gelassenheit und Ruhe, für Ablenkung und Blödeleien und natürlich für clever mit „k“... cleverk.

Das Beste kommt immer zum Schluss: meine Frau Carina. Ohne dich hätte ich es nie so weit geschafft. Gegenseitig haben wir uns durch diese schönen, wie auch harten, vier Jahre getragen und ich freue mich auf die vielen weiteren Jahre mit dir.