

Investigating the Biosynthesis of G Protein Inhibitor FR900359

Dissertation

zur Erlangung des Doktorgrades (Dr. rer. nat.)
der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Sophie Klöppel

aus Schkeuditz

Bonn

2024

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

Gutachter/Betreuer: Dr. Max Crüsemann
Gutachter: Prof. Dr. Martin Baunach
Tag der Promotion: 11.12.2024
Erscheinungsjahr: 2025

Table of Contents

1.	Abstract	1
2.	Introduction.....	3
2.1.	Specialized metabolites	3
2.2.	FR900359	4
2.3.	Gq protein inhibitor FR900359.....	9
2.3.1.	G protein-related signaling.....	9
2.3.2.	Inhibitory mechanism of FR900359.....	11
2.3.3.	Pharmacological relevance of FR900359	13
2.4.	Non-ribosomal peptide FR900359.....	15
2.4.1.	Non-ribosomal peptide synthetases	15
2.4.1.1.	Adenylation domain	16
2.4.1.2.	Condensation domain.....	17
2.4.1.3.	Thiolation domain	18
2.4.1.4.	Thioesterase domain.....	19
2.4.1.5.	Secondary/Modifying domains	20
2.4.1.5.1.	Epimerization domain	20
2.4.1.5.2.	Methyltransferase domain	20
2.5.	FR900359 biosynthesis	21
2.5.1.	The monomodular FrsA.....	23
2.5.2.	MbtH like protein FrsB.....	24
2.5.3.	Modifying enzyme FrsH.....	25
2.5.4.	Modifying enzyme FrsC	26
3.	Aims of study.....	27
4.	Results and Discussion	29
4.1.	Crystallization.....	29
4.1.1.	Construct FrsA _{TE} Ed	30
4.1.1.1.	Construct design by proteolysis	30
4.1.1.2.	Expression and Purification.....	34
4.1.1.3.	Crystallization trials	37
4.1.2.	Constructs FrsA _{TE} short and FrsA _{TE} long.....	37
4.1.2.1.	Construct design	37
4.1.2.2.	Cloning.....	39
4.1.2.3.	Expression and Purification.....	40
4.1.2.4.	Activity test.....	41
4.1.2.5.	Optimization approaches	44
4.1.2.5.1.	Maltose Binding Protein Fusion tag.....	44
4.1.2.5.2.	mVenus Fusion tag.....	46
4.1.3.	Didomain FrsA _{T+TE}	50
4.1.3.1.	Construct design	50
4.1.3.2.	Cloning.....	51
4.1.3.3.	Expression and Purification.....	51
4.1.3.4.	Crystallization trials	53

4.1.3.5.	Structure elucidation of FrsA _{T+TE}	55
4.1.4.	Generation of a Substrate Bound FrsA _{T+TE}	66
4.2.	Cryo-Electron Microscopy	69
4.3.	NMR structural analysis of FrsA _{TE}	72
4.3.1.	Expression of FrsA _{TE} in M9 medium	73
4.3.1.1.	Optimization of Expression in M9 medium.....	74
4.3.2.	Purification of ¹⁵ N-FrsA _{TE}	75
4.3.2.1.	Optimization of Purification	76
4.3.2.1.1.	Dimerization	76
4.3.2.1.2.	Degradation	77
4.3.2.1.3.	Storage	79
4.3.3.	NMR measurement	81
4.3.3.1.	NMR assignment and elucidation of structural dynamics.....	82
4.4.	Investigation of FrsA substrate specificity	87
4.4.1.	Purification of enzymes FrsH, FrsA _{CAT} , and FrsA _{TE} for in vitro assay	88
4.4.2.	<i>In vitro</i> assay optimizations	89
4.4.3.	<i>In vitro</i> synthesis of FR analogs.....	91
4.4.4.	Exchange of FR-Core as substrate for the TE domain	95
4.4.4.1.	Construction of FrsG8 _{TE} knock-out mutant.....	96
4.4.4.2.	Exchange of FR-Core by linearized FR-Core	97
4.5.	The Modifying Enzyme FrsH	103
4.6.	The Modifying Enzyme FrsC.....	106
5.	Conclusion and Outlook	109
5.1.	Construct generation of FrsA _{TE}	109
5.2.	Structural elucidation of FrsA _{T+TE}	111
5.3.	Optimization approaches for FrsA _{TE} expression in labeled M9 medium	114
5.4.	Investigation of protein dynamics of FrsA _{TE}	115
5.5.	Exploration of substrate specificity of the FrsA domains and obtainment of new FR analogs ..	116
6.	Material and Methods	119
6.1.	Chemical and reagents	119
6.2.	Vectors and plasmids.....	119
6.3.	Media, Buffers, and Solutions.....	120
6.4.	Microbiological methods.....	124
6.4.1.	Cultivation of bacteria	124
6.4.2.	Optical density measurement	125
6.4.3.	Strain maintenance	125
6.4.4.	Preparation of chemically competent cells	125
6.4.4.1.	Enabling chemical competence with calcium chloride.....	125
6.4.4.2.	Preparation of super-competent cells	126
6.4.5.	Transformation.....	126

6.5.	Molecular biological methods.....	126
6.5.1.	Primer design	126
6.5.2.	Polymerase chain reaction	129
6.5.2.1.	Q5 PCR	129
6.5.2.2.	GoTaq PCR.....	130
6.5.2.3.	Gradient PCR.....	131
6.5.3.	Agarose gel electrophoresis.....	132
6.5.4.	DNA isolation.....	132
6.5.4.1.	Isolation of genomic DNA.....	132
6.5.4.2.	Purification of PCR products.....	132
6.5.4.3.	DNA isolation from agarose gel	133
6.5.4.4.	Plasmid isolation.....	133
6.5.5.	DNA concentration determination.....	133
6.5.6.	Restriction digest.....	133
6.5.7.	Ligation	134
6.5.8.	Gibson assembly	135
6.5.9.	Sanger sequencing	135
6.5.10.	Generation of <i>E. coli</i> expression strains.....	136
6.5.11.	Construction of knockout mutants	136
6.5.11.1.	Preparation of the knockout vector.....	136
6.5.11.2.	Multistep conjugation	137
6.6.	Protein analysis.....	138
6.6.1.	Protein overexpression.....	138
6.6.1.1.	Overexpression in TB medium.....	138
6.6.1.2.	Overexpression in M9 medium with a spin-down approach	138
6.6.1.3.	M9 expression optimization	138
6.6.1.4.	Expression in labeled M9 medium	139
6.6.2.	Protein purification via nickel-nitrilotriacetic acid (Ni-NTA) affinity chromatography	140
6.6.2.1.	Purification optimization	140
6.6.3.	Desalting and buffer system exchange.....	141
6.6.4.	Concentration measurement and concentration.....	141
6.6.5.	Protein storage	141
6.6.6.	His-Tag removal utilizing TEV-protease.....	142
6.6.7.	Size Exclusion-Fast protein liquid chromatography	142
6.6.8.	Sodium Dodecyl Sulfate-Polyacrylamide gel electrophoresis.....	143
6.6.9.	Proteolytic digest.....	144
6.6.10.	Protein sequencing	144
6.6.11.	Analytical size-exclusion chromatography	144
6.6.12.	Crystallography and X-ray analysis.....	145
6.6.13.	CryoEM: sample preparation, data collection and processing	147
6.6.14.	NMR structural analysis	147
6.7.	<i>In vitro</i> assays.....	148
6.7.1.	Side chain assembly assay	148
6.7.2.	TE domain assay	149
6.7.2.1.	TE domain assay with synthetic side chain.....	149
6.7.3.	Substrate exchanges in side chain assembly assay.....	150
6.7.4.	Generation of substrate-bound FrsA TE domain.....	150
6.8.	<i>In vivo</i> investigations of <i>C. vaccinii</i>	151
6.8.1.	Fermentation and extraction.....	151
6.8.2.	Isolation of FR, FR-Core, and linearized FR-Core	151
6.9.	Bioinformatic methods	152

6.9.1.	AntiSMASH	152
6.9.2.	I-TASSER.....	152
6.9.3.	ExpasyProtParam	153
6.9.4.	SIM (Expasy Alignment tool).....	153
6.9.5.	AlphaFold	153
6.9.6.	SwissDock	154
6.9.7.	Dali server	154
6.9.8.	Visualization with PyMOL.....	154
7.	References	156
8.	List of Abbreviations.....	169
9.	Appendix.....	172
9.1.	Protein sequences	172
9.2.	Plasmids	176
9.3.	SDS-PAGE	178
10.	Publications	181

1. Abstract

FR900359 (FR) is a potent and selective inhibitor of Gq proteins, thereby obstructing intracellular signaling of numerous G protein-coupled receptors (GPCRs). This makes it a promising pharmacological tool for investigating Gq-coupled signaling and potentially addressing diseases associated with aberrant Gq protein activation. FR is a cyclic depsipeptide composed of eight building blocks. Of particular importance are the side chain, *N*-propionyl-hydroxy-leucine, and *D*-phenyllactate, which play crucial roles in FR's inhibitory properties by interacting with the α and β subunit of the heterotrimeric Gq protein. FR is produced by an octamodular non-ribosomal peptide synthetase (NRPS) machinery encoded by the *frs* biosynthetic gene cluster (BGC), which was found to be harbored by the bacterial plant endosymbiont "*Candidatus Burkholderia crenata*" and the soil bacterium *Chromobacterium vaccinii*. The latter is cultivable under laboratory conditions and enables heterologous expression in *Escherichia coli*, thus allowing *in vivo* and *in vitro* investigations. This facilitates the study of enzymes that are critically involved in FR biosynthesis and catalyze unique reactions.

FrsA_{TE} is a noncanonical thioesterase domain within the FR NRPS assembly line, catalyzing a rare intermolecular transesterification reaction between the side chain and FR-Core. To elucidate the mechanism of this unique reaction, experiments towards the comprehensive structural and functional characterization of FrsA_{TE} were conducted. After protein construct optimization using controlled proteolysis and bioinformatic analysis, and crystallization attempts with different constructs, the *apo*-structure of the FrsA_{T+TE} didomain could be successfully obtained. Comparative analysis with homologous TE domains revealed conserved structural features and potential substrate binding sites. Despite crystal soaking and cryogenic electron microscopy approaches, the *holo*-structure with the noncanonical substrate FR-Core remains elusive. Following extensive optimizations of expression and purification conditions, three-dimensional nuclear magnetic resonance (3D NMR) experiments provided insights into protein dynamics during substrate interactions. The results lay the foundation for future studies aimed at understanding the substrate recognition and catalytic mechanism of FrsA_{TE} in molecular detail. Additionally, the substrate specificity of the FrsA adenylation and condensation domains was investigated using *in vitro* assays, demonstrating flexibility in incorporating different leucine isomers and acyl-CoAs of various lengths. Furthermore, the acceptance of the TE domain towards the modified side chains was confirmed, resulting in the formation of new FR analogs. The generation of a linearized FR-Core, which could be obtained by constructing a *C. vaccinii* knock-out mutant, could not be proven by high-performance liquid chromatography-mass spectrometry (HPLC-MS) measurements. Taken together, these results provide profound insights into FrsA_{TE}'s

potential for chemoenzymatic syntheses and structure-activity relationship investigations of cyclic peptide G protein inhibitors.

To elucidate the molecular mechanisms of the nonheme diiron monooxygenases (NHDM) FrsH, biochemical characterizations were conducted based on the recently elucidated crystal structure. As the focus was particularly on the interaction between the leucine β -hydroxylase FrsH and the NRPS FrsA, several mutations were introduced based on bioinformatic analysis and investigated by *in vitro* assays and dynamic light scattering (DLS) experiments. This allowed the confirmation of the complex interface, and the identification of specific residues significantly involved in the interaction of FrsH and the FrsA adenylation domain providing insights into domain recognition during leucine β -hydroxylation and its potential for targeted engineering of NRPS systems.

The NAD(P)H-dependent dehydrogenase, FrsC, was hypothesized to catalyze the reduction of phenylpyruvate (PPA) to L-phenyllactate (PLA). Comprehensive biochemical characterization of FrsC revealed optimized reaction parameters, reaction kinetics, high specificity for PPA, and preference for NADPH as a cofactor. *In vitro* and *in vivo* experiments demonstrated the FrsC-catalyzed formation of L-PLA and its incorporation into the FR peptide chain, followed by online epimerization to D-PLA by the FrsE3 module. Comparative phylogenetic analysis positioned FrsC within the cytosolic-like malate dehydrogenase family, suggesting its evolutionary adaptation to accept aromatic substrates. These findings provide valuable insights into the specialized metabolism of FrsC and its role in the biosynthesis of D-PLA, a unique building block essential for the potent biological activity of FR900359.

2. Introduction

2.1. Specialized metabolites

Specialized metabolites (SMs), also known as natural products (NPs), are synthesized by every domain of the phylogenetic tree, with the main producers being bacteria, fungi, and plants. They are of small molecular weight (<5 kDa) and possess diverse chemical structures and biological activities. Unlike primary metabolites, their production is not essential for growth and reproduction, however, SMs play a key role in the survival of the producing organisms.^[1,2] SMs e.g. can serve as competitive agents against other organisms^[2,3], metal chelators^[4], agents of plant-microbe symbioses^[3], sexual hormones, and differentiation effectors.^[4] Thus, SMs provide adaption to adverse environments, enable intercellular communication, and contribute to protection from predators.^[5]

Through their broad range of bioactivities, SMs have become increasingly relevant in other areas such as agriculture, environmental remediation, and bio-medicals.^[6] The best-known and most popular examples of SMs in human health are antibiotics with some of the most prominent examples being penicillin^[7] and streptomycin^[8]. It has to be highlighted, that most modern clinical antibiotics are derivatives of SM originating from microorganisms.^[9] However, the areas of application have expanded in many ways over the last decades and now include the utilization of SMs as pesticides, pigments, toxins, pheromones, enzyme inhibitors, immunomodulating agents, receptor agonists, or antagonists, antitumor agents, immunosuppressives, cholesterol-lowering agent, and plant protectants.^[4]

SMs provide an incredible variety of structures, many of which show enormous complexity comprising for example β -lactams, cyclic peptides, and depsipeptides, containing non-proteinogenic amino acids, structurally complex sugars, and nucleosides to name just a few. The vast array of SMs is a result of the diverse producing organisms and their evolutionary history and ecological requirements. These drove the development of numerous and highly variable biosynthetic pathways encoding e.g. mega enzymatic machineries like polyketide synthetases (PKS) and non-ribosomal peptide synthetases (NRPS).^[4] The latter will be discussed in more detail in the next chapters as the focus of this thesis lies on the biosynthesis of the depsipeptide FR900359, a non-ribosomal peptide (NRP), consequently formed by an NRPS system.

2.2. FR900359

The complex cyclic depsipeptide FR900359 (FR) was first isolated from the higher plant *Ardisia crenata* in 1988.^[10] *A. crenata* is a low-growing evergreen shrub, endemic to south-eastern subtropical and tropical regions of Asia. The roots of the plant are used in traditional Chinese medicine as treatment for infections of the respiratory tract, toothaches, menstrual disorders, and tonsillitis.^[11–14] It cannot be assumed that these effects are caused by FR, considering it is mainly located in the leaves of the plant.^[15] The pharmacological effect attributed to the root may alternatively be ascribed to other bioactive compounds, including triterpenoids and saponins, which are present in high variety and concentration within the plant's root system.^[14]

Structural elucidation of FR was performed by hydrolytic cleavage of the building blocks, followed by mass spectrometric (MS) studies, and nuclear magnetic resonance (NMR) spectroscopy, revealing a complex cyclic depsipeptide (CD).^[10]

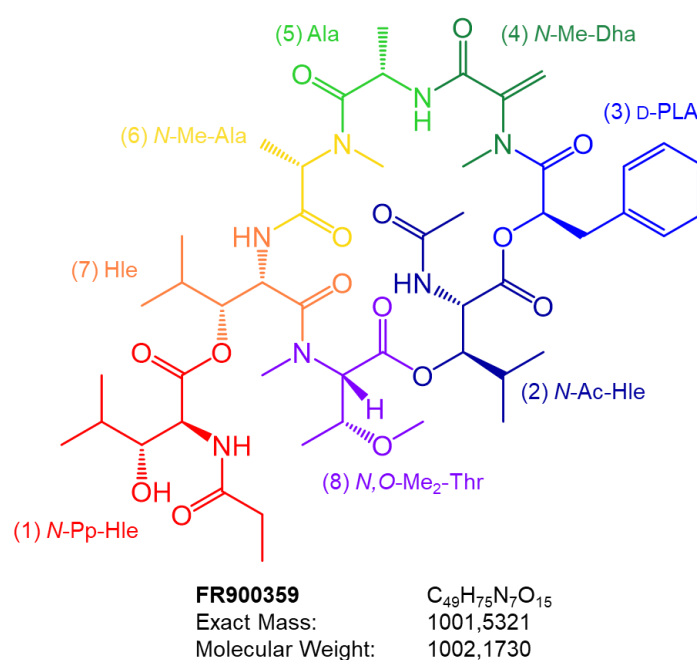


Figure 1: Structure of FR900359. Color-coded amino acid building blocks: (1) *N*-propionyl-(2*S*,3*R*)-3-hydroxyleucine (*N*-Pp-Hle), (2) *N*-acetylhydroxy-*L*-leucine (*N*-Ac-Hle), (3) *D*-phenyllactate (*D*-PLA), (4) *N*-methyl-*L*-dehydroalanine (*N*-Me-Dha), (5) *L*-alanine (Ala), (6) *N*-methyl-*L*-alanine (*N*-Me-Ala), (7) hydroxy-*L*-leucine (Hle) and (8) *N*,*O*-dimethyl-*L*-threonine (*N*,*O*-Me₂-Thr). The numeration corresponds to the respective NRPS module coordinating the incorporation of the building block (**Chapter 2.5**).

FR consists of eight building blocks, of which seven form the cyclic precursor molecule FR-Core: (2) *N*-acetylhydroxy-*L*-leucine (*N*-Ac-Hle), (3) *D*-phenyllactate (*D*-PLA), (4) *N*-methyl-*L*-dehydroalanine (*N*-Me-Dha), (5) *L*-alanine (Ala), (6) *N*-methyl-*L*-alanine (*N*-Me-Ala), (7) hydroxy-

L-leucine (Hle) and (8) *N,O*-dimethyl-L-threonine (*N,O*-Me₂-Thr). The side chain (1) *N*-propionyl-(2*S*,3*R*)-3-hydroxyleucine (*N*-Pp-Hle) is attached to the hydroxyl group of building block (7) Hle. The three-dimensional structure of FR and the absolute conformations of its stereo centers were elucidated using gas chromatography-mass spectrometry (GC-MS) and X-ray crystallographic analysis and described by Miyamae *et al.* (**Figure 1**).^[16]

FR shows structural features that are untypical for plant metabolites: non-proteinogenic amino acids, two adjacent ester bonds, and two *cis*-configured peptide bonds^[16] make a bacterial origin more plausible. Indeed, it was found that *Ardisia* species harbor endosymbiotic living bacteria in their leaf nodules^[17,18], and Carlier *et al.* were able to isolate bacterial DNA from these nodules in 2015. Genome sequencing and analysis allowed the classification into the genus *Burkholderia*, thus the designation “*Candidatus Burkholderia crenata*” (“*Ca. B. crenata*”) was proposed. The genome comprises 2.85 Mbp, thus being the second smallest genome of the *Burkholderia* genus known to date. The genome size of endosymbiotic living *Burkholderia* ranges from 2.4-6.1 Mbp, while free-living varieties harbor genomes of an average size of approx. 8 Mbp.^[19] Furthermore, the genome sequence of “*Ca. B. crenata*” is eroded, suggesting a symbiotic lifestyle. The bacteria are thus highly adapted to their obligate host, hampering cultivation and further investigation under laboratory conditions.^[20] Consequently, cultivation of “*Ca. B. crenata*” under laboratory conditions was not successful.

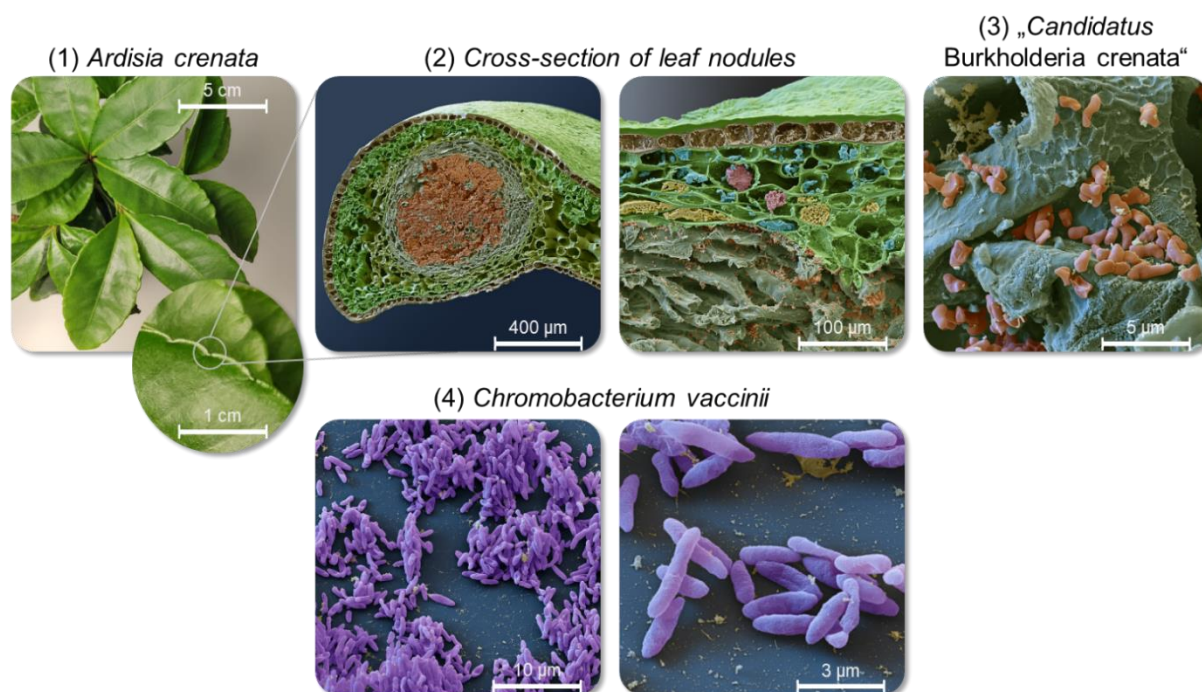


Figure 2: (1) Plant *Ardisia crenata*, with focus on leaf nodules located on the leaf margin. (2) Cross-section of leaf nodules, natural FR900359 producer (3) “*Candidatus Burkholderia crenata*” and (4) cultivable FR producer *Chromobacterium vaccinii*. Scale bars are shown in the pictures. Picture 1 is an own record and pictures 2-4 were kindly provided by Eye of Science.

On an extrachromosomal plasmid, an NRPS biosynthetic gene cluster (BGC) of 34 kb was identified, which comprises 8 genes, five of them encoding an eight modular NRPS machinery. Based on a primarily performed bioinformatic analysis, this BGC was assumed to be responsible for FR biosynthesis and named *frs*. This hypothesis was confirmed through the establishment of a heterologous expression system for *frs* in *Escherichia coli* (*E. coli*), leading to production of FR, albeit with low yields (1 µg/L). The *frs* BGC and the biosynthesis of FR is described in detail in **Chapter 2.5**.^[15]

To overcome the limitations that the low FR yield imposes on investigations under laboratory conditions, our group performed BLAST sequence similarity searches to identify homologous *frs* BGCs in cultivable bacteria.^[21] This search yielded fragmented DNA sequences, exhibiting high similarities to *frs* genes in “*Ca. B. crenata*”, belonging to the Gram-negative soil bacterium *Chromobacterium vaccinii* MWU205. This strain was discovered and isolated from the soil and roots of cranberry plants in Massachusetts (USA).^[22,23] Whole genome sequencing enabled the elucidation of the *cv_frs* BGC, which is nearly identical to the *bc_frs* BGC of “*Ca. B. crenata*”, only differing in GC content and length of the intergenic regions. The *cv_frs* BGC comprises the same number of genes (*frsA-frsH*), with antiSMASH analysis revealing the same organization in modules and domains. First fermentation approaches of *C. vaccinii* under laboratory conditions yielded 2.5 mg/L FR, making it a promising source for the quick and reliable production of FR, which is an effective pharmacological tool (see **2.3.3**).^[21] Due to the fact that *C. vaccinii* is a cultivable bacterium, its genome allows genetic modification, including mutagenesis, cloning, and heterologous expression in *E. coli*.^[21,24–26] This enabled the detailed investigation of FR biosynthesis.^[21]

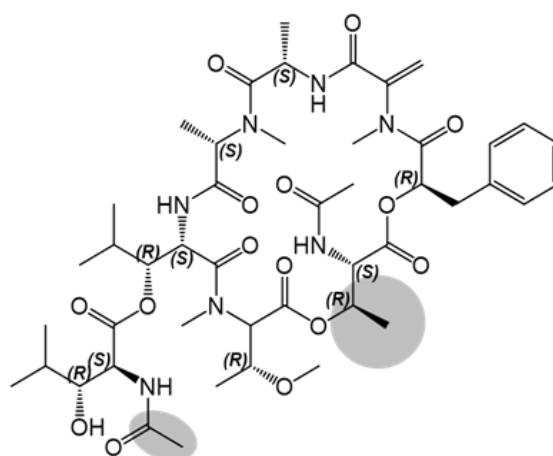


Figure 3: Structure of YM-254890. Structural differences to FR are highlighted in gray.

A structurally very similar compound to FR, called YM-254890 (YM), was isolated from the soil bacterium *Chromobacterium* sp. QS3666 (**Figure 3**), which was isolated from a soil sample gathered in Tokyo, Japan, in 2003.^[27] Structure elucidation revealed that YM differs from FR only in two building blocks: the presence of an *N*-acetyl group instead of *N*-propionyl located in the side chain (anchor 1), and the incorporation of a threonine replacing the β -hydroxyleucine in building block 7 (anchor 2), resulting in the loss of three methyl groups compared to FR.^[28] The biosynthesis of YM was recently described by Hashimoto *et al.*^[29] The BGC responsible for YM formation was called *ytf*, consisting of eight genes, which correspond in order and size to the *frs* BGC (**Chapter 2.5**). The sequence similarity of the individual genes ranges from 54-83%. The differences in the structures between FR and YM can be explained by the fact, that the NRPS modules 8 and 1 (in the *frs* BGC these correspond to modules 1 and 2) accept distinct substrates.

Besides FR and YM, further naturally occurring derivatives of FR were isolated from *A. crenata* and *C. vaccinii* by our group and are shown in **Figure 4**.^[15,27,30] They exhibit differences in the presence or length of the side chain, as well as deviant building blocks (2) *N*-Ac-Hle, (5) Ala, and (8) *N,O*-Me₂-Thr, indicating that the biosynthetic enzymes that incorporate these building blocks possess a flexible substrate specificity.^[15,31] Natural derivatives of YM were isolated from *Chromobacterium* sp. QS3666, these differ only in the side chain: either it is completely absent, or the length of the acyl moiety varies between an ethyl and a methyl group.^[27]

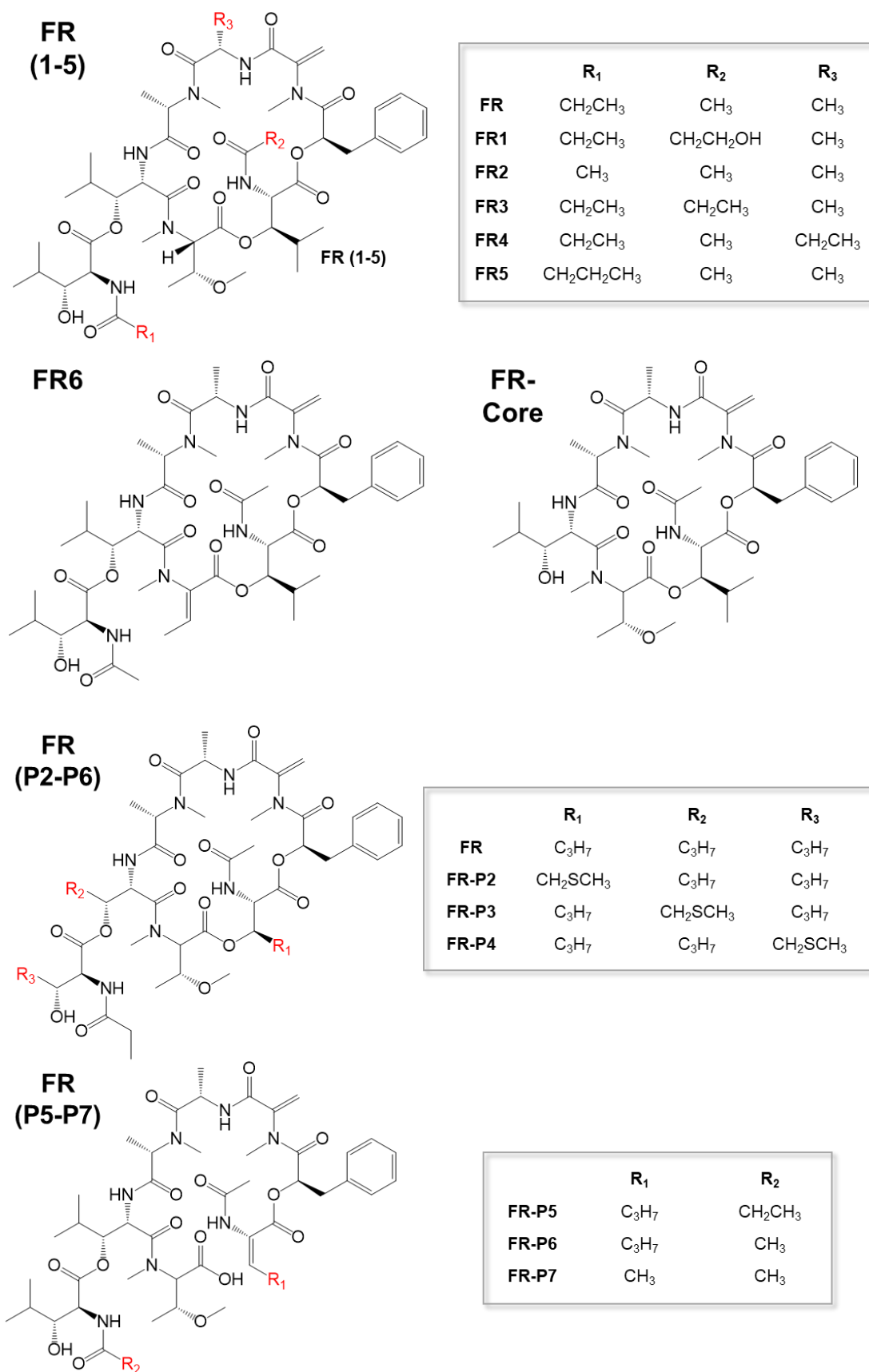


Figure 4: Structures of naturally occurring FR derivatives (P = analogs found and numbered by Pistorius *et al.*^[32]).

2.3. Gq protein inhibitor FR900359

Over the last decades, FR gained the attention of researchers, due to its potent and selective inhibitory effect on Gq protein-mediated signaling. The following chapters will provide a detailed explanation of G proteins and the related signaling pathways, as well as insights into the inhibitory effect of FR and the underlying mechanism. Further, the relevance of FR from a pharmacological point of view will be addressed.

2.3.1. G protein-related signaling

G protein-coupled receptors (GPCRs) are transmembrane proteins and belong to the largest family of membrane-bound receptors encoded in the human genome.^[33] They are crucially involved in almost all physiological processes in humans and other mammals.^[34,35] As their name indicates, on the inner side of the cell membrane, GPCRs are coupled to heterotrimeric guanine-nucleotide binding proteins (G proteins), which are responsible for the transduction from extracellular chemical signals to intracellular signaling cascades.^[36–38] When a GPCR is activated by an extracellular ligand, comprised of a wide range of chemotypes, like small molecules, peptides, proteins, and lipids, the signal is transduced to the intracellularly coupled trimeric G protein,^[19,39] consisting of an α -, β -, and γ -subunit, of which the $G\alpha$ binds GDP in its inactive form.^[40] In response to the extracellular activated GPCR, the G protein is released and dissociates into the $G\alpha$ subunit and the constitutive $G\beta\gamma$ complex. Both units interact with cognate effectors, initiating downstream signaling, whereby the $G\beta\gamma$ complex can influence various pathways by modulating the formation of second messengers, such as cyclic AMP (cAMP).^[40,41] The $G\alpha$ subunit transitions into the active form through conformational changes initiated by the exchange of GDP for GTP. Since the $G\alpha$ subunit is an intrinsic GTPase, the terminal phosphate of GTP is released and GDP remains in the binding pocket, allowing the subunit to reunite with the $G\beta\gamma$ complex and revert to the inactive form, which is then available for recruitment by the activated receptor.^[42] This closes the cycle which can subsequently start over again.

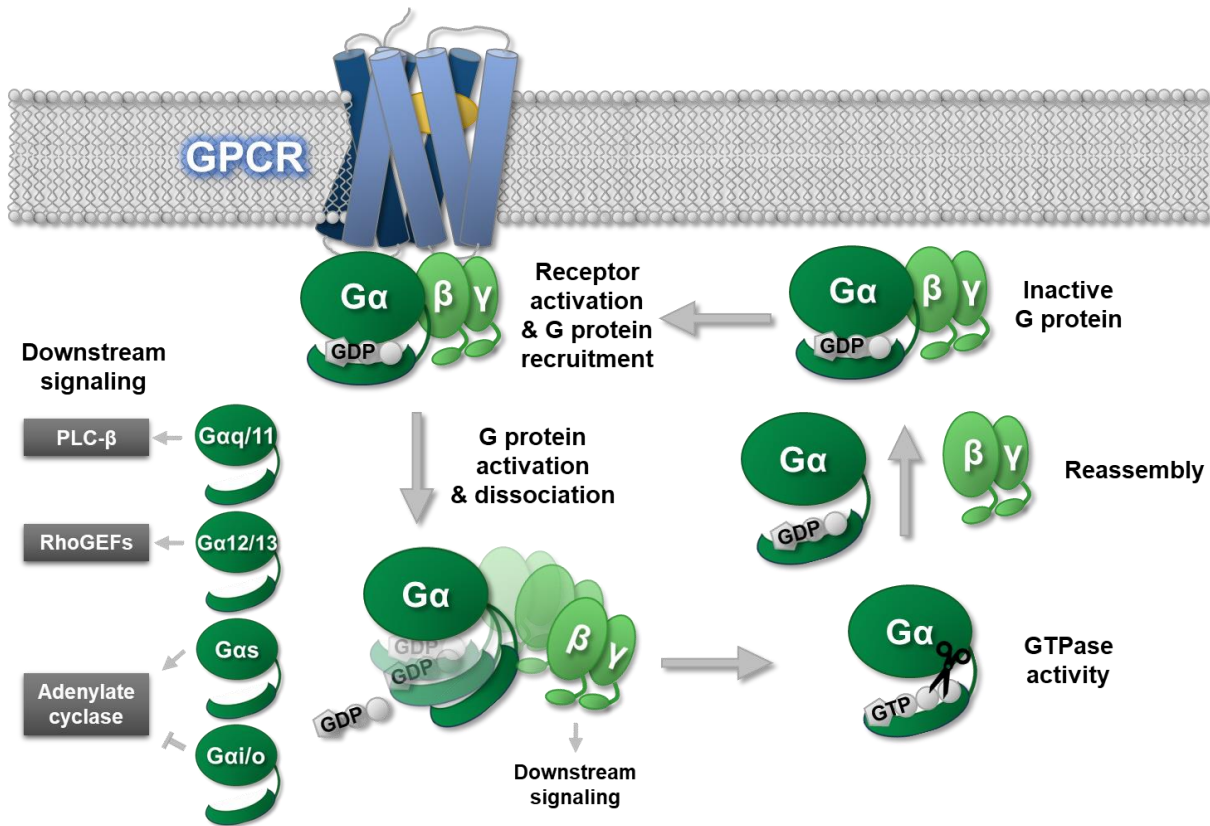


Figure 5: Schematic illustration of G protein-related signaling, adapted from Tietze *et al.*^[43] and Zhang *et al.*^[44]. After GPCR activation by an extracellular ligand, it undergoes a structural change, activating the receptor-bound heterotrimeric G protein, which dissociates into the Gα subunit and the βγ-complex, releasing GDP and activating downstream effectors. Only canonical effectors of the respective Gα protein families are depicted (PLC-β = phospholipase C-β isozymes; RhoGEFs = Rho guanine nucleotide exchange factors). The Gα subunit binds GTP and removes the terminal phosphate, enabling the reassembly with the βγ-complex. After GPCR activation, the heterotrimeric G protein gets recruited, allowing the cycle to start all over again.

There are over 800 genes encoding GPCRs present in the human genome, underlining the importance of their broad range of physiological functions. However, GPCRs and G protein underwent an asymmetric evolution^[45], resulting in only 16 different Gα protein subunits, which can be grouped depending on their sequence similarity in four subfamilies Gαs, Gαi/o, Gαq/11, and Gα12/13 (**Figure 5**).^[42,46] Activation of each of the subfamilies causes different cellular responses, as they interact with distinct effector proteins or modify the activity of the same effector in opposite directions. For example, Gαq/11 proteins activate phospholipase C-β isozymes (PLC-β), resulting in the conversion of phosphatidylinositol (3,4,5)-triphosphate (PIP3) to inositol triphosphate (IP3) and diacylglycerole (DAG).^[47] In contrast, Gα12/13 proteins can affect the cytoskeleton *via* activation of small GTPases through the interaction with Rho guanine nucleotide exchange factors (RhoGEFs).^[48] Gαs proteins stimulate the adenylate cyclases (AC), increasing the intracellular levels of the second messenger cAMP, whereas Gαi/o proteins reduce the level of

cAMP by inhibiting AC.^[42] Only the most important canonical effectors have been mentioned here, however, $G\alpha$ proteins interact with a multitude of further proteins.^[49]

As a result of this asymmetric evolution (**Figure 6**), the signaling of a vast variety of GPCRs has to converge on four main G protein families only, highlighting the relevance of pharmacologically targeting GPCR signaling on the G protein level, and opportunities that may arise from such strategy.^[45] Molecules with inhibitory activities on G proteins thus became the focus of researchers, searching for tools to investigate and understand signaling cascades associated with GPCRs and G proteins in detail as well as treatment for corresponding diseases, e.g. asthma, uveal melanoma, and cardiovascular disorders.^[50]

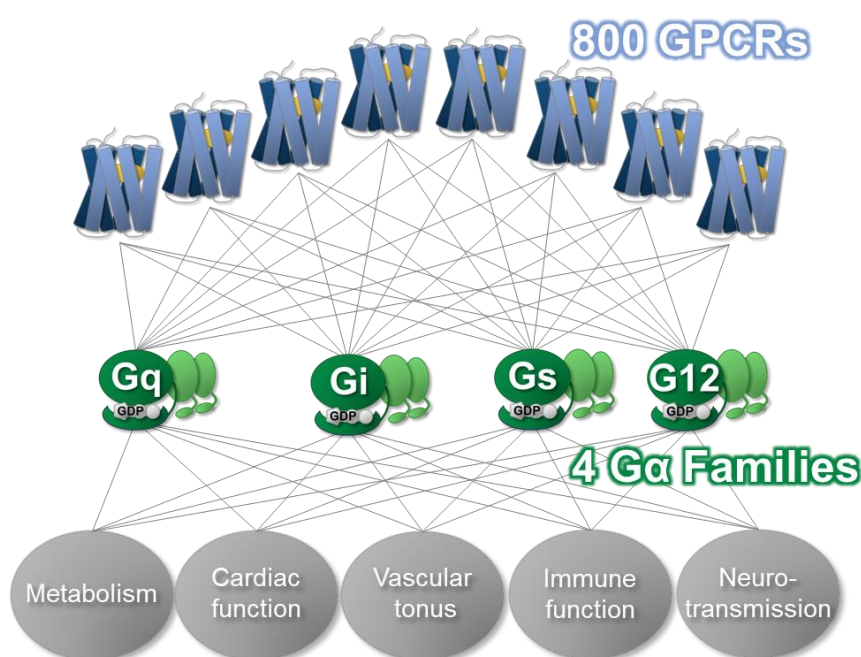


Figure 6: Schematic diagram of the coherence between GPCRs and G proteins and their influence on physiological processes.

2.3.2. Inhibitory mechanism of FR900359

The two complex natural depsipeptides FR and YM both possess a strong and selective inhibitory effect on $G\alpha_q$ proteins.^[51,52] The binding mechanism of YM was elucidated by X-ray crystallographic analysis of the YM-bound heterotrimeric $G\alpha_i/q\beta_1\gamma_2$ protein.^[53] YM binds to the inactive $G\alpha_q$ protein close to the GDP binding pocket in a cleft between the two domains of the subunit. The inactive state of the heterotrimer is stabilized *via* immobilization of the linkers, thereby preventing GDP release and dissociation. The stimulus that arrives at the GPCR and is forwarded to the G protein cannot be transmitted, whereby the downstream signal transduction is interrupted.

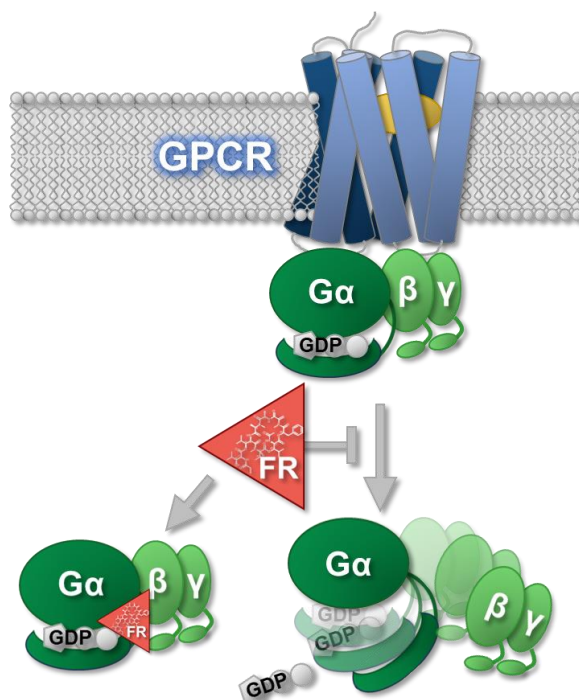


Figure 7: Molecular inhibition mechanism of FR on Gq protein, adapted from Tietze *et al.*^[43] After GPCR activation by an extracellular ligand, it undergoes a structural change, activating the receptor-bound heterotrimeric G protein, which dissociates into the G α subunit and the $\beta\gamma$ -complex, releasing GDP and activating downstream effectors. FR900359 blocks the domain opening, preventing the opening, GDP release, and dissociation of the heterotrimeric G protein. YM shows the same inhibition mechanism but is omitted in the illustration.

Like YM, FR belongs to the macrocyclic depsipeptides acting as guanine nucleotide dissociation inhibitors (GDI).^[54] Both differ only in the anchor regions: Instead of the two smaller methyl groups in YM, FR possesses a bulkier ethyl (anchor 1) and isopropyl group (anchor 2). Already in 2015, Schrage *et al.* experimentally substantiated the hypothesis, that FR and YM similarly bind to Gq and share the same inhibitory mechanism, which was, until then, only based on their high structural similarity (**Figure 7**).^[51] Recently, two high-resolution crystal structures of the G $\alpha_{11}\beta_{1}\gamma_2$ bound to FR and YM could be elucidated by Mühle *et al.*^[55] It was discovered that FR and YM not only bind to the G α subunit but also interact with the Arg96 residue of the G β subunit. This interaction contributes to the stabilization of the G protein heterotrimer, leading to the formation of a completely ‘silent’ complex and consequently prohibiting any downstream cellular signaling pathway. FR and YM showed the same binding pose with slight rearrangements in the protein backbone near the anchor regions. Despite the high similarity of the binding modes, YM seems to fit better into the binding pockets due to its smaller anchor regions, which leads to the formation of a rather transient interaction with a good binding affinity (pK_i = 8.23^[56]). In contrast, anchor 1 of FR leads to the formation of a strong hydrogen bond to Arg96, resulting in a higher binding affinity (pK_i = 9.23^[56]) Q20 and a stronger stabilization of the heterotrimer. Thus, FR shows a longer residence time, resulting in highly different dissociation kinetics of FR and YM.^[57,58]

Nevertheless, both compounds show a high affinity to Gαq, Gα11, and Gα14 proteins, all belonging to the family of Gαq proteins. In recent years, they have become an essential tool for the investigation of the Gq/11-driven signaling events. Furthermore, they both are promising treatment options against Gαq protein-associated diseases, which will be described in the next chapter.

2.3.3. Pharmacological relevance of FR900359

Over the last decades, multiple effects of FR on the cardiovascular system were discovered and investigated. In 2011, the vasorelaxant effect of FR on aortic arteries in rats was verified. This can be explained by the fact, that FR inhibits both voltage-dependent as well as receptor-dependent Ca²⁺ influx, of which the latter plays a major role in the inhibitory activity.^[59] Furthermore, FR impedes human platelet aggregation^[60] and blocks G protein-dependent vasoconstriction in mice.^[61] Moreover, an anti-hypertensive potential of chronically blocked Gαq in mouse models was investigated, revealing the potential of FR as a tool to quickly decrease blood pressure to a normal level. Nevertheless, by decreasing cardiac chronotropic and ionotropic responses, FR could cause acute changes in blood pressure, such as hypotension, which is why the applicability should be further investigated in detail.^[61]

Deviant activation of Gαq protein-dependent signaling proteins is known to cause asthma, leading to the assumption that inhibitors of those could be a promising treatment for this chronic lung disease.^[62] This hypothesis could be confirmed by Matthey *et al.*, who verified selective Gαq inhibitory activity of FR in mouse and human airway smooth muscle cells.^[63] FR thus acts as a strong broncho relaxant in mice, pigs, and humans *ex vivo*. Inhalation of FR leads to high local concentrations in the lungs and prevents the increase of airway resistance, as well as any unwanted effects on blood pressure and heart rate in a mouse model.^[63]

The first indication that FR could be a promising candidate for cancer treatment was shown in a study by Schrage *et al.*, who demonstrated that FR did not cause any direct cytotoxicity or interference with the mitochondrial metabolism in B16 melanoma cells, but instead inhibited cell proliferation and compromised cell migration.^[51] Further investigations were conducted in search of a treatment for uveal melanoma (UM), which is the most common form of intraocular cancer. Even though early-state treatment has developed considerably in recent years, the risk of distant metastasis (~50%) is unaltered.^[64] Thereby, the liver is affected in 90% of the cases.^[65] Once the UM becomes metastatic, therapy options are limited and the median overall survival numbers 10-13 months with a cure rate approaching nil.^[64,66] Investigation of the oncogenic mechanism revealed constitutively active Gαq or Gα11 mutants as cancer drivers in over 85% of UM cases^[67,68],

making them the primary therapeutic target. As a strong and selective inhibitor of Gαq/11 proteins, FR arrests UM cells in the G1 phase of the cell cycle, leading to several effects, such as evocation of melanocytic redifferentiation of UM cells with low metastatic potential, growth inhibition of xenografted UM tumors, attenuation of metabolic reprogramming, and inhibition of cytoskeletal dynamics and mechanosensing during cell migration.^[69] Despite all the mentioned effects against UM cells, FR neither kills tumor cells nor regresses UM xenografts in mouse models^[70–72], which does not diminish the value of FR as a potential new weapon in the battle against the most common malignant primary tumor of the eye.

Since it was discovered that there is also a correlation between Gαq signaling and obesity, FR could provide a treatment option for this disease with constantly increasing prevalence.^[73,74] Adults possess metabolically active brown adipose tissue, which plays an important role in basal and induced energy expenditure. In overweight individuals, these brown adipocytes exhibit reduced activity^[75], which is related to the fact that Gq signaling crucially inhibits adipocyte differentiation. *In vivo* experiments showed that Gαq signaling inhibits brown and beige adipose tissue and adipogenesis in general. Antagonizing Gαq signaling with FR enhances the formation of brown and beige adipocytes, leading to increased energy expenditure, making FR a novel approach to counteract obesity.^[73]

In this chapter, all known pharmacological opportunities for FR were depicted, highlighting a promising future for FR or FR-derived compounds as treatment against several serious diseases including cancer, and respiratory or cardiovascular ailments. Nevertheless, the utilization of FR is still limited due to its general toxicity at relevant concentrations.^[76] This derives from the fact, that FR blocks Gαq signaling throughout the whole organism, leading to a general systemic response like hypotension^[63]. To overcome this obstacle, FR needs to be applied in a target-oriented and specific manner. There are various approaches like linking potential toxic molecules to specific transport molecules, called antibody-drug conjugates^[77–79], targeting disease state Gαq proteins or using aptamers for targeted drug delivery.^[54]

In summary, FR represents a promising therapeutic and scientific tool with potential applications in the treatment and study of the aforementioned diseases. In addition, FR is outstandingly suited to investigate Gα-dependent processes, which provides new pharmacological research opportunities. Despite current limitations due to its general toxicity, targeted delivery methods may unlock its full potential, paving the way for significant advancement in modern medicine.

2.4. Non-ribosomal peptide FR900359

As mentioned in **Chapter 2.2**, FR is a cyclic depsipeptide, composed of eight building blocks, of which seven are non-proteinogenic. Its biosynthesis can be attributed to the *frs* BGC, encoding for two NRPS machineries, two tailoring enzymes, and an MbtH-like protein.

NRPs are mainly produced by bacteria and fungi.^[80] NRPs exhibit a high degree of structural variety, which is attributable to several factors: size, variety of building blocks, and their linkage. The size ranges from 2 to 25 residues which can be arranged cyclic, branched, or linear.^[80,81] It should be particularly emphasized that the high diversity derives from the fact, that the building blocks are not only proteinogenic amino acids but more than 500 monomer substrates, including L- and D-amino acids, hydroxy acids, keto acids, fatty acids, and aryl acids have been described.^[81]

NRPs were first discovered when Tatum and co-workers investigated the biosynthesis of tyrocidine in *Bacillus* species and observed that ribosomal inhibitors had no negative effect on the yield, which is why they hypothesized a mechanism independent of ribosomal biosynthesis.^[82] In 1965, Lipmann and Laland provided a more detailed insight into the biochemical mechanisms of NRPS^[83–86]. Over recent years, NRP research gained further relevance, as they are a promising resource for drug discovery and development. Their detailed structural, biochemical, genetic, and bioinformatic analysis led to an increasing understanding of their function, which is described in the following section, providing the fundamental knowledge to address FR biosynthesis in detail below.^[87]

2.4.1. Non-ribosomal peptide synthetases

Non-ribosomal peptide synthetases are multifunctional mega-enzymatic complexes, consisting of modules that are each composed of different domains.^[88] Modules are defined as sets of domains, possessing the ability to add an acyl residue to the nascent peptide chain.^[89–92] The core domains of a standard elongation module are the adenylation (A) domain, the condensation (C) domain, and the thiolation (T) domain (also known as peptidyl carrier protein (PCP)).^[81] Modules can comprise additional modifying domains or interact with tailoring enzymes, leading to an astonishing structural diversity of the domains, as well as of the building blocks.^[80] The organization of different modules in sequence is called an assembly line in which each module is responsible for the activation and incorporation of one specific building block, leading to the elongation of the peptide, which is finally released by the thioesterase (TE) domain of the terminal module.^[93] Canonical NRPS systems follow the co-linearity rule, wherein the number and order of

the resultant peptide are directly determined by the arrangement and specificity of the elongation modules, each incorporating a single defined building block.

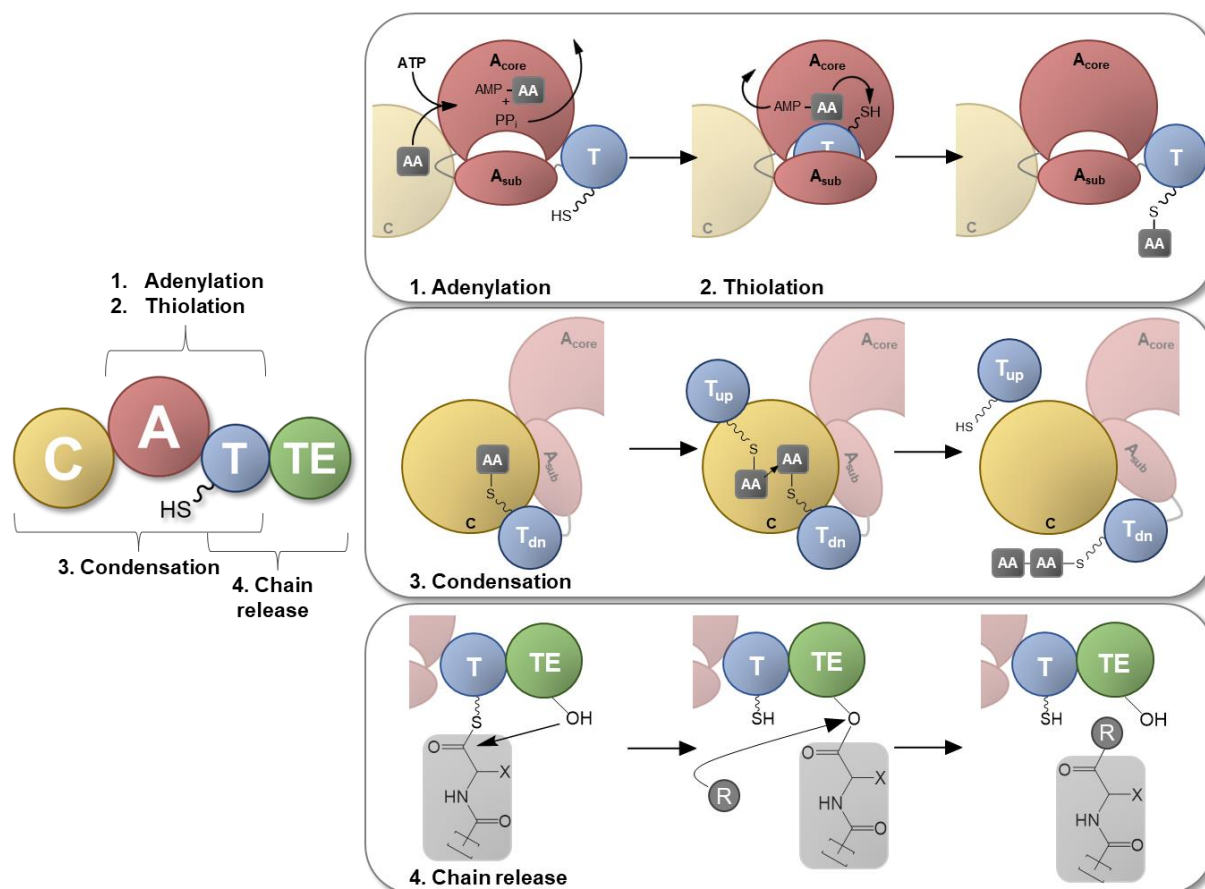


Figure 8: Simplified schematic illustration of elongation and release at a canonical NRPS module. The adenylation (A) domain (red) comprises a larger (A_{main}) and a smaller (A_{sub}) subdomains. A_{main} catalyzes the activation of a specific amino acid (AA), forming an aminoacyl-adenylate (AMP-AA), using ATP, and releasing PP_i. The A_{sub} domain uncovers the active site, leading to the entering of the thiolation (T) domain (blue). The AMP-AA is thiolated onto the phosphopantetheine moiety attached to the T domain. The loaded aminoacyl-S-T (T_{dn}) domain gets coordinated in the acceptor side of the downstream condensation (C) domain (yellow), for the condensation step. On the donor side, the C domain binds the upstream T (T_{up}) domain and catalyzes the transfer of the amino acid, leading to peptide elongation at the (T_{dn}) domain. In the final step of chain release, the peptidyl-S-T domain approaches the active site of the thioesterase (TE) domain (green), which catalyzes the attack on the thioester, resulting in the formation of an acyl-O-TE domain intermediate. In the second step, a nucleophile (R) attacks the acyl-O-TE intermediate, replacing the oxygen of the serine and causing the release.

2.4.1.1. Adenylation domain

Peptide formation starts with the adenylation (A) domain, that selects a cognate monomer substrate with a set of residues within its binding pocket. In this active center, an ATP molecule gets bound, which is necessary to catalyze the adenylation reaction, leading to the formation of an activated aminoacyl-adenylate. The A subdomain (A_{sub}), located at the C-terminus end of the A domain, is responsible for the correct positioning and binding of the T domain so that its 4'-phosphopantetheinyl (Ppant) moiety protrudes into the active site of the A domain. A covalent

link is formed between the aminoacyl moiety of the aminoacyl-adenylate and the Ppant thiol of the T domain forming a thioester (aminoacyl-S-Ppant-T domain).[81]

A domains belong to the ANL superfamily (Acyl-CoA synthetases, NRPS A domains, Luciferase enzymes).[94] They consist of 2 subdomains: the N-terminal 50 kDa core domain (A_{core}) and the C-terminal ca. 10 kDa subdomain (A_{sub}), linked by a hinge region. The selectivity of the A domain is determined by a conserved amino acid sequence of up to eight residues, which is located in the binding pocket, also known as 'Stachelhaus code', 'specificity code' or 'non-ribosomal code'.^[94-97] The Stachelhaus code enables concrete *in silico* prediction for substrate specificity for A domains.

The substrate range of A domains is not restricted to proteinogenic amino acids. Until now over 500 structurally diverse monomer substrates were described, like L- and D-amino acids, keto acids, fatty acids, hydroxy acids, and aryl acids.^[98] This is a fundamental factor for the structural diversity of NRPs, as well as a relaxed substrate specificity, that could be observed in some A domains. This provides an evolutionary advantage for the NRP-forming organism for selection and is usually the consequence of a degenerated non-ribosomal code.^[99,100]

The A core domain binds ATP and the monomer close to the A_{core} - A_{sub} interface and catalyzes an adenylation reaction of the selected monomer. The activated monomer is then attached to the adjacent T domain.

The presence of an MbtH-like protein (MLP) is of utmost importance for the functionality of many A domains of bacterial origin.^[101,102] The name derives from a small protein of mycobactin biosynthesis in *Mycobacterium tuberculosis*.^[103] Commonly, MLP genes cluster with NRPS-coding BGC indicating coexpression and specificity.^[104] *In vivo* studies were conducted to elucidate the function of MLPs, which revealed that MLPs are less specific than assumed and can also interact with multiple NRPS, thus functionally replacing each other.^[105] This hampered the investigations, as all other MLP genes in the genomes had to be inactivated.^[106,107] It was discovered, that MLPs possess a chaperoning effect, increasing the solubility of NRPS enzymes.^[105,108,109] Furthermore, it has been shown that MLPs interact with A domains and are crucial for their activity and correct functioning.^[108]

2.4.1.2. Condensation domain

The condensation (C) domain possesses binding sites for the upstream (donor) and the downstream (acceptor) T domain and induces proximity between two Ppant-bound aminoacyl moieties. This enables the catalyzation of a condensation reaction between the amino group of the acceptor aminoacyl-S-Ppant-T domain and the thioester carbonyl of the donor peptidyl-S-Ppant-T

domain, leading to the transfer of the peptidyl moiety to the aminoacyl-S-T domain. This step results in the elongation of the peptide chain.^[110] Subsequently, the elongated peptidyl-S-T domain moves to the downstream C domain, where the T domain provides the peptide chain for the next condensation reaction.

The C domain superfamily can be phylogenetically classified into several subclasses.^[111] The $^L C_L$ domains catalyze the formation of the peptide bonds between two L-configured monomers. In contrast, the $^D C_L$ domains condensate a D-configured monomer at the N-terminus of an L-configured monomer.^[112] Most initial NRPS modules do not comprise C domains, but if they do, it is a so-called starter C domain (C_{start}). These acylate the first monomer with acyl carrier protein (ACP)-bound fatty acids, polyketides, or other residues.^[113] Furthermore, some non-canonical members of the C domain superfamily were described, e.g. dual E/C domains^[114], Cy domains^[115], X domains^[116,117], 'hybrid C domains' and 'modified AA C domains' ($^{modAA}C$ domains).^[118] The latter seems to be involved in the modification of incorporated amino acids, like the dehydration of serine to dehydroalanine.^[119]

C domains have a size of about 50 kDa. Their structure represents a pseudo-dimeric domain, consisting of the C-terminal and the N-terminal section.^[120] In $^L C_L$ domains, the active site is located in the tunnel between the two segments and contains the conserved motif HHxxxDG^[121], of which the second His residue is of utmost importance for the catalytic activity. It is proposed that this residue deprotonates the amino group of the acceptor substrate to enable the nucleophilic attack, stabilizes the electronic transition state^[122], or positions the substrate.^[110] The C domain possesses binding sites for the upstream and downstream T domains on the two ends of the active site tunnel. Once the T domains are bound, the Ppant moieties protrude into the tunnel, resulting in close proximity of the bound acyl residues within the active site.^[110]

2.4.1.3. Thiolation domain

T domains have a size of approximately 10 kDa and are composed of four α -helices. The N-terminus of the second helix harbors the highly conserved core motif GxxS.^[123,124] The Ser in this motif is posttranslationally modified by a phosphopantetheinyl transferase (PPTase), which attaches the flexible Ppant arm.^[125] The Ppant-bound T domain transfers the amino acyl-/peptidyl-thioester intermediates to all catalytic centers of the C-, A-, TE- and modifying domains. The high flexibility of the Ppant arm is crucial for the functioning of the NRPS, due to its responsibility in shuttling the growing peptide chain through the NRPS assembly line.^[126] Interestingly, T domains possess a major interface, which is formed by helix $\alpha 2$, the preceding loop,

and helix $\alpha 3$. It enables the interaction with PPTases^[127] and with all domains of the respective NRPS module: the C,^[128] A,^[128] E^[129], and TE domains^[130,131].

2.4.1.4. Thioesterase domain

Another core domain of NRPS is the TE domain, which can be found at least once in the terminal module of most NRPS systems. The TE is the termination domain, which catalyzes the peptide release through hydrolysis or cyclization.

Thioesterase (TE) domains belong to the family of α/β -hydrolase proteins which are related to serine proteases.^[93,132] They are categorized into two types: Type I TE domains are located in the terminal NRPS module and are responsible for the release of the NRP from the respective T domain after the completed assembly. Type II TE domains possess a proofreading role. They recognize incorrectly loaded T domains and restore their function by hydrolysis of the Ppant moiety.^[99,133] In contrast to type I, type II TE domains can act as standalone enzymes.^[134] In this work, however, the focus is on type I TE domains, which is why only these are further discussed in detail.

The release of NRP from the terminal NRPS domains proceeds in two half-reactions (**Figure 8**), in which the catalytic triad Ser-His-Asp, located in the active site of the TE domain, plays a key role.^[81,131,132] The thioester substrate tethered to the Ppant arm of the T domain approaches the Ser of the catalytic triad from the N-terminal side of the substrate channel. The serine residue is deprotonated by the histidine residue, which allows the attack on the thioester of the peptidyl-S-T domain, resulting in the formation of an acyl-O-TE domain intermediate. This negatively charged tetrahedral transition state is stabilized by an oxyanion. In the second step, a nucleophile attacks the acyl-O-TE intermediate, replacing the oxygen of the serine and causing the release. In this way, the nucleophile determines the constitution of the product: if it is a) a water molecule, b) a hydroxy group, c) an amine group, d) located within the peptide intermediate, or e) within the next peptide intermediate, the final product is a) a carboxylic acid (linear peptide), b) ester, c) amide, d) cyclic (depsi)peptide or an e) oligomerized peptide.^[135,136] TE domains show a high structural diversity with an α/β hydrolase fold, consisting of 7 to 8 parallel β -sheets connected with α -helices and a size of approx. 240 to 290 amino acids. The catalytic triad His-Ser-Asp is highly conserved and the serine residue, which forms the center of the active site, is located in a GxSxG motif between the C-terminus of β -strand 5 and the N-terminus of the following α -helix.^[137] The active site is covered by a lid structure, composed of two α -helices between β -strands 6 and 7.^[138,139] The lid has a size of 40 to 90 amino acids with a high variability of sequence and structure between different TE domains. It is proposed that the lid works as a solvent shield, mediates interactions with the

nucleophile as well as the T domain, and leads the Ppant moiety into the active site of the TE domain.^[130] Crystallization approaches showed different conformations of the lid structure: the open conformation, enabling the interaction with the substrate-bound T domain Ppant arm and the nucleophile, and the lid helices forming a channel through the enzyme that restricts substrate and nucleophile entry.^[93,140] Although the lid structure is the largest source of sequence and size diversity between the TE domains, it cannot serve as a predictor of reactivity.^[81,122,138] Thus, neither the kind of substrates and nucleophiles accepted by the TE domains nor the release selectivity can currently be predicted based on the TE domain's structure or phylogenetic analyses.

2.4.1.5. Secondary/Modifying domains

In addition to the high substrate variability of A domains, the chemical complexity of NRPs is further enhanced by modifying domains. The most important ones for FR biosynthesis are mentioned in the following chapter.

2.4.1.5.1. Epimerization domain

Epimerization (E) domains, have a size of 50 kDa and are encoded between a T domain and a ^oC_L domain. They are related to C domains, show a similar catalytic motif as well as protein fold, and catalyze the epimerization of the T domain-bound monomer preceding the condensation.^[113,141] NRPs containing D-amino acids may confer beneficial properties like lower proteolytic susceptibility and more advantageous bioactive conformations.^[80]

2.4.1.5.2. Methyltransferase domain

Methyltransferase (MT) domains usually possess a size of approx. 45 kDa and are integrated into conserved motifs in the A_{core}-A_{sub} hinge domain.^[142] Most MT domains catalyze an *N*-methylation reaction through the transfer of a methyl group from S-adenosylmethionine (SAM) onto the backbone amide groups during the peptide assembly. They can be divided into C-MT and N-MT domains, which transfer the methyl group to a carbon or nitrogen atom, respectively.^[143]

2.5. FR900359 biosynthesis

The cyclic depsipeptide FR is formed by an NRPS system encoded by the *frs* BGC. It consists of eight genes encoding for a monomodular NRPS FrsA, a heptamodular NRPS FrsD to FrsG, the MbtH-like Protein FrsB, and the two modifying enzymes FrsC and FrsH (**Figure 9**). As is common for NRPS, each module activates and incorporates one specific amino acid into the growing peptide chain. Bioinformatic analysis revealed two type I TE domains, leading to the annotation of two distinct NRPS machineries: one for the assembly of the peptide core and the other for the formation and attachment of the side chain.^[20] FrsD is the first module of a heptamodular NRPS responsible for the assembly of the core molecule of FR. FrsD_A activates leucine, which is then attached to the adjacent FrsD T domain and hydroxylated by the non-heme diiron monooxygenase FrsH (this will be discussed in more detail below). FrsD possesses a C starter domain, which was shown to transfer an acetyl moiety onto hydroxyleucine. The resulting *N*-acetyl-OH-Leu is the first building block of the peptide, which is then condensed by the downstream C domain (FrsE3_C) to the next building block attached to its subsequent T domain (FrsE3_T). These steps are repeated throughout the whole NRPS assembly machinery, resulting in the elongation of the peptide chain. FrsE consists of two modules, of which FrsE3_A was predicted to activate and incorporate a carboxylic acid, based on its Stachelhaus code. It was postulated that FrsE3 activates and incorporates phenyllactate, which is not a common metabolic product and, thus, needs to be specifically provided for FR biosynthesis. As the modifying enzyme, FrsC shows a high sequential similarity to malate and L-lactate dehydrogenases, it is conceivable that FrsC catalyzes the reduction of phenylpyruvate derived from the phenylalanine metabolism, to phenyllactate (**Chapter 2.5.4**). Phenyllactate is then activated and incorporated by the A and T domain of FrsE3. In FR the phenyllactate is D-configured, although it is not yet clarified to what extent the epimerization domain of FrsE3 contributes to this process. It was proposed, that FrsE4 synthesizes *N*-methyldehydroalanine and introduces it into the growing peptide chain. This non-proteinogenic amino acid is a rare building block, whose formation is still not fully understood. Bioinformatic analysis suggests that the A domain activates serine, which is why it is assumed that FrsF5_C performs an 'online' dehydration of this serine residue to dehydroalanine, as known from the albopeptide^[144] and nocardicin NRPS^[145]. However, it should be emphasized that the FrsF5_C domain cannot be assigned to the C_{modAA} clade, as it might be expected, but phylogenetically belongs to the ^DC_L domains. FrsE4_{MT} then performs an *N*-methylation on this dehydroalanine (**Figure 10**). FrsF5 and FrsF6 incorporate L-alanine, however, module 6 contains an MT domain, resulting in the *N*-methylation of the second alanine, leading to the formation of the building block *N*-methylalanine. FrsG7_A shows a high similarity to the sequence of FrsD, therefore, it can be assumed that it activates L-leucine, which is then also hydroxylated by FrsH, leading to the incorporation of OH-Leu. The last module of the NRPS machinery possesses two MT domains and a TE domain besides the core domains, leading to the incorporation of *N*-O-dimethylated threonine and finally macrocyclization to the biosynthetic intermediate FR-Core.

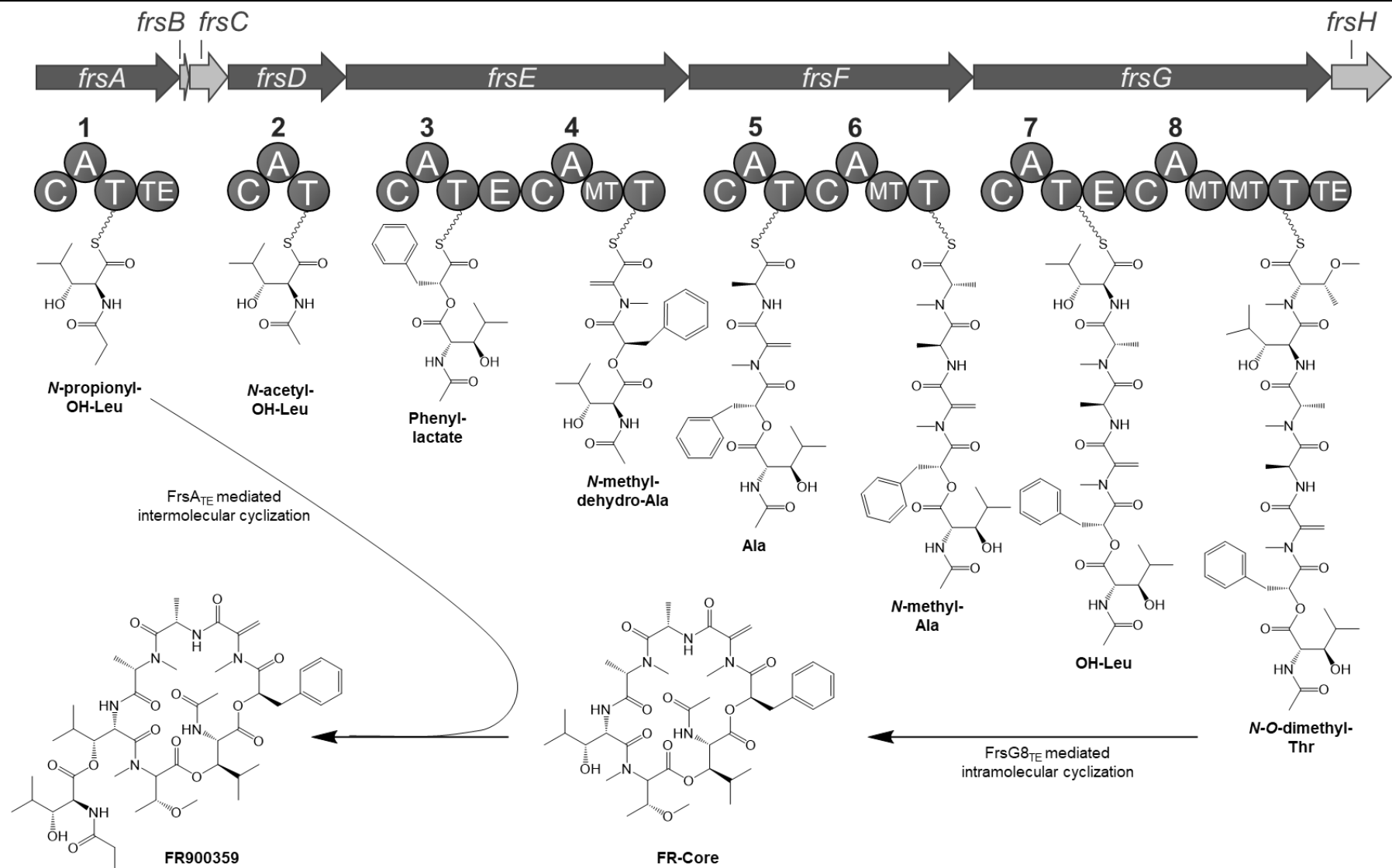


Figure 9: Biosynthesis of FR900359 (FR). Organization of *frs* biosynthetic gene cluster (dark gray = NRPS, light gray = modifying enzymes and auxiliary enzymes) and the encoded NRPS (C = condensation domain, A = adenylation domain, T = thiolation domain, TE = thioesterase domain, MT = methyltransferase domain, E = epimerization domain). The NRPS FrsD-FrsG form the linear peptide chain composed of seven building blocks (Leu = leucine, Ala = alanine, Thr = threonine), which is then cyclized by FrsG8_{TE}. The NRPS FrsA assembles the side chain *N*-propionyl-hydroxyleucine and FrsA_{TE} catalyzes its intermolecular transesterification to FR-Core, yielding FR.

2.5.1. The monomodular FrsA

Based on bioinformatic analyses, it was proposed that the monomodular NRPS FrsA assembles the side chain and attaches it onto the macrocyclic FR-Core. Experimental investigations to confirm the bioinformatic results were conducted by Hermes and Richarz *et al.*^[21] Utilizing a *frsA* deficient *C. vaccinii* knock-out (KO) mutant, it was shown that the formation of FR-Core is unimpeded in the absence of FrsA, the formation of the complete FR molecule, however, is abolished, leading to the accumulation of FR-Core. After confirming that FrsA is crucial for the incorporation of the side chain in FR, our group focused on the elucidation of the chronology and mechanism of the reactions catalyzed by FrsA. Through an γ - $^{18}\text{O}_4$ -ATP exchange assay^[146], L-leucine was identified as the favored substrate of the FrsA A domain, which nevertheless also moderately accepts D-leucine and isoleucine.^[147] These results contradict the tight substrate specificity of FrsD to L-leucine, which has a high structural similarity to FrsA (identity of amino acids of C-, A-, and T domain: 94.3%). The same holds true for the C domain of FrsA: Our group identified propionyl-Coenzyme A (CoA) as the preferred substrate, while further acyl-CoAs like acetyl-CoA and butyryl-CoA are also accepted with lower turnover rates. Preceding the acylation, the T domain-bound leucine-residue needs to be hydroxylated by the modifying enzyme FrsH. With *in vitro* assays, the side chain assembly could be sequentially elucidated (**Figure 10**): First, the A domain activates L-leucine and attaches it to the adjacent T domain. FrsH then hydroxylates the residue and, subsequently, the starter C domain attaches a propionyl group to the N-terminus. The final side chain is then attached onto FR-Core by the FrsA TE domain.

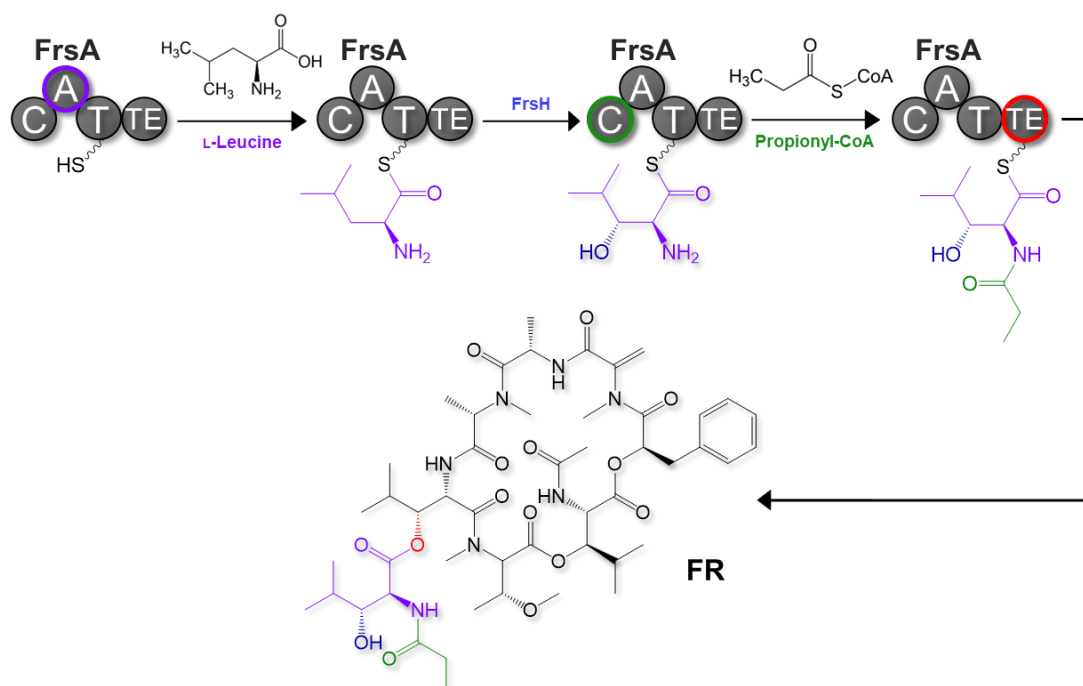


Figure 10: Reaction scheme of side chain formation and intermolecular transesterification onto FR-Core. The catalyzed reactions are color-coded according to the respective catalyzing domains. Adapted from Hermes *et al.*^[147]

Comparative studies revealed the major importance of the side chain for the Gq protein inhibitory potency of FR. In a cell-based assay, FR-Core showed a 16-fold reduced potency on Gq signaling compared to FR with an IC₅₀ value of 0.45 μ M.^[148] Additionally, binding assays with human platelet membranes demonstrated a 207-fold weaker binding affinity of FR-Core compared to FR.

Since FR originates from an endosymbiotic bacterium living in plant leaf nodules, it was generally assumed that it provides a selective advantage against predatory herbivores. Therefore, FR and FR-Core were fed to nymphs of the stinkbug *Riptortus pedestris*, with FR showing a higher *in vivo* toxicity. Bioinformatic studies were thus conducted to explain the deviations in potency.^[21] Docking-based models of FR to G α of mice revealed, that the building blocks *N,O*-dimethylthreonine, and D-phenyllactate both form hydrogen bonds with the side chain of R60 in G α , while the side chain of FR forms a strong interaction *via* two hydrogen bonds with the backbone of E191. These additional bonds explain the increased affinity and underline the extremely important function of the side chain and, thus of the incorporating NRPS FrsA for FR activity. A special focus needs to be on the TE domain of FrsA, which possesses the unusual, yet crucial function of catalyzing an intermolecular transesterification that has rarely been described in other NRPS systems^[149] and hardly been examined experimentally.

Investigations of the evolutionary origin of the TE domain confirmed the hypothesis that TE domains show a low sequence homology and broad substrate promiscuity, as no sequence homolog could be found for the FrsA_{TE} except for the FrsG8_{TE} with 42%, indicating an evolutionary connection of the two domains.^[21] Our group investigated the phylogenetic relationship between NRPS TE domains showing well-separated clades according to the mode of peptide release. The FrsA_{TE} can be found in the same clade as Sln9 and Dhbf_{TE}. The latter catalyzes trimerization in bacillibactin biosynthesis^[150], while Sln9 also catalyzes intermolecular transesterification in salinamide biosynthesis in *Streptomyces* bacteria.^[149] The mode of action and mechanism behind that reaction are yet to be understood and need to be explored.

2.5.2. MbtH like protein FrsB

The smallest gene in the *frs* BGC, *frsB*, encodes for an MbtH-like protein (MLP) (2.4.1.1).^[103] The chaperoning effect of MLPs increases the solubility of NRPS enzymes, and the interaction of MLPs with A domains was proven to be crucial for the A domain activity.^[105,108,109] The heterologous expression of FrsA and FrsC in presence and absence of FrsB revealed, that both needed coexpression with FrsB to be soluble and to exhibit adenylating activity on their substrates^[21]

2.5.3. Modifying enzyme FrsH

A widespread theme in NRPS biosynthesis are modifications on the T domain-bound amino acids performed by *cis*-acting domains encoded in the NRPS, like epimerization or methyltransferase domains.^[80] The hydroxylation of different leucine residues in FR biosynthesis, however, derives from the *trans*-acting β -hydroxylase FrsH (**Figure 11**).^[21] As already mentioned in the previous chapter, FrsH is of utmost importance for the correct formation of the side chain *N*-Pp-Hle, which is responsible for the strong affinity of FR to the Gq protein, a prerequisite for its potent inhibition.^[57] The attachment of a functional group, like a hydroxy group, to a molecule, is often not only decisive for its biological activity but makes the molecule susceptible to further modifications, e.g. transesterification, macrocyclization, and glycosylation.^[21,151] *Trans*-acting β -hydroxylases of NRPS systems can be classified into different enzyme families, including cytochrome P450-dependent monooxygenases (CYP450)^[152,153] and nonheme diiron monooxygenases (NHDMs).^[154] FrsH belongs to the sparsely characterized NHDMs. In NRPS systems, the correct assembly often depends on the correct protein-protein interactions between different domains.^[155] This could also be demonstrated in an *in vitro* experiment with a CYP450 from skyllamycin biosynthesis, showing that its correct function as a β -hydroxylase does not rely on the amino acid substrate but rather on the T domain to which the substrate is bound.^[153] One further reason for this could be the high substrate specificity of the A and C domains. Little is yet known about the underlying reaction mechanism of FrsH as a *trans*-acting β -hydroxylase, however, the domain has a high potential for being utilized for *de novo* NRPS engineering in the future.^[155]

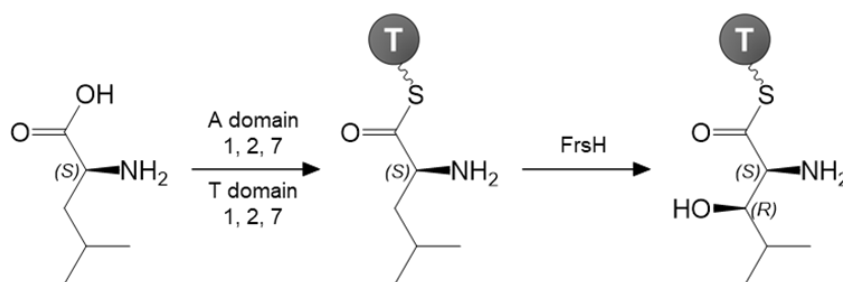


Figure 11: Proposed biosynthesis of the non-proteinogenic building blocks *N*-propionyl-(2*S*,3*R*)-3-hydroxy-L-leucine (*N*-Pp-Hle), *N*-acetylhydroxy-L-leucine (*N*-Ac-Hle), hydroxy-L-leucine (Hle).

2.5.4. Modifying enzyme FrsC

As already mentioned in **Chapter 2.5**, the phenyllactate (PLA) building block in FR and YM is of utmost importance for their inhibitory effects. Modification studies in the structurally similar YM revealed a 25 and 107-fold loss of Gq-inhibiting activity when D-PLA was substituted with D-leucine (YM-31) and phenylalanine (YM-3).^[156,157]

The biosynthetic model for FR, based on preliminary bioinformatic analysis, implies the activation and incorporation of L-PLA into FR by the A domain of FrsE3. Thus, we postulated the subsequent conversion to D-PLA catalyzed by the epimerization domain of FrsE3.

Regarding the importance of the presence of D-PLA in the final depsipeptide, the mechanism behind the incorporation and the biosynthetic origin of PLA is of great interest to be elucidated.^[15]

It was hypothesized that the modifying enzyme FrsC provides PLA, as its gene is located within the *cv_frs* BGC. A preliminary bioinformatic analysis classified FrsC as NAD(P)H-dependent dehydrogenase with high sequence similarity to malate dehydrogenases (MDH; EC 1.1.1.37)^[158,159] and lactate dehydrogenases (LDH; EC 1.1.1.27).^[160,161] Both enzymes derive from primary metabolism and catalyze the NAD(P)⁺/NAD(P)H-dependent oxidoreduction of 2-hydroxy and 2-keto carboxylic acids.^[162] Thus, it was assumed that FrsC catalyzes the NAD(P)H-dependent reduction of phenylpyruvic acid (PPA) to L-PLA, which is then accepted and incorporated by the A domain of FrsE3 and subsequently epimerized to D-PLA by the FrsE3 E domain.

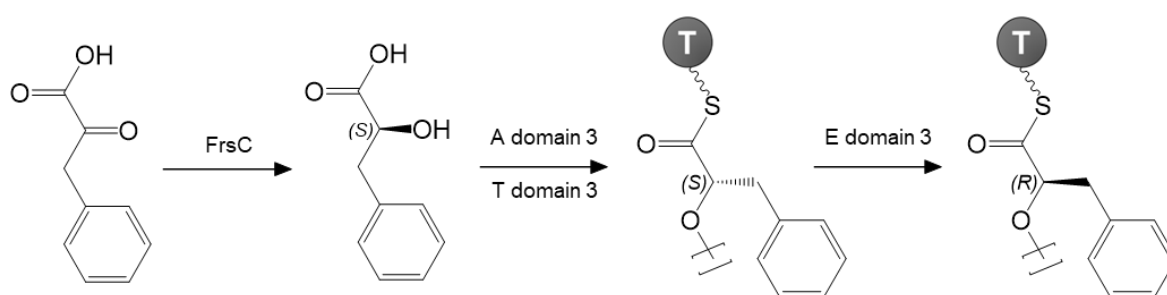


Figure 12: Proposed biosynthesis of the non-proteinogenic building block D-phenyllactate (D-PLA)

3. Aims of study

The goal of this thesis is the detailed biochemical and structural investigation of the non-ribosomal peptide synthetase system of FR900359 biosynthesis with focus on the unusual FrsA_{TE} and the modifying enzymes FrsC and FrsH.

FrsA_{TE}

The aim of this project is the detailed structural and biochemical characterization of the noncanonical FrsA thioesterase (TE) domain, catalyzing intermolecular transesterification of the FR side chain, *N*-propionylhydroxyleucine, onto the macrocyclic biosynthetic intermediate FR-Core. Not only is the catalyzed reaction scarce among all the NRPS systems known to date, but the side chain also plays a crucial role in the biological activity of the Gαq protein inhibitor FR.

Highlighting the relevance of FrsA_{TE}, its structure shall be elucidated to provide insights into the unusual reaction mechanism by analyzing its active site and the substrate binding. To achieve this, it is necessary to produce a construct of the TE domain in sufficient quantity and quality suitable for structural investigations using crystallization and X-ray analysis, as well as cryogenic electron microscopy (cryoEM).

To acquire detailed insights about its mode of action, specifically the dynamic interaction between FrsA_{TE} and its substrate FR-Core, 3D NMR measurements shall be conducted. For this purpose, labeled FrsA_{TE} has to be generated by thorough expression, purification, and storage optimization approaches. If a sufficient measurement can be performed, the enzyme dynamics during substrate binding should be investigated on this basis. This should lead to the elucidation of the reaction mechanism of this noncanonical TE domain and create the basis for further biosynthetic engineering studies.

Utilizing established *in vitro* assays, the kinetic parameters of the reaction catalyzed by FrsA_{TE} and its substrate specificity shall be investigated. Therefore, the substrates in the side chain assembly *in vitro* assay are to be exchanged, which could lead to the formation of altered side chains, which in turn can be used as substrate for the TE domain to generate FR derivatives with altered side chains *in vitro*. By constructing a FrsA and FrsG8_{TE} deficient *C. vaccinii* mutant, a linearized FR-Core shall be obtained, which could also serve as substrate for FrsA_{TE} to investigate its promiscuity for linear peptides. The detailed investigation of the promising catalyst FrsA_{TE} will provide insights into its potential for chemoenzymatic syntheses, e.g. the generation of altered, bioactive FR analogs with novel effects.

FrsH

Together with the monomodular NRPS FrsA, the nonheme diiron monooxygenases FrsH is responsible for the side chain assembly of *N*-propionylhydroxyleucine. Size exclusion-fast protein liquid chromatography (SEC-FPLC) and isothermal titration calorimetry (ITC) investigations of the protein-protein interaction revealed that an interaction is crucial for the hydroxylation reaction catalyzed by FrsH. Dr. Wirtz performed *in silico* studies with the elucidated FrsH structure and the FrsA_A and disclosed a possible binding interface.^[163] These results led to the identification of distinct residues, which potentially contribute to the interaction. The aim was to ascertain the relevance of these residues by targeted mutagenesis with subsequent analysis of product formation in the side chain assembly assay. These results could provide deeper insights into the A domain-guided recruitment of FrsH and hold great potential for future prospective investigations of NRPS protein-protein interactions.

FrsC

The specialized dehydrogenase FrsC from the FR biosynthesis and its catalyzed reaction shall be characterized in this work. This includes the detailed biochemical characterization of FrsC using the NAD(P)H-dependent activity assay, whereby the temperature and pH optimum, the substrate specificity, and the kinetic parameters are to be analyzed. This should provide insight into FrsC's relation to dehydrogenases from primary metabolism and its evolution to a highly specific dehydrogenase of aromatic substrates.

Furthermore, the configuration of the FrsC reaction product is to be determined by using high-performance liquid chromatography (HPLC) with a chiral column, biochemical characterization of FrsE3_A, and *in vivo* characterization of phenyllactate biosynthesis and incorporation by performing feeding experiments with the *frsC*-deficient *C. vaccinii* mutant.

4. Results and Discussion

4.1. Crystallization

Investigating the three-dimensional structures of proteins is of fundamental importance to provide insights into the basic principles of protein functioning. In addition, structure elucidation of enzymes is pivotal to advance protein engineering. Especially enzymes deviating from canonical structure-function paradigms offer unique catalytic mechanisms. Their structural analysis may provide insights into substrate specificity, stability determinants, and catalytic residues, enabling rational protein design and directed evolution strategies. Moreover, the detailed understanding of noncanonical enzyme architecture facilitates the engineering of enzyme complexes and the exploration of novel biocatalytic applications. Ultimately, structural characterization of these enzymes contributes to a deeper comprehension of evolutionary relationships and enzyme mechanisms, driving fundamental advancements in the field.^[164]

In the endeavor to elucidate the structure of the transesterifying FrsA_{TE} domain, the first crystallization attempts were carried out in our group by Dr. Cornelia Hermes in cooperation with Jonas Mühle of the Schertler workgroup at the Paul Scherrer Institute (Villigen, CH). First approaches using the sitting drop vapor diffusion method did not lead to the formation of crystals, thus subsequently, buffer optimization trials were conducted to increase protein stability and thermal resistance. This led to the formation of small crystals, which however were unsuitable for X-ray diffraction analysis. Optimization trials, using ‘additive screening’^[165] and random microseed matrix screening (rMMS)^[166–168], resulted in small rod-shaped crystals, which were still small, but applicable for the first X-ray diffraction analysis, showing some reflexes of FrsA_{TE} with a resolution of 9.2 Å. This resolution is insufficient for structure elucidation, but the results were quite promising for further crystallization approaches.

In this thesis, further optimizations of the expression, purification, and crystallization process were performed to obtain crystals suitable for X-ray diffraction analysis with a resolution sufficient to fully elucidate the structure of FrsA_{TE}. This was done by further optimizing the previous approach, by using crystallophores or protein constructs that are more appropriate for the sitting drop vapor diffusion crystallization method.

Many factors hampering the crystal formation of macromolecules have been described, comprising physical, chemical, and biochemical aspects.^[165] Many of these challenges can be overcome by supplementing the buffer system with so-called additives. These include physiologically or biochemically relevant small molecules, chemical protectants, solubilizing agents or detergents, poisons, osmolytes, protein stabilizing compounds, and compounds

enhancing nucleation. The protein itself is the most important and mostly unalterable variable in crystallization approaches.^[169] Once the improvement in purity of the protein and the search for crystal formation supportive additives have been exhausted, the only remaining option is the modification of the protein of interest (POI). There are different approaches in this regard: stabilizing unstable structures by macromolecular partners (e.g. substrate), combining the POI with a second protein (e.g. didomain of an NRPS module)^[170], or removing structures preventing or impeding crystal formation.^[165]

The following chapters describe the methods and approaches used to find new constructs of the POI, that could increase the success rate of crystallization and the quality of the FrsA_{TE} crystals.

4.1.1. Construct FrsA_{TE}Ed

4.1.1.1. Construct design by proteolysis

It is a frequently occurring phenomenon that protein crystals form only several months after starting the crystallization process. This is mostly attributed to the presence and activity of residual proteases in the batch, leading to spontaneous proteolysis. Thus, the native protein undergoes modifications and is more suitable for the crystallization process, due to the removal of structures that obstruct crystal formation.^[171-173] These observations are utilized to accelerate crystal formation by performing limited, controlled proteolysis of the POI. Either, the proteolysis approach can be used directly for crystallization, or proteolysis can also be carried out *in vitro* and analyzed first to investigate which fragments can be obtained and whether these are promising for a crystallization approach. Since *in vitro* proteolysis is quick, easy to carry out, and saves resources required for crystallization, we decided to perform proteolysis of FrsA_{TE} and to examine the resulting fragments. To this end, the FrsA_{TE} construct of Dr. Hermes was digested *in vitro* by a variety of proteases. The resulting fragments were then analyzed *via* sodium-dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Promising protein fragments were then extracted from the gels and analyzed using HPLC-MS to decipher their primary structure and to assess whether they were promising for a crystallization approach.

The FrsA_{TE} was heterologously expressed (6.6.1.1) and purified using affinity chromatography (6.6.2). The proteolysis was conducted with α -chemotrypsin, trypsin, elastase, and papain with varying concentrations (protease 1:10/ 1:100/ 1:1000 protein) (6.6.9). Additionally, FR-Core was added to the 1:100 approaches in the same molarity as the TE domain, aiming for the stabilization of the protein core. After incubation (1 h; 37 °C), the batches were analyzed with SDS-PAGE (6.6.8) to identify promising degradation patterns (Figure 13). The criteria for promising proteolysis

products are the formation of single bands that can be easily separated from other bands (especially the input band) in a molecular range between at least 23 to 30 kDa. In addition, the results should be reproducible to analyze the sequence coverage for identification of the remaining peptide.

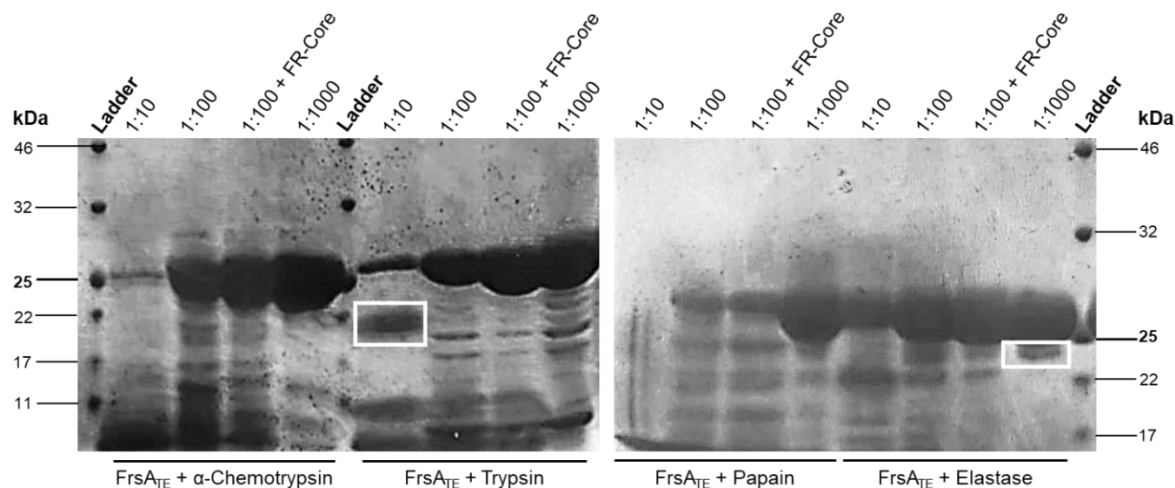


Figure 13: SDS-PAGE analysis of controlled proteolysis of His-TEV-FrsA_{TE} (30.4 kDa) with the proteases α-chemotrypsin, trypsin, papain, and elastase. Tested ratios were 1:10/ 1:100/ 1:1000 (protease:protein), while FR-Core was added to the 1:100 approaches in the same molarity as the TE domain. Applied volumes: proteolysis samples = 9 µL, ladder = 3 µL of the ladder.

Since the calculated size for FrsA_{TE} is 30.4 kDa, the input band should be detectable slightly below the height of the 32 kDa ladder band. However, the prominent band was slightly above the 25 kDa ladder band. A consistent discrepancy between calculated and observed molecular weights could be recognized for all analyzed proteins. This phenomenon was also discussed in other publications and can be attributed to the fact, that SDS preferably binds hydrophobic instead of negatively charged protein regions. This leads to a deviant migration mobility of chemically diverse proteins.^[174,175] Since it has already been proven that the proteins can be expressed by the corresponding plasmids and their activity has also been demonstrated, this effect will not be discussed any further.^[147] The following analysis therefore refers to the sizes of the protein fragments defined by the ladder, however, the actual size of the fragments is up to 5 kDa higher.

By using α-chemotrypsin in high concentrations, the analysis *via* SDS-PAGE revealed a near to complete digestion of FrsA_{TE}, as the band of approximately 25 kDa is barely visible. In the samples with a lower concentration of the protease, the protein band remains prominent, but there are also several bands distinctly below the 22 kDa ladder band visible. This indicated that the protein was digested into many fragments of minor sizes, which are not suitable for the search for new constructs of the intact core of FrsA_{TE}.

The trypsin digest in high concentration (1:10) revealed a promising band with a size of approx. 22 kDa. The lower concentrated trypsin digests showed numerous bands that were either too small in peptide size or appeared to be inseparable due to the quantity and distribution of the bands. Although this result is surprising, it was reproducible, and the distinct band could be analyzed for its sequence. The digests with elastase and papain showed a similar degradation pattern. Through the decreasing protease concentration, the number of bands also decreased. The most promising band, was right below the input band and the 25 kDa ladder band, corresponding to a size of approx. 26-28 kDa. This condition was investigated further since this was the only detectable band in the elastase 1:1000 sample, indicating a promising and easily separable peptide.

The trypsin 1:10 and the elastase 1:1000 approaches were repeated with varying temperatures (30 °C and 37 °C) and duration (1 h, 2 h, and 24 h), in presence or absence of the substrate FR-Core. The addition of FR-Core appeared to have indeed a stabilizing effect on the protein, as the resulting bands below 25 kDa were distinctly weaker than the respective bands in the digest without FR-Core. Since this was a desired effect and could prevent the loss of important structures, FR-Core was added to any further experiment.

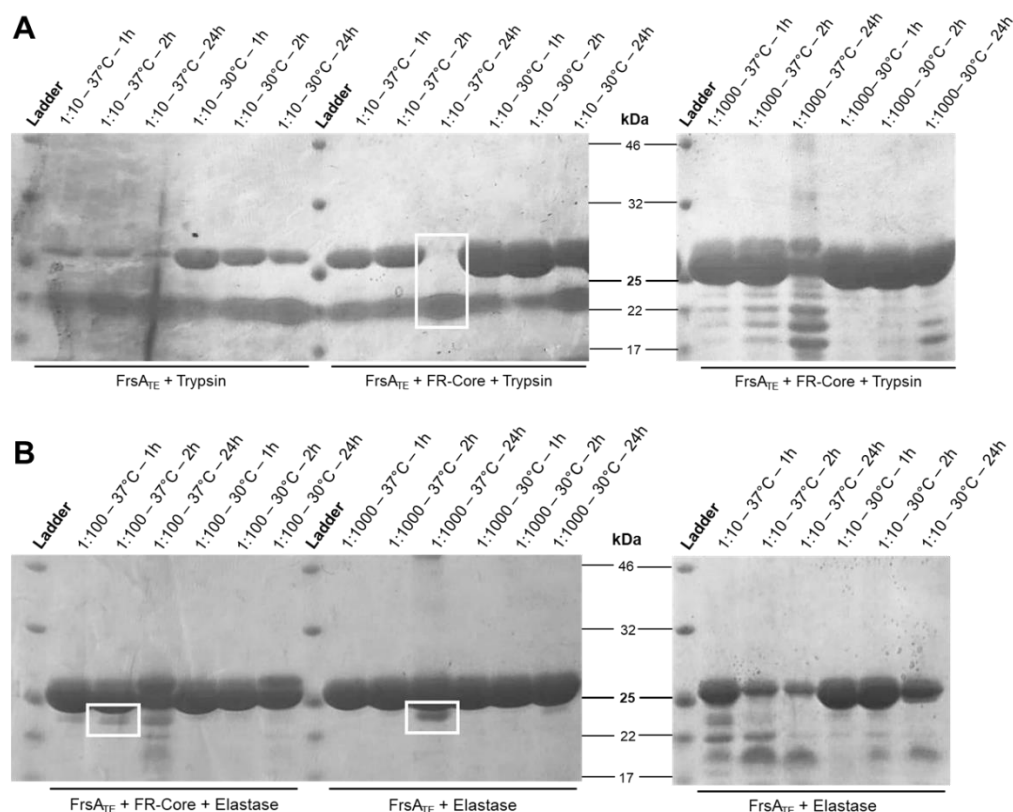


Figure 14: SDS-PAGE analysis of controlled proteolysis of His-TEV-FrsA_{TE} (30.4 kDa) with the proteases **(A)** trypsin in the ratios 1:10 and 1:1000 (protease:protein), and **(B)** elastase in the ratios 1:10, 1:100, 1:1000 (protease:protein) under varying incubation conditions (30/37 °C; 1/2/24 h). FR-Core was added in the same molarity as the TE domain. Applied volumes: proteolysis samples = 10 µL, ladder = 3 µL of the ladder.

Subsequent experiments attempted to optimize the digests with trypsin and elastase, aiming for a clear and separable band right below the native FrsA_{TE} band (**Figure 14**). For the trypsin digest, it was possible to digest the whole added FrsA_{TE} at 37 °C for 24 h with a concentration of 1:10 + FR-Core, as the prominent band of native FrsA_{TE} at 25 kDa was no longer detectable. However, the fragment at 22 kDa was evident. For the digest with elastase, the determined optimal conditions were at 37 °C for 2 h and a concentration of 1:100 + FR-Core or at 37 °C and 24 h with a concentration of 1:1000 FR-Core. The upper band exhibited a high intensity, whereas the lower band was less prominent and exhibited proximity to the input band. While these results suggested minimal structural degradation, they also indicated potential challenges in achieving complete band separation for subsequent HPLC-MS analysis. Therefore, the addition of FR-Core was maintained for both digest samples.

The digests were repeated under optimal conditions (FrsA_{TE} + FR-Core + elastase (1:100) incubated at 37 °C for 2 h; FrsA_{TE} + FR-Core + trypsin (1:100) incubated at 37 °C for 24 h) and applied to a preparative SDS-PAGE (15% acrylamide). The use of a higher concentrated gel results in better separation in the lower molecular weight range. After the SDS-PAGE the bands of interest were excised, whereby meticulous care was taken not to transfer anything from the input protein band. The gel piece containing the digested protein was lyophilized and analyzed *via* HPLC-MS in the Core Facility Analytical Proteomics of the University of Bonn (**6.6.10**). Here, electrospray ionization-liquid chromatography-mass spectrometry (ESI-LC/MS) was used to identify the protein sequence coverage after proteolytic digestion. The results are depicted in **Figure 15**.



Figure 15: *Via* HPLC-MS determined sequence coverage of FrsA_{TE} after proteolytic digestion with (A) elastase (1:100, FR-Core, 37 °C, 2 h) and (B) trypsin (1:100, FR-Core, 37 °C, 24 h). Sequence coverage was determined by peptide identification and assignment to the FrsA_{TE} protein sequence. Green highlighted regions indicate detected amino acid residues.

The elastase digestion product yielded a high recovery rate of 48.72% of the FrsA_{TE} sequence. While the C-terminus was fully sequenced, the N-terminal region, comprising the first 27 amino acids and the His-tag, was absent from the MS data. The decrease in the molecular weight correlates approximately with the proteolytically cleaved amino acids. This result led to the conclusion that the elastase protease removed accessible protein structures at the N-terminus, which supports the assumption that these structures are not highly conserved.

The trypsin digest had a low abundance and only one peptide could be found, which could be assigned to amino acids 168-181. This sequence coverage of only 5.13% suggested, that the excised band only contains an accumulation of peptides, and that the protease trypsin digests FrsA_{TE} completely, making it ineligible for the search of new constructs.

4.1.1.2. Expression and Purification

The amino acid sequence for the first construct was named FrsA_{TE}Ed, derived from the result of the elastase digest. The construct was generated and provided by Dr. Hermes. The truncated sequence of FrsA_{TE} (Appendix **Table 22**), was cloned into the pHis8-TEV plasmid, provided by Dr. René Richarz, *via* ligation as described in **Chapter 6.5**, using the gDNA of *C. vaccinii* as template and the *Bam*HI and *Hind*III restriction sites. After the sequence verification by Sanger sequencing, the final plasmid was transformed into the host strain *E. coli* BL21(DE3) for protein expression.

Expression and purification of FrsA_{TE}Ed was analyzed *via* SDS-PAGE (**6.6.1-6.6.8**), which is shown in **Figure 16**.

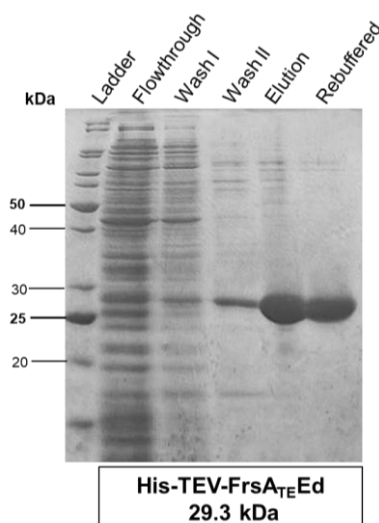


Figure 16: SDS-PAGE of His-TEV-FrsA_{TE}Ed purification after expression in *E. coli* BL21. Expected size for His-TEV-FrsA_{TE}Ed = 29.3 kDa. Eluted protein is rebuffed in TE domain buffer. Applied volumes: purification fractions = 9 μ L, ladder = 3 μ L.

After the purification process, a clear band right below the 30 kDa ladder band could be detected, which corresponds to the His8-TEV-FrsA_{TE}Ed construct with a calculated size of 30.4 kDa, leading to the conclusion, that the expression was successful, and the purification led to a high amount of His8-TEV-FrsA_{TE}Ed. However, the SDS-PAGE revealed a minor loss of the protein, while the purification process as the Wash I and the Wash II fractions, also showed comparatively weak bands at the same height as the protein band in the Elution sample. As the POI yield was still sufficient (68 mg/1 L expression culture), the loss was neglected at this point. The detected impurities of high molecular weight in the Elution sample could be separated during the re-buffering process, resulting in a high purity of the final sample. The purified protein was therefore considered for further processing.

Before the crystallization trials, the His-Tag was removed by targeted cleavage with a tobacco etch virus (TEV) protease (6.6.6). This protease recognizes the TEV cleavage site, located between the POI and the His-Tag, and cuts within this specific sequence (ENLYFQ\S). After tag removal, the protein was further purified *via* reverse affinity chromatography, whereby the undigested protein remains in the resin, while the digested protein without His-Tag passes through the column.

All purification samples were analyzed by SDS-PAGE (**Figure 17A**). The digest mix showed two prominent bands in the relevant molecular weight range. The upper band of approx. size of 29 kDa corresponds to the His-TEV-FrsA_{TE}Ed construct, whereby the lower band may originate from FrsA_{TE}Ed or the protease, as they have the same size of 27 kDa. The Flowthrough and the Wash I samples showed a high purity, only containing the desired protein band, which is why they were pooled. To ensure the complete removal of the TEV protease, the reverse affinity chromatography was repeated, whereby the amount of Nickel-nitrilotriacetic acid (Ni-NTA) was increased and the incubation time was extended to 1.5 h (6.6.6). As the Elution fraction indeed showed a band, it could be assumed that the protease could be separated from the POI without His-tag. The pooled Flowthrough 2 and Wash 2 fractions revealed a high purity, were concentrated (8.5 mg/1 L expression culture), and applied for further purification *via* preparative SEC-FPLC (6.6.7).

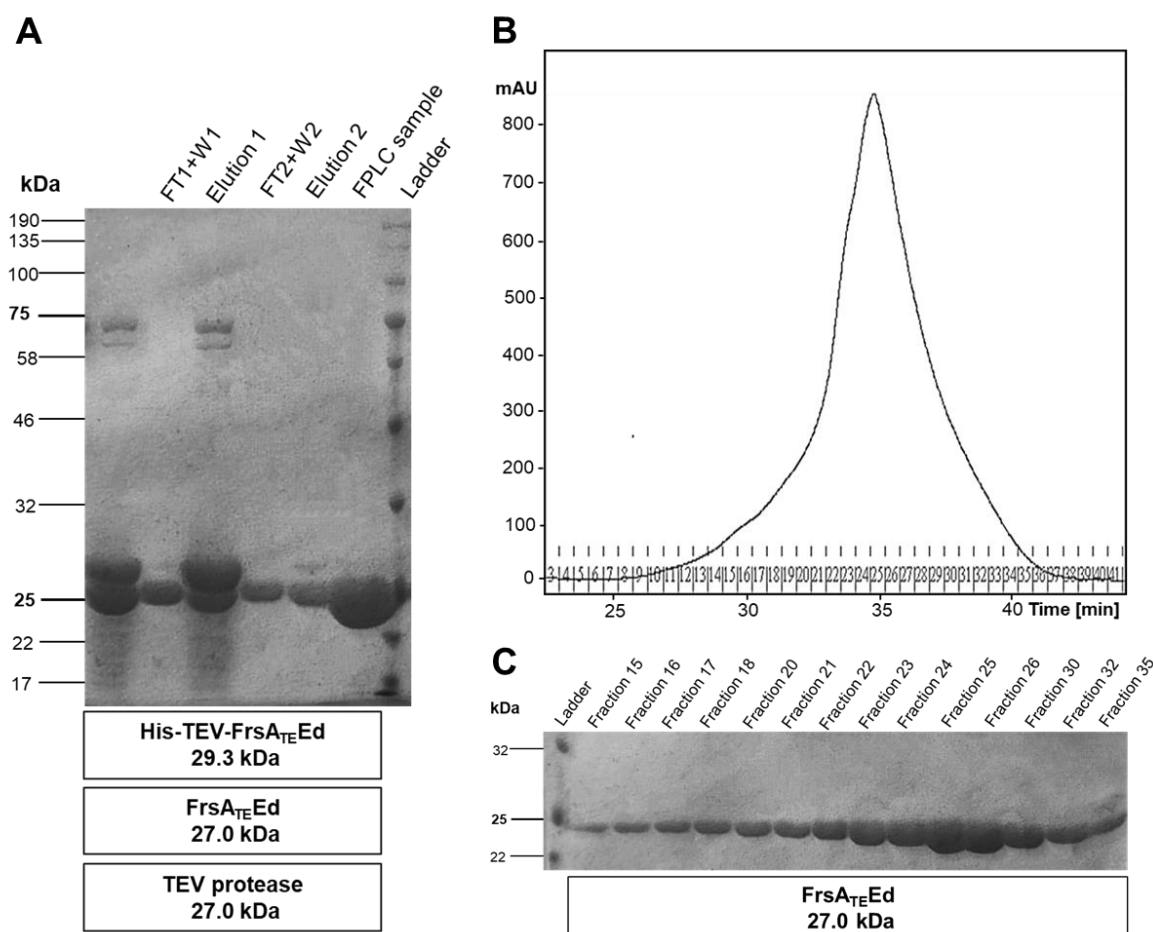


Figure 17: Sample processing of His-TEV-FrsA_{TE}Ed for crystallization approach, involving **(A)** SDS-PAGE analysis of reverse purification after His-tag removal, **(B)** SEC-FPLC chromatogram of FrsA_{TE}Ed, and **(C)** SDS-PAGE of the SEC-FPLC fractions. Expected sizes: His-FrsA_{TE}ED = 29.3 kDa, FrsA_{TE}Ed = 27.0 kDa, TEV protease = 27.0 kDa. FT=Flowthrough, W=Wash, 1=fractions of first reversed Ni-NTA purification, 2=fractions of second reversed Ni-NTA purification. Applied volumes: purification fractions = 15 μ L, ladder = 3 μ L.

The SEC-chromatogram of the FrsA_{TE}Ed purification run (**Figure 17B**) exhibited a prominent peak at a retention time of approx. 34 min. It was assumed that the peak corresponded to the desired protein, as the input sample had a high purity. However, the peak possessed an unusual appearance, comprising two shoulders in the front region. SDS-PAGE analysis of the FPLC fractions revealed however one prominent band of high intensity and the desired size, indicating high purity and protein yield (**Figure 17C**). The shoulder peaks could derive from dimers, which could not be distinguished from the monomer using SDS-PAGE. The final protein sample was of suitable concentration (~4 mg/1 L expression culture) and sufficient purity for crystallization approaches.

4.1.1.3. Crystallization trials

The crystallization trials of FrsA_{TE}Ed were performed by Jonas Mühle of the Schertler group at the Paul-Scherrer-Institute (PSI; Schweiz) as described in 4.1. By using the sitting drop vapor diffusion method, the protein solution (10 mg/mL) was mixed 1:1 with the precipitant solution in a drop above a precipitant-filled reservoir. For the first trial, a standard screening of crystallization buffers was conducted, not leading to any crystal formation. Moreover, the same conditions resulting in the FrsA_{TE} crystals were tested for FrsA_{TE}Ed, which also did not lead to crystal formation. Further crystallization additive screenings, were also not successful, leading to the conclusion, that the newly generated construct is not suitable for the crystallization process as the previous crystallization successes could not be reproduced and certainly not improved.

4.1.2. Constructs FrsA_{TE}short and FrsA_{TE}long

4.1.2.1. Construct design

To get a deeper insight into the structure of FrsA and to confirm the results of the proteolysis product, bioinformatic analyses were performed. Therefore, the structure of the whole monomodular NRPS FrsA was predicted using I-TASSER (6.9.2). The single domains were colored according to the antiSMASH (6.9.1) identification and annotation of modules and domains of the whole *frs* gene cluster (Figure 18).

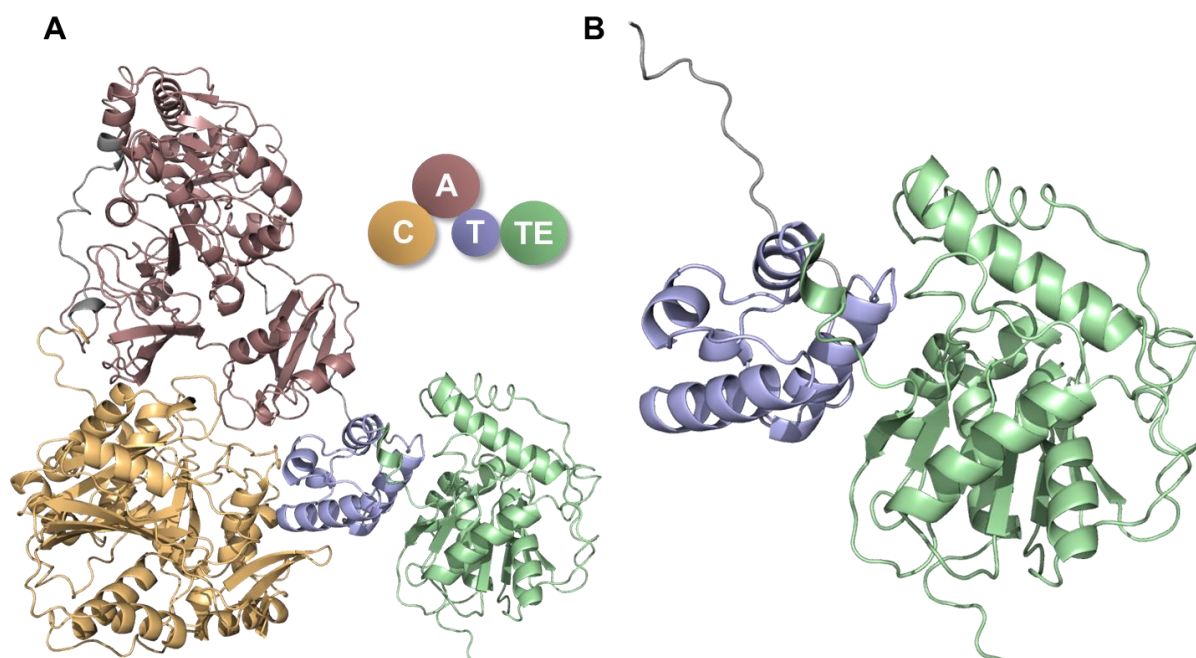


Figure 18: I-TASSER-predicted structures based on antiSMASH annotation of **(A)** the monomodular NRPS FrsA (color-coded: yellow = condensation (C) domain, red = adenylation (A) domain, blue = thiolation (T) domain, green = thioesterase (TE) domain, gray = linker regions) and **(B)** the T and TE didomain (color-coded: blue = thiolation (T) domain, green = thioesterase (TE) domain, gray = linker regions).

The automatic annotation of the domains of FrsA by antiSMASH corresponds to the expected core domains C- (blue), A- (red), T- (yellow), and TE-domain (green). Furthermore, hinge or linker regions were predicted between the C and the A domain, as well as between the A and the T domain. Conspicuously, this linker region was not considered between the T and the TE domain. The TE domain sequence is predicted to start immediately after the last α -helix of the T domain. The crystallization approaches of FrsA_{TE} and FrsA_{TE}Ed were conducted in the presence of these long N-terminus, not harboring any structure or condensed motif. Therefore, it was reasonable to assume that the structure of the 25 N-terminal amino acids is the linker region, which possesses high flexibility/low stability and might prevent crystal formation.

For confirmation of this assumption, the predicted structures of FrsA_{TE} and FrsG8_{TE} were compared. It is a well-known fact that NRPS systems develop evolutionarily by duplication, resulting in many domains showing high similarity in terms of sequence and conserved motifs.^[176,177]

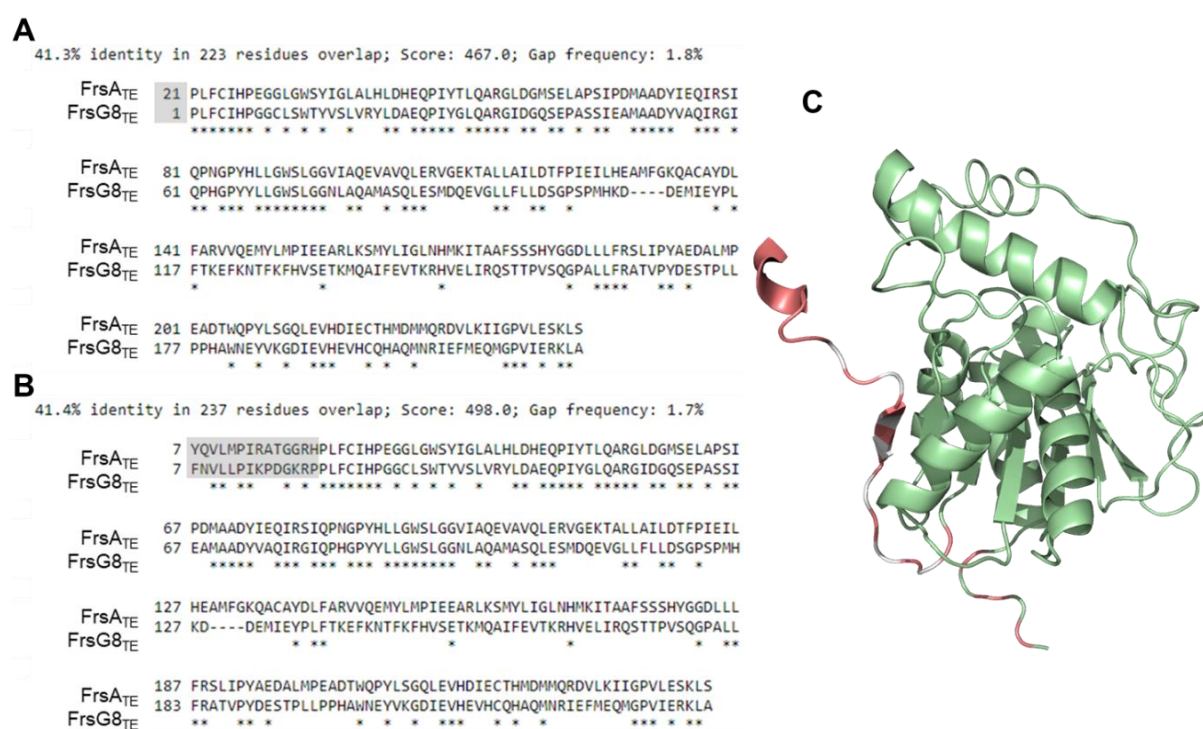


Figure 19: Expaty sequence alignment of FrsA_{TE} and FrsG8_{TE} **(A)** based on the annotation of FrsG8_{TE} excluding the upstream sequence and **(B)** and based on the annotation of FrsA_{TE} including the upstream region of FrsG8_{TE}. Highlighted in grey are the amino acids, which were not assigned to the FrsG8_{TE} sequence. **(C)** shows the I-TASSER predicted FrsA_{TE} structure (color-coded according to construct design based on the sequence comparison with FrsG8_{TE}, green = retained structure, red = truncated structure in FrsA_{TE}short (lacking the N-terminus up to the first β -fold structure (G1-H20) and the C-terminus right after the last α -helix structure (V244-Q249), white = truncated structure in FrsA_{TE}long lacking the N-terminus (G1-H20)).

The alignment of the automatically annotated sequences of FrsA_{TE} and FrsG8_{TE} using the ExPASy Alignment tool (6.9.4) revealed a sequence identity of 41.4%, highlighting the presumed similarity between both domains. Particularly worth mentioning, is that the automatically annotated FrsG8_{TE} aa sequence is several amino acids shorter than FrsA_{TE} (Figure 19A). If the approx. 20 missing amino acids upstream of FrsG8_{TE} are included in the alignment, it is noticeable that 14 of them still have a high similarity to FrsA_{TE} (Figure 19B), but the six amino acids further upstream could not be aligned. This analysis led to two conclusions: (1) *frsA_{TE}* is probably incorrectly annotated and comprises the flexible linker region between FrsA_T and FrsA_{TE} and (2) the bioinformatical analysis verified the proteolysis result, showing that the first 21 amino acids are easily accessible and targetable for proteases, equally confirming the first conclusion.

In addition, the sequences of FrsA_{TE} and FrsG8_{TE} showed a high variability in the C-terminal sequence. These have approx. the same length, but the predicted structures showed high differences, whereby it must be emphasized that the C-terminus of the FrsA_{TE} seemed to be disordered. The underlying assumption was that the TE domain of FrsA evolutionary arose from a duplication event with FrsG8_{TE} domain and then underwent structural changes, especially in the C-terminal region. Since the core structures of TE domains are known and it could be concluded that the termini do not harbor any catalytic activity but high flexibility, two constructs were designed (Figure 19C): FrsA_{TE}short, lacking the N-terminus up to the first β -fold structure (G1-H20) and the C-terminus right after the last α -helix structure (V244-Q249), and FrsA_{TE}long, only lacking the N-terminus (G1-H20).

4.1.2.2. Cloning

The cloning of the constructs FrsA_{TE}short/long was performed using the Gibson assembly method as described in Chapter 6.5.8. Briefly, the gene of interest (GOI) was amplified by Q5 PCR using the gDNA of *C. vaccinii* as template and specific primers. The latter also comprised a sequence complementary to the plasmid pHis8-TEV (provided by Dr. René Richarz). The resulting plasmid was transformed into an *E. coli* α SS strain for amplification, re-isolated and the sequence was verified by Sanger sequencing. The plasmid was then transformed into the expression strain *E. coli* BL21 (6.5).

In the process of further expression and purification optimizations, the genes encoding for the fusion proteins MBP and mVenus were inserted *via* Gibson Assembly into the already generated pHis8-TEV-FrsA_{TE}short/long plasmid (4.1.2.5). Therefore, the genes of the fusion proteins were amplified using the plasmids pET28a:MBP-TEV-PadeA and pCDF:His6-mVenus provided by Dr. Julian Hegemann. The sequences were inserted between the His-tag and the TEV cleavage site

sequence. Downstream of the latter, a sequence encoding for a 'GGGSGGG' linker region was embedded to ensure that the protease has access to its specific recognition site for its correct function. Thus, the primers for amplifying the inserts contained complementary regions for the pHis8-TEV-FrsA_{TE}short/long plasmid and the linker regions, downstream the genes of the fusion proteins. The Gibson assembly experiments of these approaches led to the generation of the plasmids pHis8-MBP-TEV-Linker-FrsA_{TE}short/long and pHis8-mVenus-TEV-Linker-FrsA_{TE}short/long, which were then transformed into chemically competent *E. coli* αSS cells for amplification. After re-isolation and sequence verification *via* Sanger sequencing, the plasmids were transformed into the expression strain *E. coli* LOBSTR (6.4).

4.1.2.3. Expression and Purification

SDS-PAGE analysis of the first expression and purification (6.6.1-6.6.8) attempt of His8-TEV-FrsA_{TE}short and His8-TEV-FrsA_{TE}long revealed a weak band in the Elution and the final product sample (Figure 20). In both cases, the band indicated the desired size of 24.9 and 25.5 kDa, as it is slightly below the 25 kDa ladder band. Furthermore, the purity seemed sufficient because the Elution sample exhibited a high purity, without detectable unspecific protein bands. In the first wash step, most proteins were removed from the protein mix so that the second wash step showed only hardly visible bands.

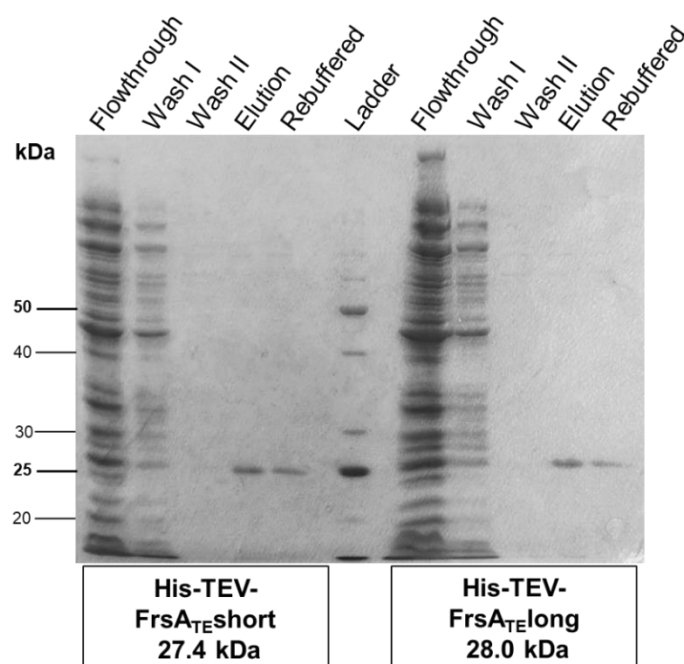


Figure 20: SDS-PAGE of His-TEV-FrsA_{TE}short (left, expected size = 27.4 kDa) and His-TEV-FrsA_{TE}long (right, expected size = 28.0 kDa) purification after expression in *E. coli* LOBSTR. Eluted protein is rebuffed in TE domain buffer. Applied volumes: purification samples = 15 µL, ladder = 3 µL.

The bands of the Elution and the rebuffered sample could be assigned to the POI, however, the weak intensity was a sign of the low yield, which was confirmed *via* photometric concentration measurement (0.2 mg/100 mL expression culture). Assuming a high protein loss of 50-75% due to His-tag removal, reverse affinity chromatography, and subsequent SEC-FPLC, the expression should yield high amounts for compensation (>50 mg/1 L expression culture). So, it could be concluded that although the purification was sufficient and the desired protein was expressed, the yield was not suitable for crystallization attempts, as high concentrations of at least 10 mg/mL are needed.

4.1.2.4. Activity test

Since the truncated constructs, FrsA_{TE}short/long could be successfully expressed and purified (6.6.1-6.6.4), their catalytic activity was investigated in advance of further optimization approaches regarding their expression yield. The demonstration of persistent activity is important to prove that the removed structures do not play a crucial role in protein-protein interaction and interfere with the catalytic activity.

Dr. Cornelia Hermes and Dr. Daniel Wirtz established the side chain assembly assay in which the enzymes FrsA_{CAT} (10 μ M) and FrsH (10 μ M) are incubated with their cofactors and substrates (Methyl viologen, NADH, L-leucine, Propionyl-CoA, ATP) to assemble the side chain *in vitro*. This assay mix then provides the substrate, *N*-propionylhydroxyleucine, for the TE domain assay, which is used to investigate the transesterifying activity of the TE domain (6.7.1, 6.7.2). To perform the side chain assembly assay, also the (multi)enzymes FrsA_{CAT} and FrsH had to be expressed and purified. The used *E. coli* expression strains were prepared and provided by Dr. Wirtz and Dr. Hermes.

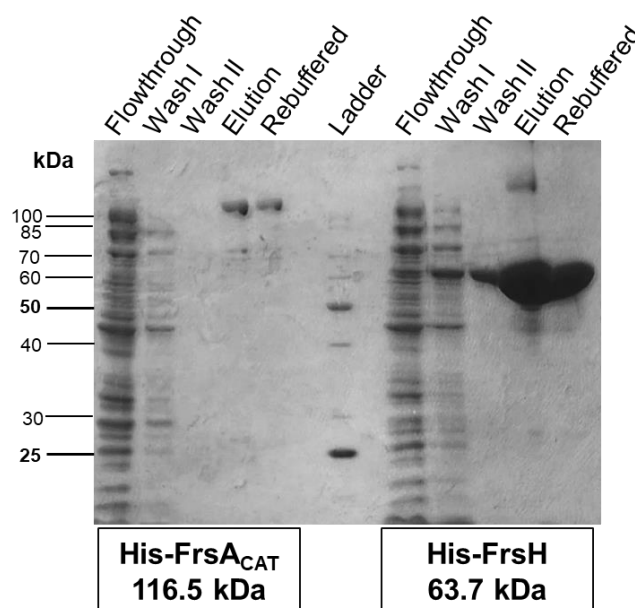


Figure 21: SDS-PAGE of His-FrsA_{CAT} (left, expected size = 116.5 kDa, expressed in *E. coli* BAPI) and His-FrsH (right, expected size = 63.7 kDa, expressed in *E. coli* BL21) purification. Eluted protein is rebuffered in C domain assay buffer. Applied volume: Flowthrough and Wash samples = 15 μ L, ladder = 3 μ L. As a high protein concentration was expected in the Elution and Rebuffered sample, the application volume was adjusted to 5 μ L.

The SDS-PAGE analysis (6.6.8) of the FrsA_{CAT} purification process (Figure 21) revealed a clear band above the 100 kDa ladder band, according to the expected size of 116 kDa. The wash steps seemed to be effective, as nearly every unspecific band could be removed, resulting in a high purity of the final samples. The yield was about 6 mg/250 mL expression culture.

As the MLP FrsB is crucial for the correct function of the A domain within the purified FrsA_{CAT}, they were coexpressed *via* an additional plasmid in the same expression strain. However, FrsB could not be evidenced by SDS-PAGE analysis, as it has a size of approximately 8 kDa and the run time was extended to achieve good separation in the higher molecular weight ranges, resulting in the loss of lower molecular size. Nevertheless, previous experiments could show that the A domains exhibited their amino acid activating activity, leading to the assumption of the presence of FrsB in the assay mixture.^[147]

The samples of the FrsH purification fractions (Figure 21) showed a prominent band at the height of the 60 kDa ladder band, corresponding to the predicted size of the desired protein. The FrsH expression yielded 11 mg/250 mL expression culture. While the Elution fraction showed some unspecific bands, the re-buffering process removed the contaminant proteins of higher molecular weight. The purity of the final sample was considered sufficient to perform the assay.

The second part of the reaction is started by the addition of FrsA_{TE} (5 μ M) and its nucleophilic substrate, FR-Core, is added to the side chain assembly assay mix. Every approach was performed

in triplicates, and for each set of purified proteins, a positive control was prepared with an unmodified FrsA_{TE} domain.

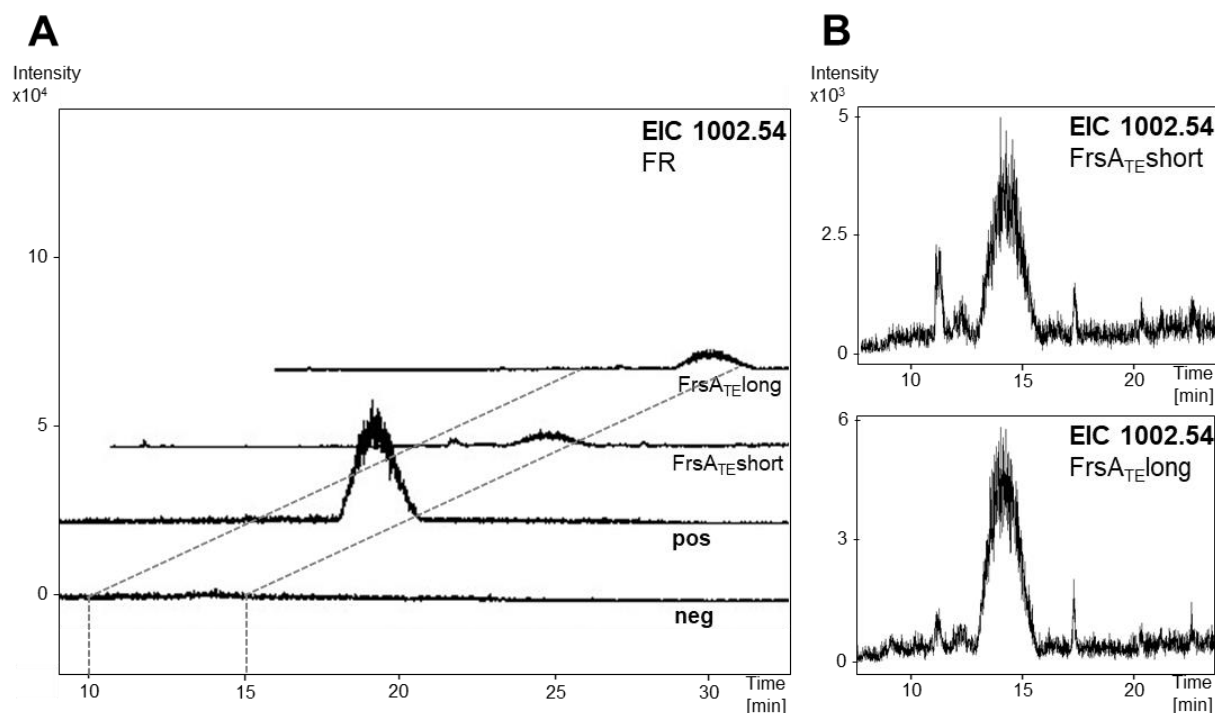


Figure 22: Extracted ion chromatograms (m/z 1002.54) of the HPLC-MS measurements of the TE domain activity assays. **(A)** EIC comparison (m/z 1002.54) of negative control (neg), positive control (pos), and the assays conducted with FrsA_{TE}short and FrsA_{TE}long. **(B)** shows the same EICs of the assays conducted with FrsA_{TE}short and FrsA_{TE}long as in (A) focusing the product peak. All assays were performed under the same conditions (Chapters 6.7.1/6.7.2) only differing in the used TE domain (pos = native FrsA_{TE}, neg = heat-inactivated native FrsA_{TE})

The results depicted in **Figure 22** verified the remaining activity of FrsA_{TE}short and FrsA_{TE}long, as in both cases a measurable peak could be detected at the retention time and m/z of FR, also matching the typical shape of the FR peak. However, it should be mentioned that the catalytic activity turns out to be rather low, as depicted in **Figure 22A**. The peak intensities of the measurements of the assays performed with the modified enzymes showed only a fractional amount of the positive control, which allows direct conclusions about the activity. What led to the activity loss could only be assumed so far. One possibility is the diminished interaction between the T and the TE domain, as the truncated region might be the linker, enabling the contact and proper positioning between both domains. The catalytic center of the core protein appeared to be complete and functional, as FR as the reaction product was detectable, which was not the case in the negative control, where the TE domain was heat-inactivated. Thus, the constructs FrsA_{TE}short and FrsA_{TE}long were rated as suitable for crystallization approaches, if the optimization of expression and purification proved successful.

4.1.2.5. Optimization approaches

The low yield of FrsA_{TEshort}/long appears to be due to reduced solubility since a prominent band could also be detected in the pellet sample (not shown). Therefore, the first optimization attempts focused on the support of protein folding and aggregation prevention by utilizing fusion proteins (FPs). These are often multifunctional and thus a simple and practical solution to overcome the occurring hindrances of expression and purification. They are utilized to improve the heterologous protein expression in *E. coli*, facilitating proper protein folding through a chaperon-like behavior. Furthermore, some FPs, like the maltose binding protein (MBP), additionally give the opportunity to serve as purification tag itself. Moreover, some FPs like mVenus possess fluorescence, enabling the observation of the POI in living cells or a simplified expression and purification optimization, which is of more importance for this work.

4.1.2.5.1. Maltose Binding Protein Fusion tag

The maltose binding protein (MBP) derives from *E. coli* and is a highly soluble protein of 43 kDa^[178–180], playing an important role in the translocation of maltose and maltodextrins^[181]. It possesses a chaperone intrinsic activity, leading to an increased solubility of the tagged protein when used as fusion protein tag. It supports proper protein folding, correlated with a large hydrophobic area on its surface. Furthermore, it can be used as purification tag, because of its ability to bind immobilized amylose resins, enabling a separation of the POI and proteins without MBP fusion tag.^[182]

The gene encoding the MBP was inserted into the existing plasmid pHis-TEV-FrsA_{TEshort}/long. The initial test expressions revealed an increased yield (~0.8 mg/100 mL expression culture), which could be further increased by purification optimization to ~2 mg/100 mL expression culture in both cases (Appendix **Figure 68**). Upscaling experiments with 100/250/500 mL expression cultures led to the conclusion that a 100 mL volume was most suitable; hence this culture volume was used for all further expressions. Unfortunately, the SDS-PAGE analysis (Appendix **Figure 69** and **Figure 70**) showed that the removal of the His-tag using the TEV digest was not sufficient. It was assumed that the protease did not have space for proper binding at the cleavage site due to the fusion protein. Thus, a single ('GGGS') and a double ('GGGSGGGS'; Linker²/L²) linker was inserted between the TEV cleavage site and the POI. The insertion of a 'GGGSGGGS' linker resulted in a significant increase in digested protein, hence it was retained for all following experiments. The purification *via* Ni-NTA affinity chromatography before and after the removal of the His-tag is depicted in **Figure 23**.

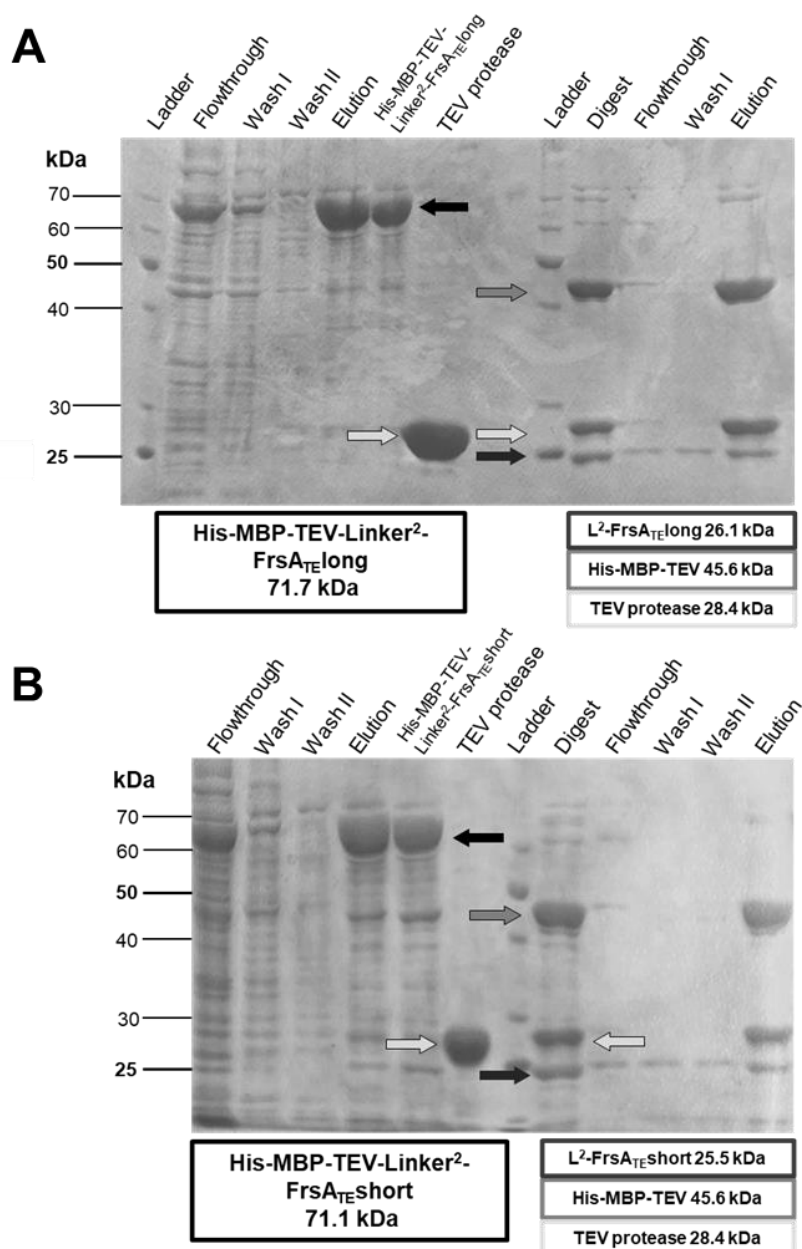


Figure 23: SDS-PAGE analysis of the expression, purification, and reverse affinity chromatography after His-tag removal of the constructs with MBP fusion tag **(A)** FrsA_{TE}long and **(B)** FrsA_{TE}short. On the left side the Ni-NTA affinity purification after expression in *E. coli* LOBSTR and on the right side the reverse purification after His-tag removal. Expected sizes: His-MBP-TEV-Linker²-FrsA_{TE}long = 71.7 kDa, Linker²=FrsA_{TE}long = 26.1 kDa, His-MBP-TEV-Linker²-FrsA_{TE}short = 71.1 kDa, Linker²=FrsA_{TE}short = 25.5 kDa, His-MBP-TEV = 45.6 kDa, TEV protease = 28.4 kDa. Color-coded arrows highlight the expected protein bands. Applied volumes: Flowthrough, Wash, digest samples = 15 μ L elution and final protein samples = 10 μ L, ladder = 4 μ L.

The analysis *via* SDS-PAGE (6.6.8) revealed that the expression of the His-MBP-TEV-Linker²-FrsA_{TE}short/long was successful for both FrsA_{TE} constructs. A distinct band can be found in the Elution and the Rebuffered sample at a height between the 60 and 70 kDa ladder band. The POI has a size of 71.1/71.7 kDa, which corresponds to the result, considering that the protein bands run slightly faster than the ladder. Regarding the purification, it should be noted, that there was a marginal loss of POI, but the rebuffered sample also showed a multitude of unspecific protein

bands. Therefore, the volume of the first wash step was adjusted to 8 mL, leading to a more sufficient removal of the contaminative proteins. In addition, the purity of the sample was neglected at this step, as it will increase as a result of further processing. The sample of the self-produced TEV protease showed a high purity, as no unspecific band could be detected. The Digest sample was taken after 16 h incubation at 4 °C but before the reverse affinity chromatography, for the purification of the digested protein. The band of the whole construct showed a clear loss of intensity, while the bands that correspond to the His-MBP-TEV and Linker²-FrsA_{TE}short/long could be detected and showed a high intensity. However, the purification by reversed affinity chromatography was not successful in both cases. The Flowthrough sample still showed bands corresponding to the undigested construct, the cleavage product His-MBP-TEV, and the TEV protease, whereby the band of the desired cleavage product Linker²-FrsA_{TE}short/long had a low intensity. Furthermore, the POI could also be detected in the Wash and the Elution sample, indicating a high loss, which had to be diminished, as the measured amount in the Flowthrough sample was 0.1 mg/1 L expression culture, which is too low. To increase the yield, the purification was optimized by increasing the incubation time with Ni-NTA after the TEV protease digest to enable a proper affinity binding. The purification fractions of interest could also be repeatedly mixed with Ni-NTA and purified. Furthermore, the digest could also be purified with an amylose column, as the MBP works as an affinity tag itself. The proteins without MBP should be found in the flowthrough, separated from the undigested construct as well as the cleavage product His-MBP-TEV. However, it should be noted that the protease cannot be separated by this method, as it does not harbor an MBP. Therefore, purification using MBP could only be an additional step and should not be used as a stand-alone application.

Apart from the optimization of the purification, another way to increase the yield of the POI significantly is upscaling and optimization of expression parameters. As this can be a time-consuming process due to the conjunction of many expressions, purifications, and SDS-PAGE analyses, the workflow may be accelerated by using a fusion protein with fluorescence, like mVenus.

4.1.2.5.2. mVenus Fusion tag

The green fluorescent protein (GFP) gene derives from the jellyfish *Aequorea victoria*^[183,184] and serves as the base for the generation of a variety of visibly fluorescent proteins^[185]. The monomeric Venus (mVenus) has a size of 26.9 kDa and is a yellow variant of monomeric enhanced green fluorescent protein (mEGFP), obtained by protein engineering.^[186–188] As a fluorescent fusion tag, designed for live-cell imaging, it can also be used for the optimization of expression and

purification, as it enables the determination of the protein yield by simply measuring the fluorescence of the cell crude extract.

The gene encoding for mVenus (*mVenus*) was amplified using a plasmid, provided by Dr. Julian Hegemann. It was then inserted into the already existing plasmid pHis8-TEV-FrsA_{TE}short/long, between the His-tag and the TEV cleavage site using Gibson assembly, resulting in the plasmid pHis8-mVenus-TEV-FrsA_{TE}short/long. The primers were designed to additionally insert a 'GGGSGGG' linker adjacent to the TEV cleavage site upstream to the POI, to promote proper cleavage by the TEV protease (**Chapter 6.6.6**). The test expression and digest were analyzed by SDS-PAGE, shown in **Figure 24**.

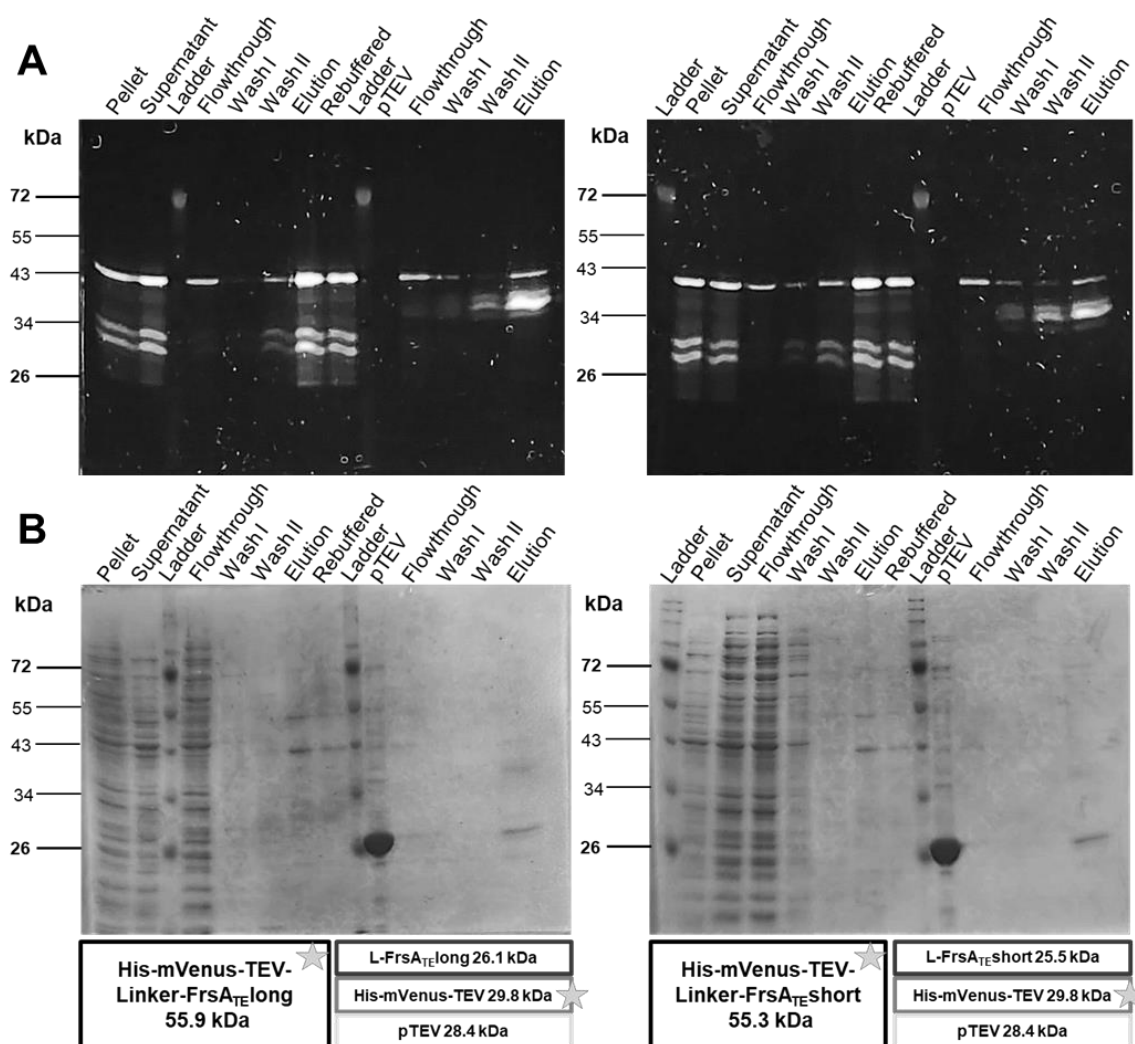


Figure 24: SDS-PAGE analysis of the expression, purification, and reverse purification after His-tag removal of the constructs FrsA_{TE}long (left) and FrsA_{TE}short (right) with mVenus fusion tag. **(A)** shows the detection via UV light, whereby **(B)** shows the same gel, which was subsequently stained with Coomassie Blue and detected under visible light. On the left side of each gel the Ni-NTA affinity purification samples after expression in *E. coli* LOBSTR and on the right side the reverse purification samples after His-tag removal were applied. Expected sizes: His-mVenus-TEV-Linker²-FrsA_{TE}long = 55.9 kDa, Linker²-FrsA_{TE}long = 26.1 kDa, His-mVenus-TEV-Linker²-FrsA_{TE}short = 55.3 kDa, Linker²-FrsA_{TE}short = 25.5 kDa, His-mVenus-TEV = 29.8 kDa, TEV protease = 28.4 kDa. Stars mark the fluorescent protein fragments. Applied volumes: Flowthrough, Wash, digest samples = 15 µL, Elution and final protein samples = 10 µL, ladder = 4 µL.

The fluorescent proteins were detected by UV light (**Figure 24A**), subsequently, the gels were stained using Coomassie blue, and evaluated under visible light (**Figure 24B**).

The expected band of the construct His-mVenus-TEV-Linker-FrsA_{TE}short/long should have a size of 55.9/55.3 kDa and a visible fluorescence, due to the comprised mVenus-tag. This band is slightly visible in the stained gel but is missing under UV light, under which the gel showed a clear fluorescent band at a height of the 43 kDa ladder band, which did not correspond to any expected protein sizes.

One possible explanation could be the structure of mVenus, which can be described as β -barrel protein (PDB: 7PNN). This specific tertiary structure can affect the detergent-loading level and the migration rate, resulting in 'gel shifting' behavior, whereby the migration while SDS-PAGE does not correlate with the formula molecular weight.^[189] In literature, it is often described, that SDS-PAGE analysis of GFP often leads to the formation of two bands of different sizes.^[190] If the protein was not fully degraded by heating or the used detergent, it maintained its β -barrel structure and its chromophore, resulting in a smaller fluorescent band, whereas a complete degradation led to the tear of the tertiary structure, associated with a loss of fluorescence and a normal behavior on SDS-PAGE.

To maintain the fluorescence, the SDS-PAGE samples were only heated up to 70 °C, which led to the detection of the non-fluorescent band at the expected size and a fluorescent band of lower height. If the fluorescence is not necessary for SDS-PAGE analysis, it can be considered to heat the samples at 95 °C, leading to a complete degradation of the tertiary structure and a single band at the expected size.

Furthermore, each purification sample containing the approx. 40 kDa band also showed two fluorescent bands between the sizes 34 kDa and 26 kDa with a correlating intensity. It was unclear where these bands derived from, but Coomassie staining did not lead to the visualization of these bands with the deviating band shape. This effect led to two different assumptions: 1. the fluorescence detection is a lot more sensitive or 2. that the bands derived from some fluorescent artifacts, which could be neglected.

However, based on the assumption that both upper bands with a size of ~55 and ~43 kDa derived from the His-mVenus-TEV-Linker-FrsA_{TE}short/long construct, a clear purification effect could be ascertained, as was to be expected, even if the final sample also showed some contaminating protein bands.

The digest with the TEV protease should result in a fluorescent 29.8 kDa-sized His-FP-tag and the non-fluorescent POI fragment with a size of 26.1/25.5 kDa.

The His-FP fragment with the desired size could neither be found by using UV light nor at the Coomassie-stained gel. However, it should be mentioned that another band occurred at a height of approx. 35-40 kDa after the TEV digest, which seemed to have the wrong size, but showed the expected purification pattern. Through the comprised His-tag, the fragment is expected to be arrested with the Ni-NTA resin and eluted by the wash and elution steps with an increasing imidazole concentration. This could be detected in both cases.

The band with a size of approx. 40 kDa, was detected in each sample with different intensities. Based on the assumption, that this was the construct His-mVenus-TEV-Linker-FrsA_{TEshort}/long, it should only be found in the Elution sample due to the His-tag. The appearance in the flowthrough sample indicated that the amount of Ni-NTA resin or the affinity to the resin was insufficient. The two unspecific bands with a size of approx. 30 kDa disappeared completely, which could be a sign of degradation or a confirmation of the artifact theory.

The Linker-POI fragment should not be detectable under UV light, due to the removal of mVenus. Unfortunately, the stained gel also showed no band that could be distinctly assigned to this fragment. The expected band would have a size of 26.1/25.5 kDa and should thus be found below the protease band (28.4 kDa), detectable in the Elution sample. On closer examination, a very weak band was detected in the Flowthrough samples. However, it should be noted, that the purification samples marginally showed bands after digestion, which is conclusive when examining the low protein yield of the input sample 'Rebuffered'.

After the SDS-PAGE analysis of the test expression and purification, the DNA and amino acid sequences of the POI were controlled repeatedly. No discrepancies were found, which could lead to another interpretation of the results. Since the plasmid was sequenced and was used in previous experiments in which the protein could be identified, it was ensured that the amino acid sequence of the POI was correct. The detection of a fluorescent product also indicates the successful insertion and expression of mVenus. As the SDS-PAGE anomalies were already described in other publications and the purification patterns also fit with previous results, it was concluded that expression, purification, and removal of the His-FP-tag was successful, although this attempt did not lead to an increased yield of FrsA_{TEshort}/long. In addition, the inconclusive evaluation using SDS-PAGE and the design of a more promising construct led to the decision not to pursue this approach further.

4.1.3. Didomain FrsA_{T+TE}

4.1.3.1. Construct design

NRPS are organized in modules comprising multiple domains, functioning as an assembly line-like enzymatic machinery. Therefore, specific protein-protein interactions are of utmost importance in NRPS biosynthesis^[155] and could also play a critical role in protein stability. The complexes are of high mobility, as the flexible linkers enable the interaction between alternate domains during peptide assembly.^[170] Thus, it can be advantageous to work with single domains for crystallization approaches. However, if the first crystallization trials and further optimization approaches do not lead to the intended results, it may be promising to consider crystallizing larger constructs comprising more domains, didomains.^[191,192] In the case of the TE domain of the *frs* BGC, the crucial interaction partner is FrsA_T the T domain of the same module. So, it was assumed that the T-TE didomain, FrsA_{T+TE}, could be a promising construct for crystallization. Furthermore, the addition of the substrate FR-Core should lead to the formation of *holo*-form crystals with co-crystallized FR-Core.

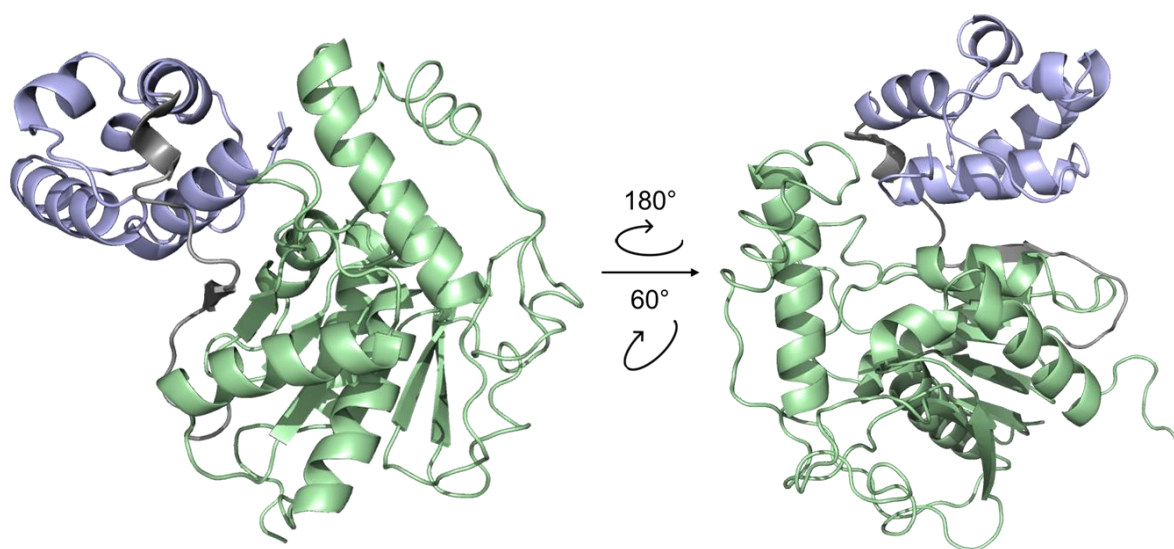


Figure 25: I-TASSER-predicted structure of the designed didomain construct FrsA_{T+TE} in different orientations. (color-coded: blue = thiolation (T) domain, green = thioesterase (TE) domain, gray = linker regions)

The construct FrsA_{T+TE} was designed based on the antiSMASH annotation of the FrsA domains. It comprises the amino acids Q1-G75 of the T domain (FrsA Q948-G1023) and the amino acids G76-Q325 of the annotated TE domain (FrsA G1024-Q1272), whereby it must be considered that the amino acids E75-H96 (FrsA E1022-H1049) belong to the linker region, as already discussed in Chapter 4.1.2.1.

4.1.3.2. Cloning

The gene *frsA_{T+TE}* was cloned using the restriction and ligation method. Therefore, *frsA_{T+TE}* was amplified by PCR using gDNA of *C. vaccinii*, inserted into the pHis8-TEV vector (linearized with *Bam*HI and *Hind*III), and transformed into the expression strain *E. coli* LOBSTR (6.5).

4.1.3.3. Expression and Purification

The SDS-PAGE analysis (6.6.8) of the first expression and purification (6.6.1, 6.6.2) attempt of FrsA_{T+TE} is shown in **Figure 26**. The calculated size for the His-TEV-FrsA_{T+TE} construct amounts to 38.7 kDa and the SDS-PAGE showed a corresponding band with high intensity in the middle of the 30 and 40 kDa ladder band. Therefore, it could be concluded that the expression and purification of the POI were sufficient.

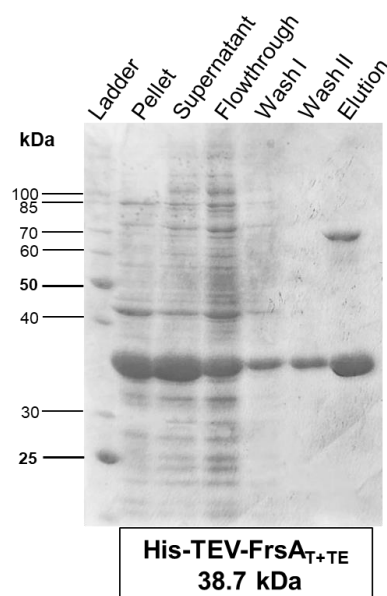


Figure 26: SDS-PAGE of His-TEV-FrsA_{T+TE} purification after expression in *E. coli* LOBSTR. Expected size for His-TEV-FrsA_{T+TE} = 38.7 kDa. Eluted protein is rebuffed in TE domain buffer. Applied volumes: purification fractions = 15 µL, ladder = 4 µL.

SDS-PAGE analysis also revealed a high protein loss during the purification, as a prominent band of the desired size could be detected in each fraction. However, the loss could be neglected, since the yield and purity of FrsA_{T+TE} were sufficient for further processing (5 mg/100 mL expression culture). The purity of the Wash II fraction was suitable for pooling with the Elution fraction, increasing the protein yield. The Elution sample showed two bands: one of high intensity and a height of approx. 38 kDa, deriving from the POI, and one band at a height slightly above the 70 kDa ladder band. This second band could derive from a dimer, which is most likely to occur at higher

protein concentrations.^[193] The band could also be the result of an insufficient denaturation during preparation of the SDS-PAGE samples. Through upscaling experiments, the optimal expression volume was defined as 100 mL.

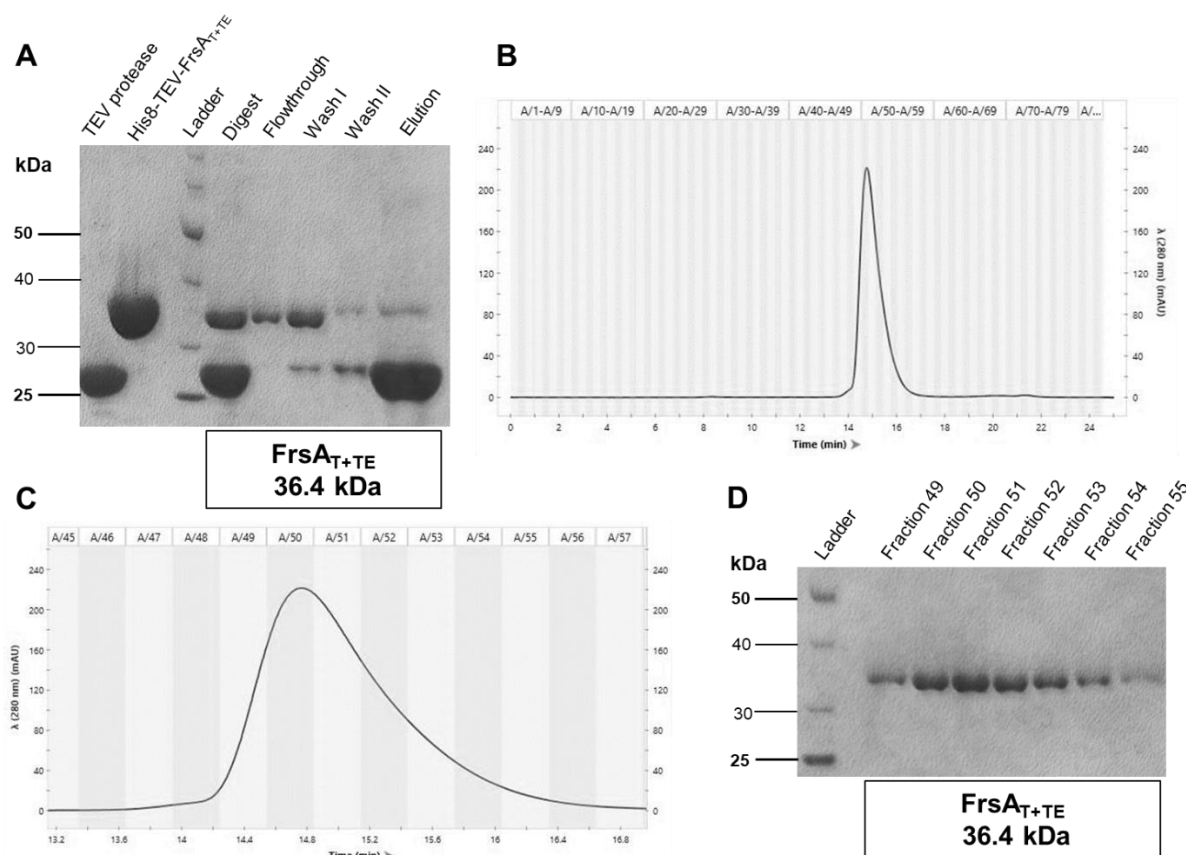


Figure 27: Sample processing of His-TEV-FrsA_{T+TE} for crystallization approach, involving (A) SDS-PAGE analysis of reverse purification after His-tag removal, (B)+(C) SEC-FPLC chromatogram of FrsA_{T+TE}, and (D) SDS-PAGE of the SEC-FPLC fractions. Expected sizes: His-TEV-FrsA_{T+TE} = 38.7 kDa, FrsA_{T+TE} = 36.4 kDa, TEV protease = 28.4 kDa. Applied volumes: protein samples = 5 µl, purification fractions, SEC-FPLC fractions = 15 µl, ladder = 4 µl.

The next step was the efficient removal of the His-tag (6.6.6). Accordingly, the His-TEV-FrsA_{T+TE} construct was incubated with the TEV protease in a 5:1 ratio overnight (~16 h) at 4 °C. In **Figure 27A**, the reverse purification after the digest with the TEV protease is depicted. The first samples were the input for the digest: the TEV protease and the His-TEV-FrsA_{T+TE}. Both showed a high purity with no detectable bands originating from protein contaminations. The 'digest' sample was taken after overnight incubation right before the reverse affinity purification. The lower band originates from the TEV protease and the upper could be the desired FrsA_{T+TE} without His-tag. As the undigested and the digested POI differ only in 2 kDa, they were hardly distinguishable. However, comparing the ~35 kDa band of the His-TEV-FrsA_{T+TE} and the flowthrough sample, it can be assumed that the His-tag removal was sufficient. The comparison of the product and the input band was not suitable to determine if the size of the band decreased, as the protein band of the

input sample had too high intensity to be able to make a verified prediction about the size. However, the direct comparison of the flowthrough and Elution sample provided clarification that the bands indeed possessed different sizes and consequently resulted from the His-TEV-FrsA_{T+TE} and FrsA_{T+TE}. Nevertheless, the SDS-PAGE analysis also revealed protein loss caused by wash step I, as the sample showed a prominent band at the desired height. Despite the slight contamination with TEV protease, the Flowthrough, and the Wash I sample were pooled and proceeded further. This should prohibit protein loss, and the contamination could also be removed during FPLC purification. In following purifications, the incubation time with Ni-NTA was extended, and the volume of Wash buffer I was increased to achieve a better separation. The total protein loss during His-tag removal was between 40-50%.

To obtain FrsA_{T+TE} in sufficient purity for the crystallization trials, the protein was processed using SEC-FPLC (6.6.7). The pooled Flowthrough and Wash I samples were concentrated to a concentration of up to 30 mg/mL where 200 µL were gradually injected into the SEC-FPLC. The entire chromatogram of a single run of FrsA_{T+TE} is shown in **Figure 27B**. The protein eluted as a single peak and was clearly assigned to the POI. This is a sign of the high purity of the sample. Only a small degree of dimerization was detected, eluting as a small shoulder directly before the main peak (**Figure 27C**). Before pooling the fractions, they were analyzed for contaminations *via* SDS-PAGE (**Figure 27D**). Fractions 49-55 were pooled and concentrated to 35-40 mg/mL for crystallization trials. Purification by SEC-FPLC with subsequent concentration resulted in a protein loss of 20-30%, depending on the number of runs.

4.1.3.4. Crystallization trials

Crystallization trials of FrsA_{T+TE} were conducted in cooperation with Philipp Hendricks and PD Dr. Gregor Hagelüken from the Institute of Structural Biology, University of Bonn.

The crystallization was performed by Philipp Hendricks using the sitting drop vapor diffusion method. Thereby, the protein solution with a concentration of 35-40 mg/mL was mixed 1:1 with the precipitation solution in a drop above a reservoir filled with 20 µL pure precipitation solution (6.6.12). After sealing the chamber, the water diffuses from the drop into the reservoir and vice versa, leading to a slow concentration change until a state of equilibrium of osmolarity is reached. If this state is reached in the nucleation zone of the phase diagram in **Figure 28**, crystal formation should occur.

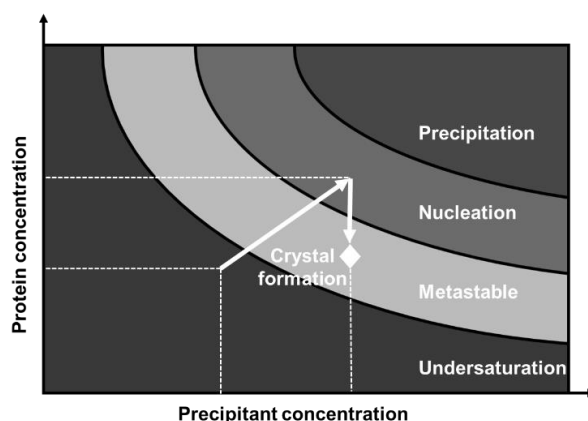


Figure 28: Phase diagram for protein crystal formation conditions adapted from Dessau *et al.*^[194]

For each tested crystallization buffer, a batch of *apo* FrsA_{T+TE} and a batch with FrsA_{T+TE} supplemented with a 2-fold molar excess of FR-Core was prepared. On the one hand, it was intended to stabilize the TE domain in the presence of its natural substrate, on the other hand, it would precipitate in *holo*-form, facilitating the elucidation of the reaction mechanism.

Initial crystallization trials were conducted using a standard buffer with additives screening was performed using commercially available buffer packages (Ligandfriendly, PACT++, ProPlex, etc.). The screening was conducted to identify buffer conditions that would facilitate protein crystallization. Afterwards, the most promising buffer undergoes profound optimization approaches to gain the optimal conditions for the formation of well-diffracting crystals. It should be highlighted that numerous conditions were found during buffer screening in which many promising crystals grew, especially in the approaches, in which FR-Core was added for co-crystallization. Cubic crystals grew in buffers consisting of 0.1 M Bis-Tris, 20% polyethylene glycol 3350 (PEG), and 0.2 M Na₂SO₄, pH 7.5, within 12 h (**Figure 29**).

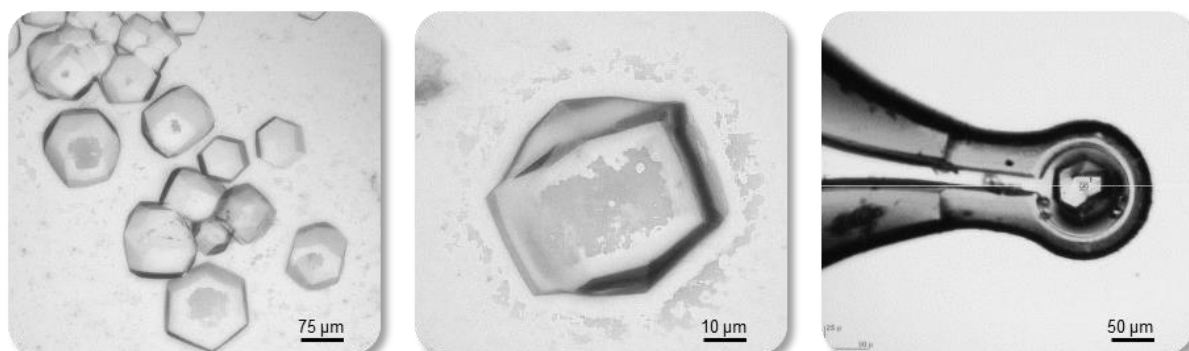


Figure 29: Protein crystals of the construct FrsA_{T+TE} in a co-crystallization trial with a twofold molar excess of FR-Core. Buffer conditions: 0.1 M Bis-Tris, 20% PEG 3350, and 0.2 M Na₂SO₄, pH 7.5. Right = protein crystal in a nylon loop. Pictures recorded by Philipp Hendricks.

Suitable crystals were selected based on size and shape, transferred to a nylon loop, and flash-frozen in liquid nitrogen. The crystal-containing loop was mounted on a goniometer head and placed in the X-ray beam. Through rotation, multiple diffraction patterns were recorded. The crystals had the ideal requirements for X-ray diffraction analysis and showed an excellent diffraction pattern, sufficient to generate reasonable three-dimensional data. The 3D structure was generated through molecular replacement based on the AlphaFold-predicted structure of the didomain.

4.1.3.5. Structure elucidation of FrsA_{T+TE}

The crystal diffracted X-rays to a maximum resolution of 2.1 Å, which was sufficient to elucidate the three-dimensional structure of FrsA_{T+TE}. Although the addition of FR-Core supported crystal formation, FR-Core itself was not detected in any of the crystals measured, hence the existing structure merely represents the *apo* form of the didomain.

The T domain has a length of 73 amino acids and shows the typical structure consisting of four α-helices (**Figure 30**). The helices α1, α2, and α4 exhibit 14, 15, and 10 amino acids a greater length and have approximately the same orientation. Helix α3 possesses only 6 amino acids and tends rather perpendicular to the other helices. As active center, the serine residue at the beginning of α2. was described, which can also be found at the same position (S35) in the structure of FrsA_{T+TE}. Phosphopantetheinylation of this amino acid is crucial for the T domain functionality, as it enables binding and translocation of the growing NRP. Located at the N-terminus

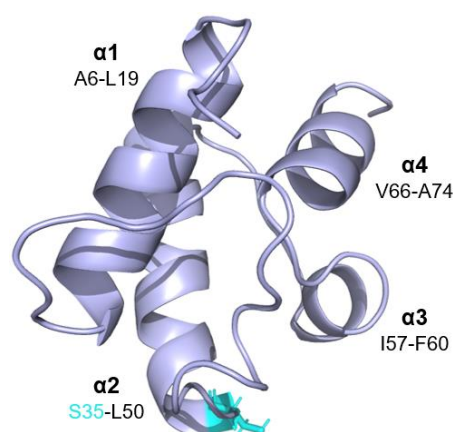


Figure 30: Crystal structure of FrsA_T, obtained by X-ray diffraction measurements. The catalytic serine is highlighted in blue and given in sticks.

of α2 is a long loop, the sequence of which differs considerably between T domains. In the case of the FrsA_T, the loop comprises 18 amino acids (L19-H36). In general, it can be emphasized that the structure of the FrsA T domain exhibits a high similarity to previously described T domains.^[195]

The TE domain of FrsA shows a typical structure for a member of the α/β hydrolase superfamily (**Figure 31A**). It consists of seven parallel β-sheets (β1-β7), surrounded by five α-helices (αA-αE). The active site is covered by a lid structure, formed by three α-helices (L1-L3) between β5 and β6. The active site harbors the catalytic triad, consisting of S170, D197, and H300 (**Figure 31B**). The serine is located in the highly conserved motif 'GXSXG' in the turn between β4 and αC, while the aspartate can be found between β5 and the first lid α-helix (αL1), and the histidine is right behind

$\beta 7$. Through the protein fold the residues are located close to each other in the active center: The central histidine is 3.0 and 4.3 Å apart from the serine and the aspartate residue. In canonical TE domains, histidine and aspartate activate the nucleophile serine by forming a hydrogen bond network.^[93,196] Thus, the serine residue can form an acyl-enzyme intermediate^[197] in the first half-reaction.

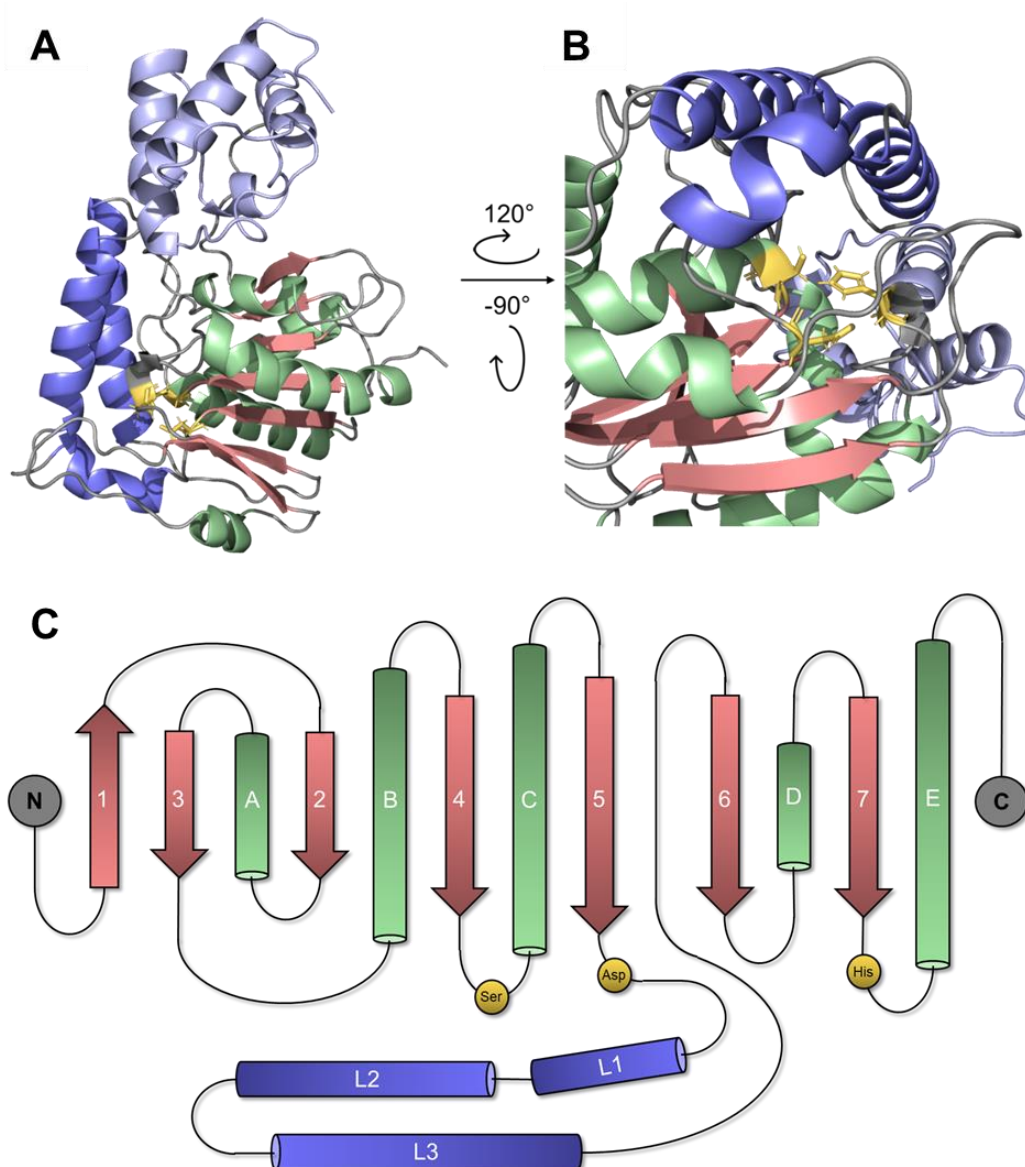


Figure 31: Crystal structure of FrsA_{T+TE} obtained by X-ray diffraction measurements. **(A)** Depiction of the monomeric structure of the didomain with color-coded structures. **(B)** Focus on the active center and the catalytic triad. **(C)** Schematic illustration of the protein structure of FrsA_{TE}. Color-coded structures: light blue = T domain, green = α -helices, red = β -sheets, blue = lid α -helices, yellow = catalytic residues given in sticks.

The structural foundation of the second half-reaction catalyzed by the TE domain, where the ester bond is formed by nucleophilic attack, is much less characterized.

The elucidated crystal structures of ten thioesterase domains^[128,130,131,140,198–203] deriving from NRPS systems provide structural insights: SulM_{TE} (8W2C; sulfazecin biosynthesis^[203]), Vlm2_{TE} (6ECF; valinomycin biosynthesis^[199]), NocB_{TE} (6OJD; nocardicin biosynthesis^[198]), EntF_{TE} (3TEJ; enterobactin biosynthesis^[130]), Srf_{TE} (1JMK; surfactin biosynthesis^[140]), Feng_{TE} (2CB9; fengycin biosynthesis^[200]), Skyxy_{TE} (7CRN; skyllamycin biosynthesis^[201]), Ooc_{TE} (7R0X; oocycin biosynthesis^[204]), AB3403 (4ZXH; NRPS from *Acinetobacter baumannii*^[128]), holoObiF1_{TE} (6N8E; NRPS from *Burkholderia diffusa*^[202]).

A structural comparison of FrsA_{T+TE} with the studied NRPS domains was conducted using PyMOL (6.9.8) and DaliServer (6.9.7) (**Figure 32**). The root mean square displacement (rmlds) of a many-to-one comparison revealed high structural similarities, except Ooc_{TE}, AB3403, and holoObiF1_{TE}, which showed an rmsd of 7–11 Å. Thus, they were excluded from further analysis. Additionally, it should be mentioned, that the structure of SulM_{TE} could not be included in the analysis, since the structure is not available until the corresponding preprint is published in a peer-reviewed journal. The remaining thioesterases showed rmsd values of 0.9–2.7 Å, whereby the core structures showed even higher similarities. By deleting the lid structure, the rmsd could be decreased to about 0.5 Å.

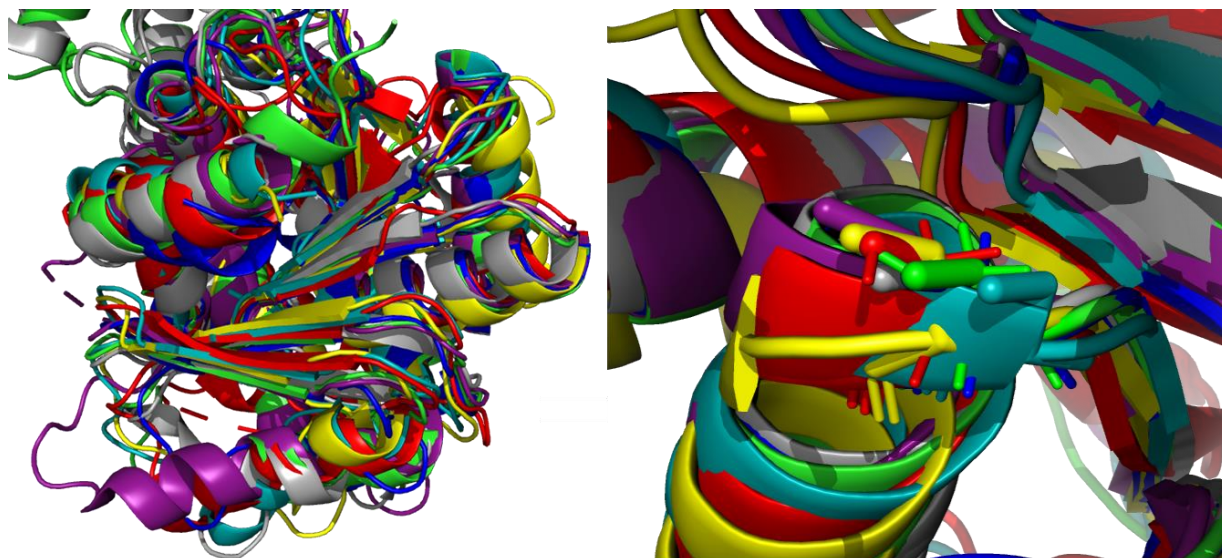


Figure 32: Structural comparison of (**left**) conserved core structures and (**right**) the catalytic serine (given in sticks) the of FrsA_{TE} (green) with other NRPS TE domains (Vlm2_{TE} (6ECF, red), NocB_{TE} (6OJD, blue), EntF_{TE} (3TEJ, grey), Srf_{TE} (1JMK, yellow), Feng_{TE} (2CB9, cyan), Skyxy_{TE} (7CRN, purple).

This result correlates with the fact that the lid shows a high structural diversity among TE domains and a distinct dynamic (**Figure 33**).

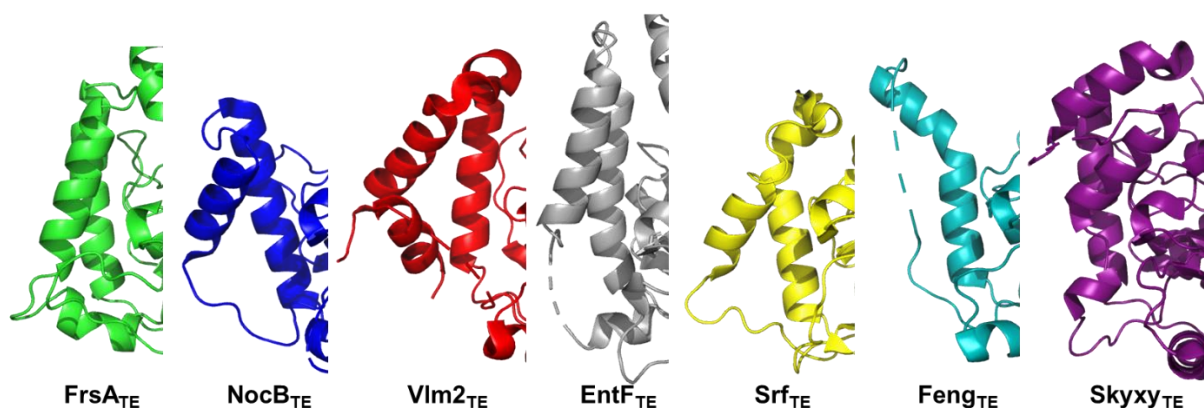


Figure 33: Structural comparison of the lid structure of FrsA_{TE} (green) with the lid structures of other NRPS TE domains (Vlm2_{TE} (6ECF, red), NocB_{TE} (6OJD, blue), EntF_{TE} (3TEJ, grey), Srf_{TE} (1JMK, yellow), Feng_{TE} (2CB9, cyan), Skyxy_{TE} (7CRN, purple).

Generally, the lid structure is located between the β_6 and β_7 and is organized in 2 to 5 α -helices. The lid of FrsA_{TE} has a length of 51 amino acids (I201-A252) and is composed of three α -helices. The first one is the smallest with 8 amino acids (α L1: I201-F209), followed by α L2, comprising 14 amino acids (C214-L228), and the longest helix α L3 (E233-A252). It can be assumed that the relatively small lid enables the acceptance of bulkier substrates, but so far no correlation between the lid character and the substrate length or the release mechanism could be confirmed.^[93] However, what was observed repeatedly is that the lid structure is highly mobile.^[199,205] Thus, it was proposed that the lid structure interacts with the upstream T domain and the substrate, and acts as a solvent shield during the thioesterification reaction.^[130]

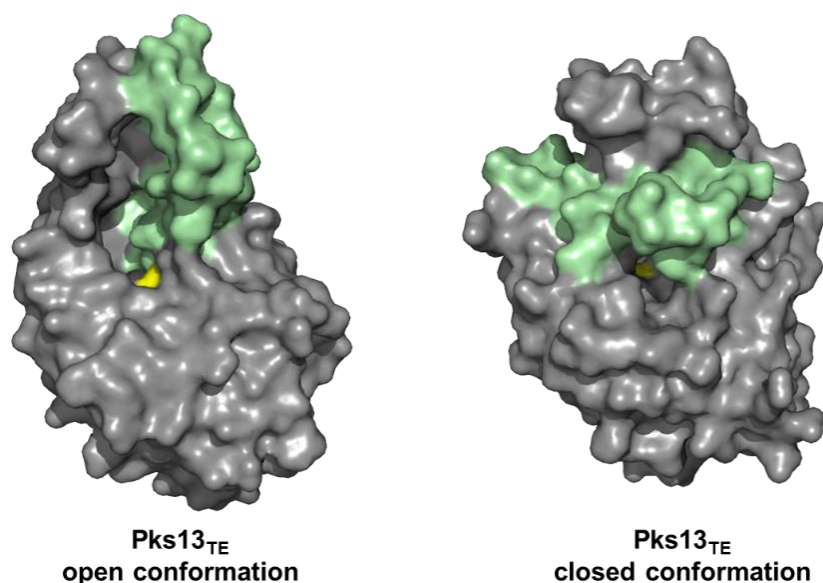


Figure 34: Illustration of the surface structure of the Pks13_{TE} from *Mycobacterium tuberculosis* in the 'open' and 'closed' conformation of the lid structure. The lid structure is colored in green, and the catalytic serine is highlighted in yellow.

There are two existing structures of the PKS 13 TE domain from *Mycobacterium tuberculosis*, superposing as 'open' (5V3Y) and 'closed' (3LCR) (**Figure 34**). In the 'open' conformation, the substrate tunnel is expansive, and the catalytic residue S1533 is accessible. Conversely, in the 'closed' conformation, both the substrate tunnel and the catalytic serine are barely accessible, what confirms the function as a solvent shield. While the overall structure of the TE domain undergoes subtle alterations, the most pronounced conformational changes occur within the lid region.

Based on these results, it was assumed that the FrsA TE domain crystallized in the 'closed' conformation, as there is only a narrow cleavage through which the active center is barely visible and thus hardly accessible (**Figure 35**).

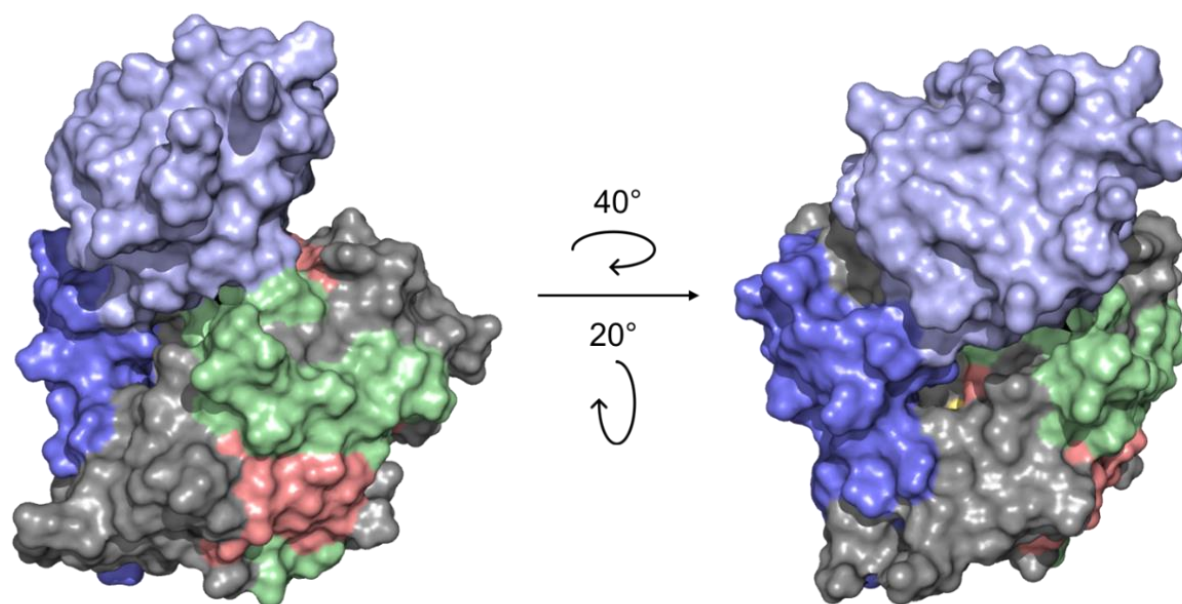


Figure 35: Visualization of the FrsA_{T+TE} surface from different angles. The right structure allows a view from above through the substrate binding pocket into the active center. Color-coded structures: light blue = T domain, green = α -helices, red = β -sheets, blue = lid α -helices, yellow = catalytic residues.

Even after several crystallization attempts with FR-Core and the side chain mimic, no *holo*-FrsA_{TE} could be ascertained. This could be due to the fact, that the acyl-*O*-TE intermediate is relatively unstable and therefore has a short lifetime.^[140,200] The elucidated complexes were generated through the utilization of phosphonate analogs^[198,206,207] or by mutagenesis of the genetic code to employ stable amide intermediates.^[199]

Due to the lack of a *holo*-FrsA_{TE} structure, the elucidation of the reaction mechanism is aggravated. Nevertheless, in order to gain an insight into the potential substrate interaction, a docking study of the elucidated FrsA_{T+TE} structure and FR-core was conducted using SwissDock. The results are depicted in **Figure 36**. FR-Core seems to bind at the C-terminal region of α L2, where many

hydrophobic side chains (V, M, L, Y) are localized. According to SwissDock, also the amino acids between the $\beta 7$ and αE appeared to be involved in the substrate interaction. Furthermore, FR-Core showed points of contact with the hydrophobic amino acids at the end of the $\alpha 2$ of the T domain.

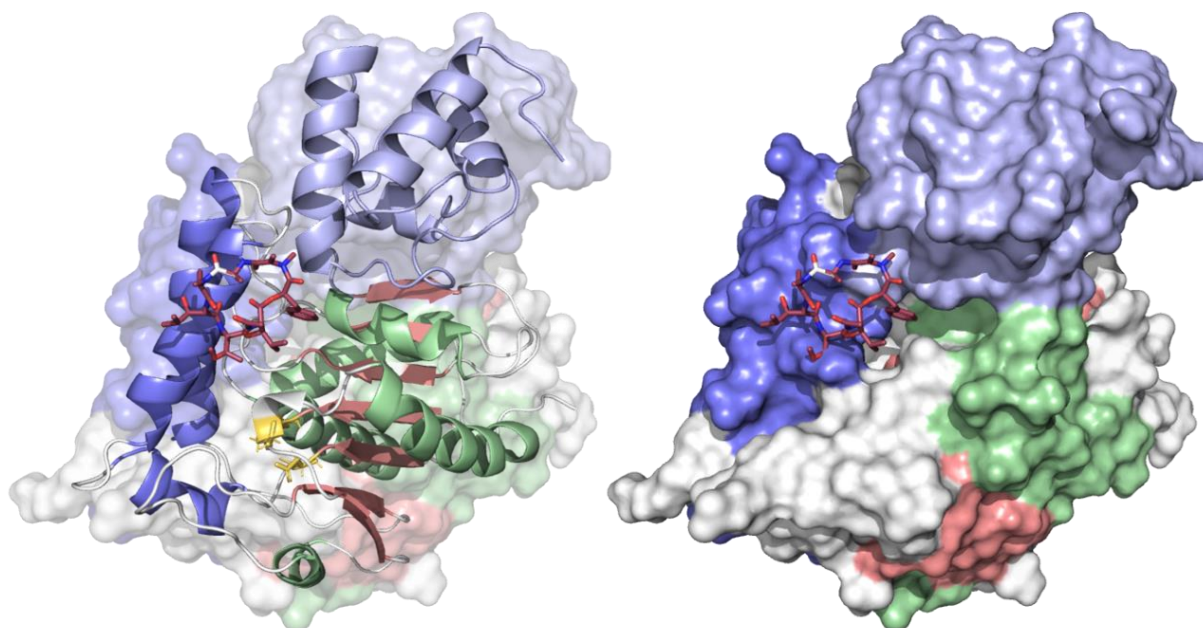


Figure 36: Depiction of the docking studies of FrsA_{T+TE} with FR-Core performed with SwissDock. Color-coded structures: light blue = T domain, green = α -helices, red = β -sheets, blue = lid α -helices, yellow = catalytic residues given in sticks, Dark red = FR-Core structure given in sticks.

Nevertheless, the reliability of the docking results is questionable for several reasons. Notably, the employed enzyme structure likely represents its ‘closed’ conformation, as previously discussed. Consequently, large parts of the lid (especially the inner region) and the active center were not included in the calculation of the program. This becomes particularly clear by calculating the distance between the substrate and the catalytic Ser using PyMOL, which amounts to 14.4 Å. The carbon atom used for the calculation originates from the phenyl ring of the (3) D-PLA building block, as this extends furthest into the substrate tunnel, which points directly to the next problem of the docking study. It is not the (3) D-PLA building block that is used for the transesterification by FrsA_{TE}, but the side chain is transferred to the building block (7) Hle, which, according to the docking results, is situated most distant from the substrate tunnel and thus also from the active center. However, the docking prompted, that the hydrophobic residues in the lid region may interact with the hydrophobic building blocks with FR-Core ((2) *N*-Ac-Hle and (3) D-PLA), coordinating FR-Core in the substrate tunnel in a way, where the (7) Hle reaches near the catalytical Ser. With the sculpting function in PyMOL, this possibility was investigated and the hydroxy group of (7) Hle reached the Ser170 with a distance of 4.1 Å, when the (3) D-PLA interacted with the C-terminal end of $\alpha L2$.

The comparison with other substrate-bound TE domains further confirmed that FR-Core requires binding within the active center. Only two structures contributed to the understanding of substrate interaction: the TE domains of the nocardicin and valinomycin biosynthesis. In the case of NocB_{TE}, the phosphate analog of the peptide substrate was designed, which reacts with the catalytic serine residue. Thus, the location of the terminal three residues of the peptide right before the release could be investigated.^[198] In the Vlm2_{TE}, the catalytic residue was replaced by 2,4-diaminopropionate, which reacts stably with the peptide and traps it for the crystallization process. **Figure 37** depicts both crystal structures and revealed that the substrates are located deeply in the active center, near or bound to the catalytic serine.

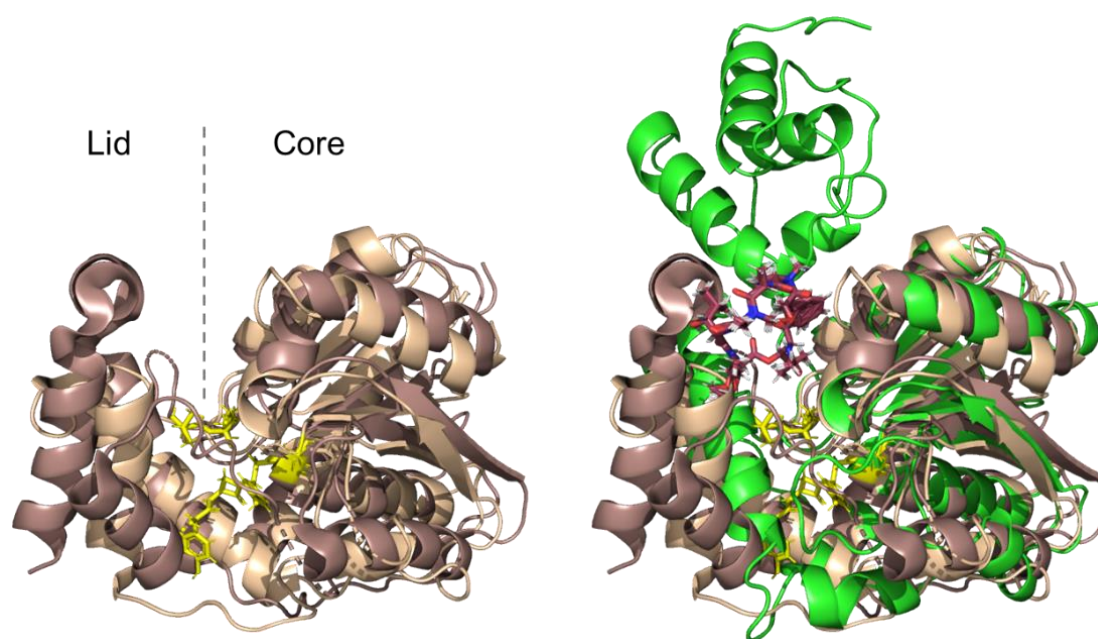


Figure 37: Comparison of the structures of the predicted *holo*-FrsA_{T+TE} (green, FR-Core in dark red) with *holo*-Vlm2_{TE} (6ECE, brown, substrate highlighted in yellow) and *holo*-NocB_{TE} (6OJD, beige, substrate highlighted in yellow) with focus on the substrate binding pocket.

It should be noted that the bound structures derive from the T domain-bound Ppant moiety, which then binds to the nucleophilic serine and is so ‘taken over’. The bound substrates of NocB_{TE} and Vlm2_{TE} thus correspond more to the side chain which is assembled by FrsA and then attached to FR-Core by the TE domain. Consequently, FR-Core embodies the nucleophile to which the TE-bound intermediate is transferred. In both analyzed *holo*-protein structures, the nucleophile is a water molecule and therefore cannot be used for comparisons or analysis. The intermolecular transesterification of FrsA_{TE} is an unusual special process, which is why it is especially important to elucidate its mode of action, by determining the substrate organization in the active center and which residues are involved.

Two NRPS TE domains were structurally described in complex with the associated T domain: the EntF_{TE} from enterobactin biosynthesis and SulM_{TE} from the sulfamycin biosynthesis.^[130,131] Furthermore, the EntF_{TE} structure also contains the phosphopantetheine (Ppant) moiety, which is bound to S1005. Due to the high structural similarity (rmsd = 1.741 Å) of the EntF TE and FrsA TE domains, they can be depicted comparatively, which allows assumptions about the interaction between the domains as well as substrate binding and arrangement in the FrsA_{T+TE} complex.

For the interaction of the T domain with other catalytic domains the residues from the N-terminal end of $\alpha 2$, $\alpha 3$, and the loop between $\alpha 1$ and $\alpha 2$ play a major role.^[208] This also appears to be the case for the T-TE-complexes of EntF and FrsA, as both show approximately the same configuration, in which the T domain seems to interact with $\alpha 1$ and $\beta 1$ from the TE domain. It can be assumed that the surface patch of the T domain consists of the hydrophobic and negatively charged residues F63, E63, and R60 on $\alpha 3$ and F33, and D34 joining $\alpha 1$ and $\alpha 2$. The potential residues of the TE domain that could be involved in the interaction are hydrophobic and positively charged. This includes the I114, A117, L118 of αA , L88, M89, I91 of $\beta 1$, and R306 and L310 of αE . It is striking, that the lid structure of EntF_{TE} is considerably longer than that of FrsA_{TE} and interacts with the entire $\alpha 2$ of the T domain. The shorter lid structure of FrsA_{TE} can at most interact with the N-terminal part of the $\alpha 2$ of FrsA_T.

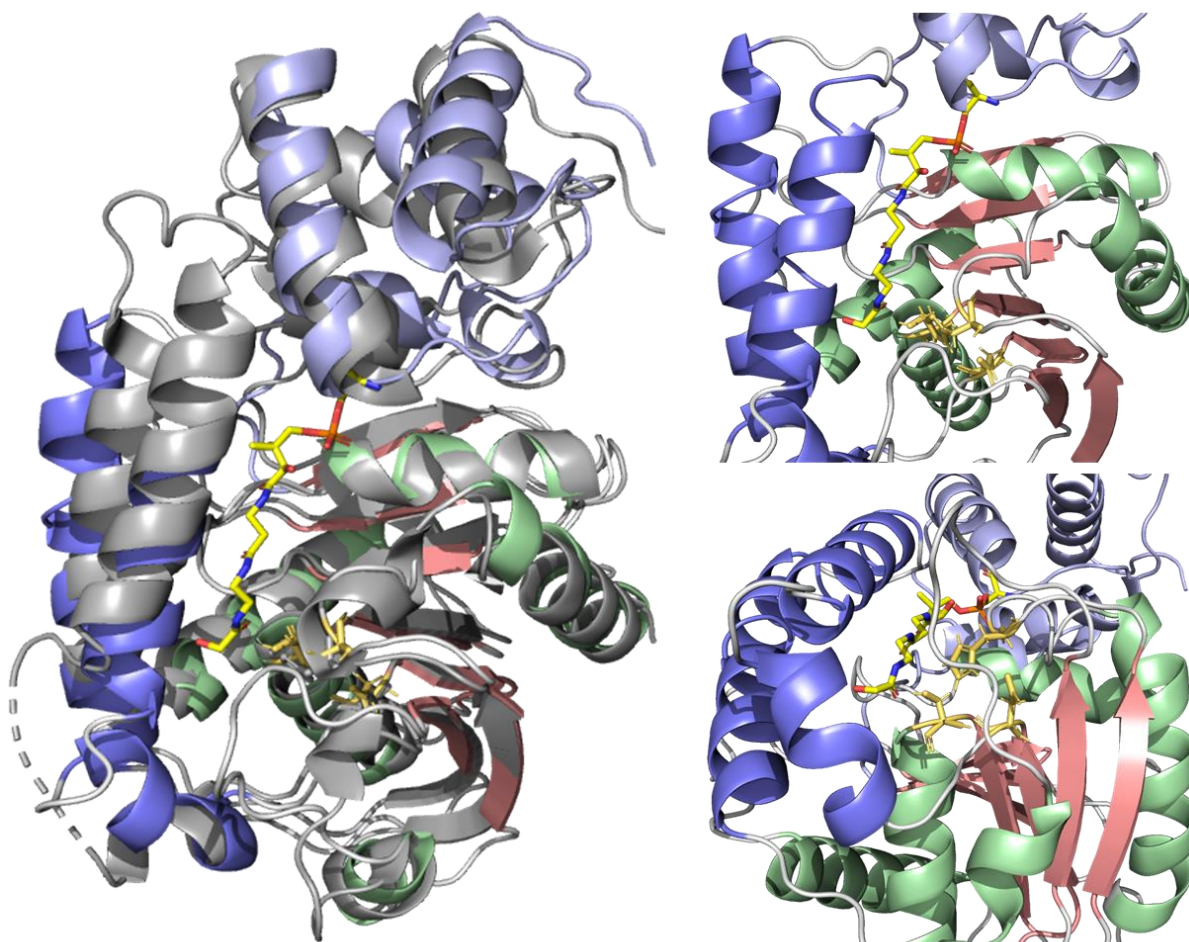


Figure 38: Structural comparison of the FrsA_{T+TE} with EntF_{T+TE} (3TEJ, grey) containing the phosphopantetheine moiety (yellow, given in sticks) bound to the catalytic serine of the T domain, from different angles. Color-coded structures of FrsA_{T+TE}: light blue = T domain, green = α -helices, red = β -sheets, blue = lid α -helices, yellow = catalytic residues given in sticks)

The structure of EntF_{T+TE} also contains the phosphopantetheine which is modelled as free thiol (**Figure 38**). The T domain binds on the cleft of the substrate tunnel to enable the permeation of the Ppant. Bound to the T domain, it reaches deep into the substrate tunnel formed between the lid and the core structure and reaches the active site. The transfer of the Ppant structure to that of FrsA_{T+TE} shows that the unit extends exactly to the catalytic serine residue, even despite a slight shift between the structures. The distances numbers only 3.2 Å. The comparison of both crystal structures provides insight into how the Ppant arm could be organized in the substrate tunnel of FrsA_{TE}. The pantetheine unit of SulM_{TE} shows a similar orientation in the active site, further confirming the assumptions. The substrate tunnel is formed by the mainly hydrophobic regions between β 2 and α A (P105-G110) and α L2 (C214-M229), indicating that these structures are involved in the Ppant positioning. For example, the hydrophobic M229 is most likely to interact

with the dimethyl group of the Ppant and R221 with the phosphonate moiety. The hydroxy groups seem to interact with the polar negatively charged residues E106 and E225.

Another well-preserved characteristic throughout α/β -hydrolases is the oxyanion hole.^[200] It shields the hydroxyl group of the catalytic serine, thus stabilizing the negatively charged tetrahedral intermediate. For FrsA_{TE} it was identified by comparison with the structure of EntF_{TE} (**Figure 39**).^[130,209] P105 plays a crucial role in orienting the backbone amide hydrogen into the oxyanion hole. Therefore, it is conserved in all TE domains.^[130] The hydroxyl sidechain of Ser170 attacks the terminal carbonyl of the Ppant and the resulting oxyanion is stabilized by the amide hydrogens of Glu105 and Leu171.

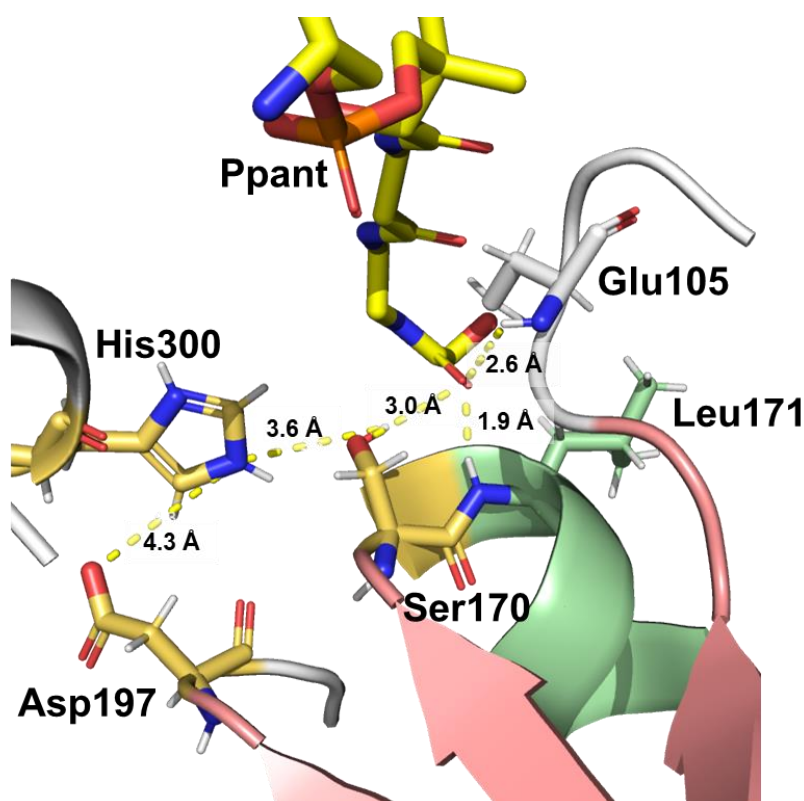


Figure 39: Illustration of the oxyanion hole of FrsA_{TE}, modelled with the pantetheine moiety of EntF_{T+TE} (3TEJ, yellow and given in sticks) with distance measurement. Color-coded structures of FrsA_{T+TE}: light blue = T domain, green = α -helices, red = β -sheets, blue = lid α -helices, yellow = catalytic residues. The involved residues Glu105 and Leu171 and the catalytic triad is given in sticks.

Below the active center, a second loop is located between $\beta 6$ and $\beta 7$. After the lid structure, it shows the highest variability between the TE structures (**Figure 40**). Generally, this region comprises 19-26 residues, containing a short α helix in most cases.

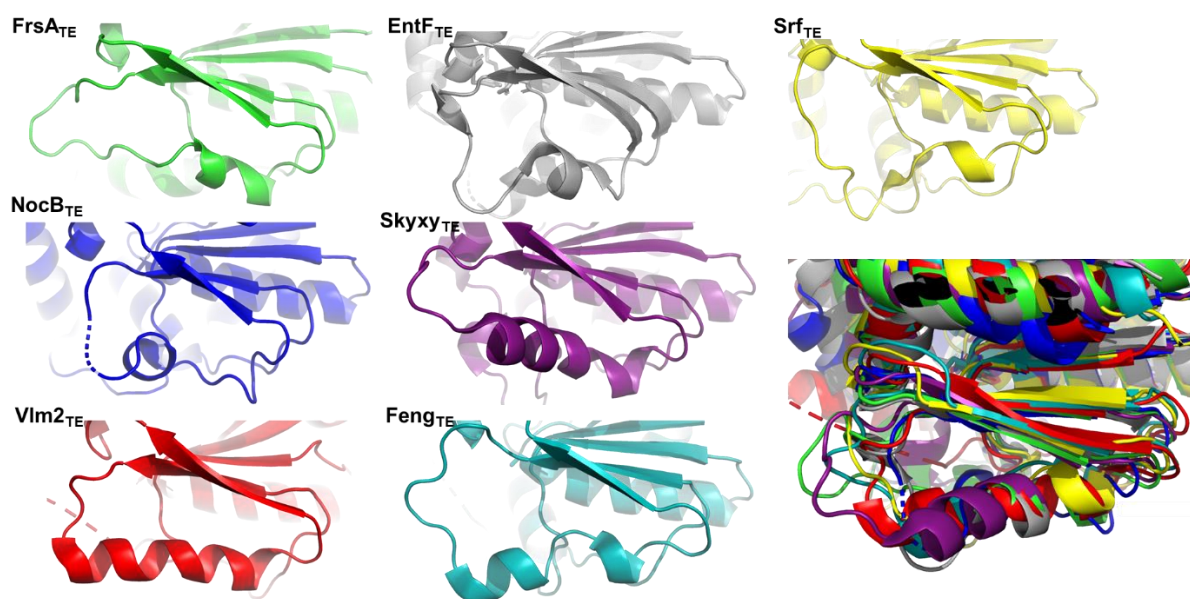


Figure 40: Comparative presentation of the loop between the $\beta 6$ and $\beta 7$ that border the active center from below of FrsA_{TE} with other NRPS TE domains (Vlm2_{TE} (6ECF, red), NocB_{TE} (6OJD, blue), EntF_{TE} (3TEJ, grey), Srf_{TE} (1JMK, yellow), Feng_{TE} (2CB9, cyan), Skyxy_{TE} (7CRN, purple)).

In FrsA_{TE}, this loop has a length of 23 amino acids (L266-Q289) with a helix of six amino acids (A278-Y284). In NocB_{TE} and Vlm2_{TE}, no ligand interaction could be observed, but since the residues L266-E277 form the lower area of the substrate binding pocket, it could not be excluded that it may play an important role in the interaction with FR-Core in this noncanonical reaction. An in-depth analysis of enzyme-substrate interaction should also focus on this particular area.

Given the unsuccessful co-crystallization trials, crystal soaking was pursued as an alternative strategy to investigate protein-substrate interaction. This technique is commonly preferred for its time- and cost-effectiveness. It involves exposing the pre-manufactured protein crystals to a solution containing the substrate of interest, which should diffuse through the water channels passing through the crystalline packing of the assembled protein molecules. Upon binding to a suitable protein surface, the ligand can occupy a well-ordered position within the crystal, and its presence can be detected through differences in the crystal's diffraction pattern.^[210]

However, the crystal soaking experiments with FrsA_{T+TE} and FR-Core revealed several issues: The majority of the pre-manufactured crystals were destroyed during the soaking process, limiting the number of diffraction datasets. Furthermore, none of the datasets provided evidence of FR-Core binding, indicating a lack of complex formation.

It is important to admit that even if *holo*-form crystals had been obtained, the reliability of the data would have been questionable. The pre-existing crystalline lattice may constrain the protein's ability to undergo conformational changes necessary to accommodate ligand binding. This

limitation is particularly pronounced for flexible proteins, where ligand binding is likely to induce substantial structural rearrangements. Furthermore, the size of the ligand itself can pose challenges. Large ligands, such as FR-Core, may be unable to efficiently access the binding pocket during the soaking process due to steric hindrance and the potential for inducing protein conformational changes.^[210] These observations align with our previous analysis and NMR measurements presented in **Chapter 4.3.3.1**.

4.1.4. Generation of a Substrate Bound FrsA_{T+TE}

Since the crystallization trials only led to the formation of *apo*-form crystals, the aim was to gain a substrate-bound T-TE didomain. The aim was to load the T domain of the FrsA_{T+TE} construct (**4.1.3**) with the side chain. For this purpose, the side chain assembly assay with subsequent TE domain assay was slightly modified: FrsA was separated into the parts FrsA_{C+A} and FrsA_{T+TE} (**Figure 41A**). FrsA_{C+A} and FrsH were cloned with His-tag for affinity purification. The FrsA_{T+TE} construct was the same one that has already been described in **Chapter 4.1.3** and contained a His-tag and a TEV cleavage site for His-tag removal. To gain the substrate-bound TE domain, the side chain assembly assay was conducted in a one-pot assay, as before, but the FrsA_{T+TE} already underwent His-tag removal and was added at the beginning of the assay (**6.7.4**). Even though the TE domain was needed at a later point in time, the presence of the T domain is crucial for the sidechain formation, as it gets assembled on the Ppant arm of the T domain. Afterwards, the whole one-pot assay was incubated with Ni-NTA, to remove the His-tag-bound enzymes from the assay mix. Thereby needs to be mentioned, that the coexpressed FrsB does possess a His-tag itself, but has a high affinity to the A domain, which is why it sticks to it, and can also be removed. The other assay components should not lead to an unwanted reaction and could remain in the assay. According to the hypothesis, the Flowthrough sample should contain the FrsA_{T+TE} with the attached sidechain. If this sample comprised the expected purity, it could be used for a crystallization approach, where FR-Core is added, resulting in the formation of crystals of the FrsA_{T+TE} with the side chain bound to the T domain and FR-Core at the TE domain.

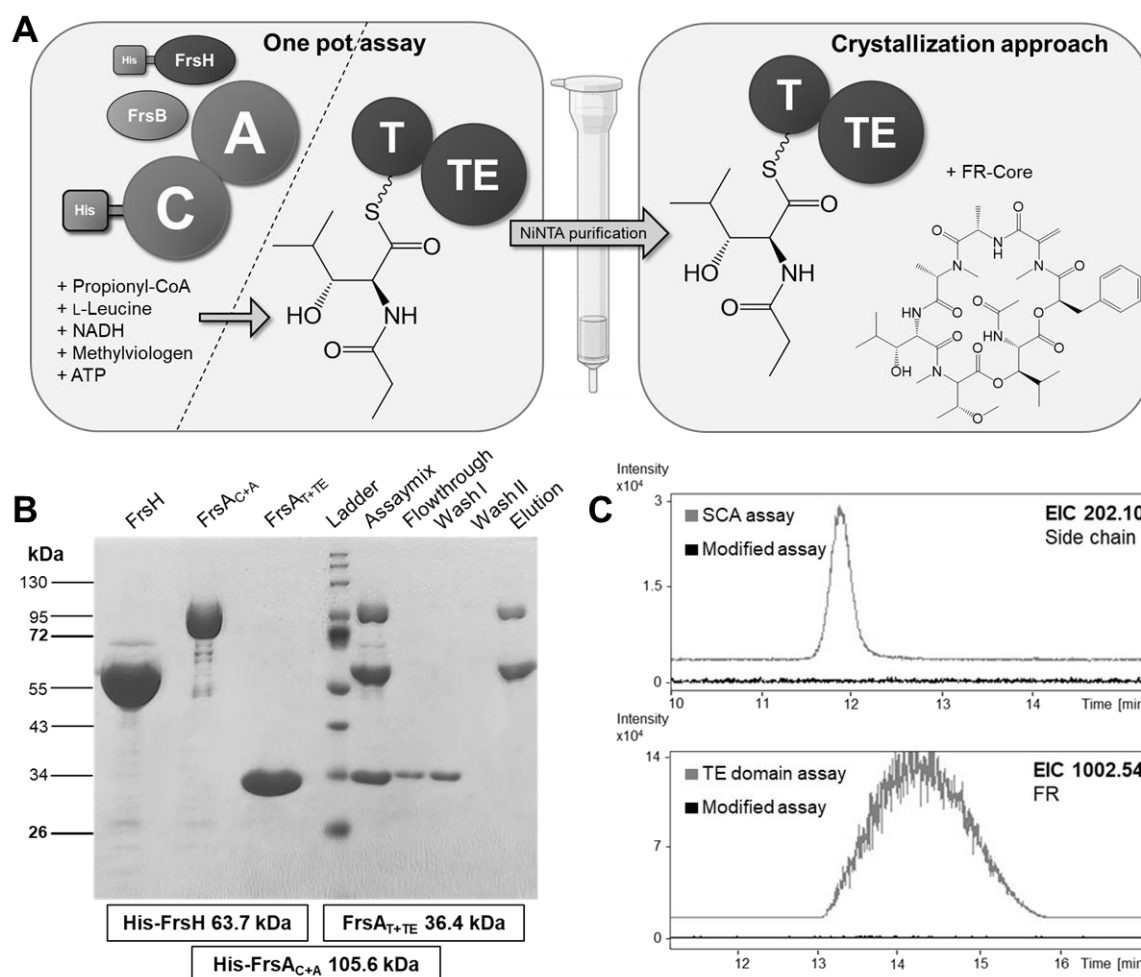


Figure 41: (A) Schematic illustration of the modified assay approach to gain a substrate-bound FrsA_{T+TE} (C = condensation domain, A = adenylation domain, T = thiolation domain, TE = thioesterase domain). (B) SDS-PAGE analysis of the purified assay enzymes (left) and the reverse affinity chromatography after assay (right). Expected protein sizes: His-FrsH = 63.7 kDa, FrsA_{T+TE} = 36.4 kDa, His-FrsA_{C+A} = 105.6 kDa. Applied volumes: protein samples and purification fractions = 10 μ L, ladder = 4 μ L. (C) Extracted ion chromatograms of the side chain assembly (m/z 202.10) and the transesterification assays (m/z 1002.54). The established assays are depicted in grey and the modified assays in black.

Figure 41B shows the SDS-PAGE analysis of the purified enzymes and the reverse affinity purification of the assay mix. The FrsH, FrsA_{C+A}, and FrsA_{T+TE} samples showed sufficient quality and quantity since the protein bands of the desired size had a high intensity and neglectable contaminations. The assay mix showed bands for every crucial protein. The reverse affinity purification achieved the expected result. The FrsA_{T+TE}, not possessing a His-tag, can be found in the flowthrough, and the enzymes harboring a His-tag could be found in the Elution sample. Here, it was also noticeable that a partial yield of FrsA_{T+TE} was found in the first Wash sample. However, this was a minor issue, as the sample did not show any contamination and could be combined with the Flowthrough sample before further processing. The SDS-PAGE results of the approach looked quite promising, as the desired proteins were detected in high concentrations and the separation was successful.

The next step was the investigation of the reaction products *via* HPLC-MS measurement (6.7.1, 6.7.2). The extracted ion chromatograms are depicted in **Figure 41C**. To estimate at which reaction step problems might have occurred, the assay was performed several times and extracted at two different points: after the side chain assembly and after the TE domain assay. To verify the correct execution of the assay, a positive control was conducted using the standard enzymes FrsA_{CAT} and FrsA_{TE}. The EICs of the positive controls showed a clear peak for the side chain and FR proton adducts in both assays. Unfortunately, those peaks were missing in the EICs of the modified assays, which led to the distinct conclusion that neither the side chain nor FR was formed by the modified enzymes. So, the absence of reaction products must be due to the modified reaction enzymes. As the only difference was the cleavage site, which was moved from the linker between the T and the TE domain to the linker between the T and the TE domain to the linker between the A and T domain, the assumption was that direct contact was needed between the A and the T domain. Through the separation of those two domains, the protein-protein interaction could be hampered by conformational changes or flexible ends, and thus prohibit the formation of the side chain. The absence of FR could be caused mainly due to the absence of the side chain. Nevertheless, there is the possibility, that the formation of FR does not occur despite the production of the side chain, since the stability of the bond between the T domain and the side chain is yet unknown and the repeated purification *via* Ni-NTA could also lead to the disruption of this bond.

Several optimization strategies could be explored to facilitate the isolation of a stable, substrate-bound didomain.

Since the transfer of the side chain onto FR-Core occurred in the initial *in vitro* assays, it could be assumed that sufficient phosphopantetheinylation of the T domain was achieved by the phosphopantetheinyltransferases (PPTases) encoded in *E. coli*. However, the structural analysis failed to confirm the presence of the Ppant arm. Studies providing elucidated structures of T domains harboring a Ppant arm indicate that incubation with a purified PPTase can ensure adequate phosphopantetheinylation.^[130] Consequently, incubating the T and TE domains with PPTase could enhance the formation of the Ppant arm, thus increasing the chance of generating a substrate-bound didomain.

To ensure protein-protein interactions within the FrsA module, the whole module could be expressed and subjected to *in vitro* assays. This approach would ensure that the proteins interact under native-like conditions. However, to facilitate subsequent purification steps and crystallization, a TEV protease cleavage site should be introduced into the linker region between FrsA_{C+A} and FrsA_{T+TE} domains. This would enable the proteolytic removal of the His-tag connected to the didomain FrsA_{C+A}, that are not relevant for structural studies.

It is conceivable, however, that introducing such a cleavage site could potentially disrupt the protein-protein interactions, as linker regions often play a crucial role in coordinating these interactions. Additional amino acids, such as "GGGS," might need to be incorporated into the linker to enhance the accessibility of the cleavage site for the TEV protease and ensure efficient cleavage. The significance of linker regions for protease accessibility has been previously discussed in **Chapter 4.1.2.5**.

4.2. Cryo-Electron Microscopy

Many TE domains occur as biological dimers.^[204] All data sets recorded by FrsA_{T+TE} confirmed that it also primarily forms dimers (**Figure 42**).

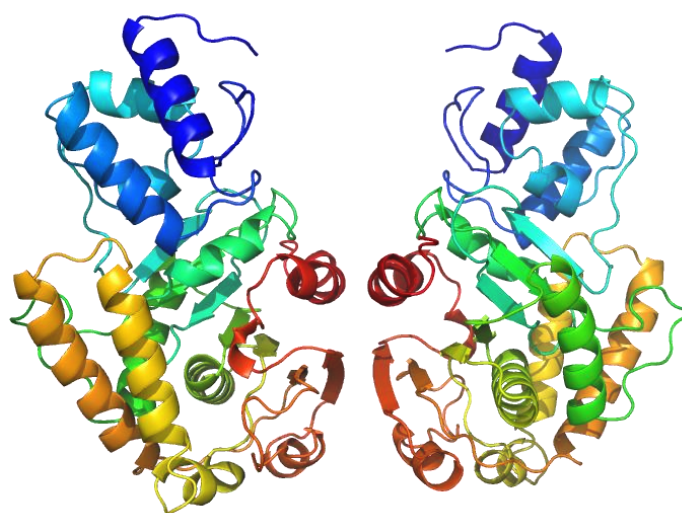


Figure 42: Illustration of the dimeric structure of FrsA_{T+TE} (rainbow color gradient: N-terminus = blue, C-terminus = red).

The, with 36 kDa, rather small didomain, reaches through its dimeric character size of 72 kDa, enabling a further method for structural elucidation: CryoEM.

Over the last decades, cryoEM became one of the most powerful tools for structure elucidation of biomolecules besides X-ray crystallography and NMR spectroscopy.^[211]

In this method, the aqueous biological sample is applied to an EM grid. The common way for grid preparation is the manual application on a support grid. The excess liquid is then automatically removed by blotting, resulting in a ~0.1 picolitres thin layer on the grid, which is subsequently vitrified by plunge-freezing in liquid ethane. Thus, the sample is trapped in a frozen-hydrated state without ice crystallization.^[212] The vitrified grids must be maintained in liquid nitrogen to prevent

recrystallization. The sample is then loaded in the cryo-electron microscope using a cryo-transfer holder. The high-resolution data collection is performed on a semi-automated system. Subsequently, the data are evaluated by the single-particle workflow, including the pre-processing by motion correction, defocus estimation, and particle collection. The latter is supported by machine learning and neural networking methods, however, a training set of manually picked particles is needed as guide for automated picking. With a suitable set of selected particles, initial 2D alignments and classification can be conducted, providing insights into sample homogeneity and orientation distribution. Appropriate 2D classes can further be used for the generation of a 3D map *ab initio*. Refinements of motion and optical distortion using intermediate structures provide better models for correlation. Thus, the resolution can be significantly increased by repetition of the pre-processing steps.^[213] The structure can then be generated using atomic model building with subsequent model refinement and validation.

The cryoEM method can overcome several drawbacks of the other methods and thus advance the structural elucidation of challenging macromolecules. Flash-freezing, for example, makes it feasible to preserve the protein in a near-native environment.^[211] But the probably greatest advantage over X-ray crystallography is the sample preparation. CryoEM requires relatively small amounts of sample^[214] and is not hampered by an increase in heterogeneity. Although processing is more straightforward and faster at higher purity and stability, heterogeneity often provides a deeper understanding of the macromolecular machinery.^[213] Before performing the cryoEM sample preparation, the sample purity and composition are analyzed using standard methods like SEC-HPLC, SDS-PAGE, and negative staining. CryoEM enables the structure determination of proteins with the size of several ten kDa to particles with many MDa up to whole cells^[211] but the lower molecular weight limitation is the key challenge for cryoEM. The main limitation is the minimum size of the macromolecule to be analyzed, which should not be less than approx. 50 kDa. The FrsA_{T+TE} didomain has a size of 36 kDa and is thus below the threshold. However, the crystallization experiments (4.1.3.5), as well as the NMR experiments (4.3.2.1.1), revealed the natural occurrence of FrsA_{T+TE} as a dimer, which consequently doubled the size up to 72 kDa, exceeding the threshold.

For the cryoEM experiment (6.6.13), the at -80 °C stored FrsA_{T+TE} was used, which had been purified earlier for the crystallization experiments (4.1.3.3). After gentle thawing, the samples were purified *via* SEC-HPLC (6.6.7). The chromatogram showed two not baseline-separated peaks, which were assumed to correspond to the dimer and the monomer. The 'dimer' peak was collected and concentrated. To achieve sufficient substrate binding, the protein was mixed with a 2-fold molar excess of FR-Core before it was transported on ice to the Ernst Ruska-Centrum for Microscopy and Spectroscopy with Electrons at the Research Center in Jülich (Germany). The

cryoEM grid was prepared under the supervision of Dr. Thomas Heidler, who performed the data collection trials. The data evaluation was conducted by Philipp Hendricks using cryoSPARC.

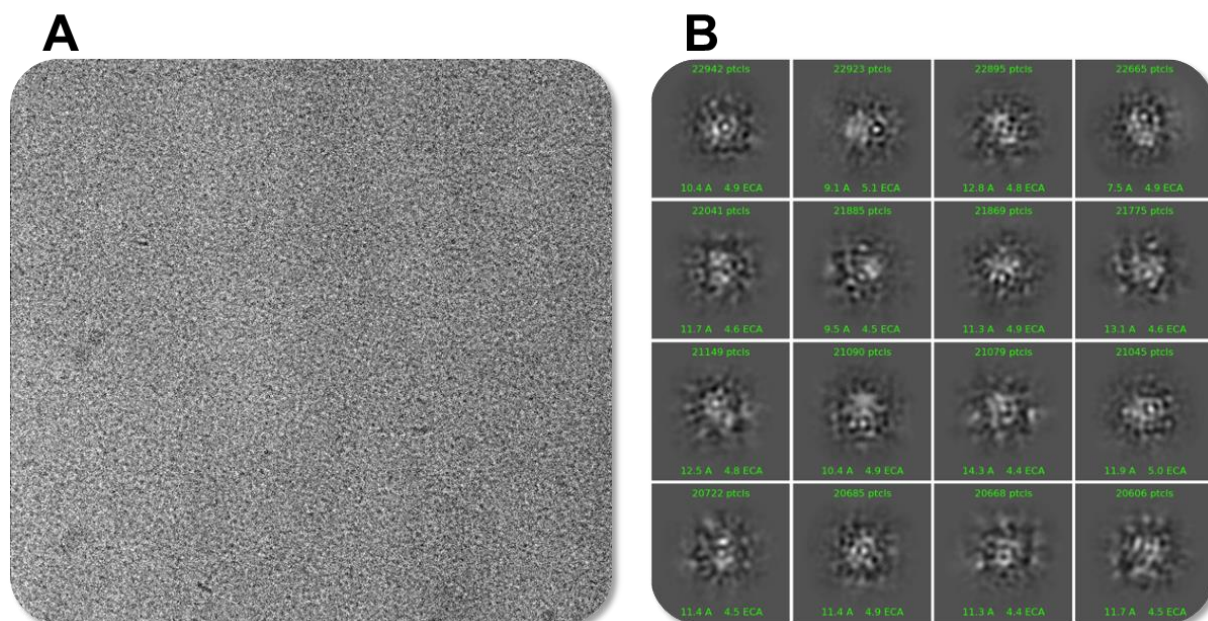


Figure 43: Results of the first cryoEM trial of FrsA_{T+TE}. The picture (**left**) shows the raw micrograph and the (**right**) shows a snippet of the 2D classification.

In the raw micrograph of one hole of the cryoEM grid (**Figure 43A**), distinct particles could be identified in high concentrations. This confirmed the successful sample preparation and the suitability of the blotting method. Nevertheless, it should be noted, that the observed particles seemed to have a minor size, likely corresponding to the FrsA_{T+TE} monomer respectively. This was confirmed by the fact, that the particle picking was significantly impeded by low resolutions (**Figure 43B**). Therefore, no 2D classes could be generated that would have been sufficient for three-dimensional structure reconstruction.

To overcome the size limitation, different approaches could be pursued: performing cryoEM with the whole NRPS module FrsA or supporting the dimerization of FrsA_{T+TE}. Dr. Hermes of our group already performed cryoEM experiments with FrsA in cooperation with Jonas Mühle of the Schertler workgroup at the Paul Scherrer Institute (Villigen, CH), resulting in the observation of particles of deviant size. The proposed improvement approach was the utilization of chaperones, such as the designed ankyrin repeat proteins (DARPin). They can be fused to the POI, thereby increasing the product size and supporting asymmetry.^[215]

4.3. NMR structural analysis of FrsA_{TE}

Proteins are inherently flexible macromolecules and continuously undergo conformational changes, depending on the environment and their reaction state. It is important to note that conventional protein crystallography yields a static snapshot of the protein structure. This approach may not fully capture the dynamic nature of proteins in their native environment. In contrast, Nuclear Magnetic Resonance (NMR) spectroscopy offers a powerful tool for investigating protein dynamics and conformational heterogeneity, providing a more comprehensive understanding of the macromolecular function and substrate interaction, and could support protein engineering.^[216] NMR spectroscopy is a well-established method to comprehend conformational and interactional dynamics at atomic resolution, which may also provide a detailed three-dimensional structure of the protein in solution. In contrast to X-ray and cryoEM methods, NMR is highly sensitive to chemical environment changes and more tolerant of different sample conditions.^[217] Nevertheless, the sample preparation differs, as isotope-labeled proteins are required for the measurement. Thus, proteins are expressed in minimal media with a defined chemical composition enabling the addition of labeled precursors. The initial aim was to investigate whether an expression of FrsA_{TE} in M9 minimal medium is possible (**Figure 44**).

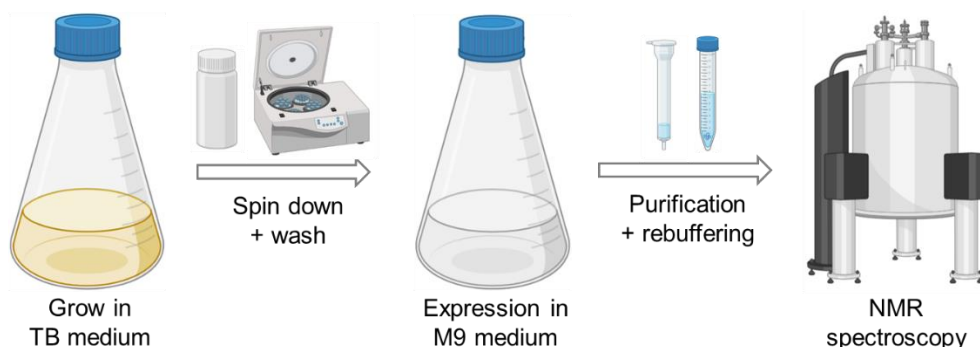


Figure 44: Schematic illustration of the protein preparation workflow for NMR structural analysis.

As a minimal medium, M9 is known to cause slow growth of *E. coli* cultures. To circumvent long incubation times, a spin-down approach was applied. Thus, the TB medium was inoculated with *E. coli* BL21 pHis8-TEV-FrsA_{TE} and incubated to the desired OD. Afterwards, cells were harvested, washed, transferred into M9 expression media, induced with IPTG, and incubated overnight. The protein was isolated, purified, and stored (6.6.1-6.6.6). NMR measurements were conducted by Christian Bonifer of the Glaubitz research group at the Institute of Biophysical Chemistry and Centre for Biomolecular Magnetic Resonance of the Goethe University Frankfurt (6.6.14).

If the M9 expression was successful, the next step was the expression optimization, resulting in a suitable yield for further analysis. In addition, the buffer conditions for the purified POI had to be adjusted with regard to protein stability and compatibility with NMR measurements. It was also necessary to determine the optimal storage conditions for shipping to preserve protein structure and functionality.

In the following chapter, all optimizations are described and discussed in detail and the results of the NMR spectroscopy are finally presented and analyzed.

4.3.1. Expression of FrsA_{TE} in M9 medium

For the first expression attempt of the FrsA TE domain in M9 medium, the *E. coli* BL21 pHis8-TEV-FrsA_{TE} was used. According to the spin-down protocol, the cultures were grown to an optical density at 600 nm (OD₆₀₀) of 0.6 at 37 °C and 200 rpm. The cells were then harvested by centrifugation at 4000 g, washed two times with M9 medium, and transferred into a baffled flask with 600 mL M9 medium. After induction with 0.4 mM IPTG (final concentration), cells were incubated for 16 h at 16 °C and 180 rpm. The protein was isolated by Ni-NTA affinity chromatography using the standard protocol (6.6.2). The purified protein was analyzed by SDS-PAGE (6.6.8) (Figure 45).

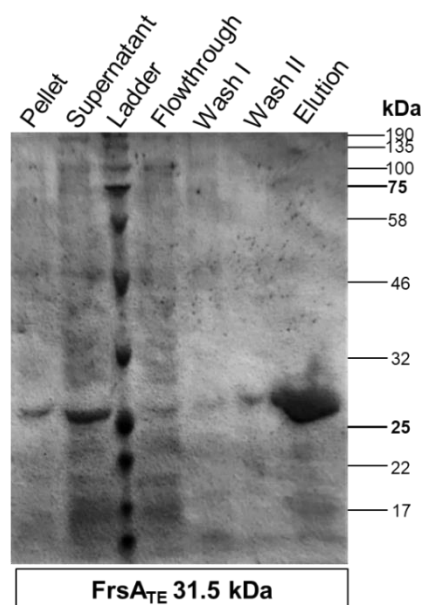


Figure 45: SDS-PAGE of first ¹⁵N-expression of FrsA_{TE} in *E. coli* BL21. Expected size for FrsA_{TE} = 31.5 kDa. Applied volumes: purification fractions = 15 µL, ladder = 4 µL.

The construct His-TEV-FrsA_{TE} possessed a size of 31.5 kDa and should be approximately at the height of the 32 kDa ladder band. As already shown in previous experiments, the band might occur

slightly below the expected height due to the faster migration of the sample compared to the ladder bands. Thus, a prominent band detectable in each purification fraction could be assigned to the desired construct His-TEV-FrsA_{TE}.

The cell lysis seemed to be successful, as the vast majority of the proteins could be found in the supernatant sample. Nevertheless, a minor percentage of the POI could also be detected in the pellet sample, leading to the conclusion, that the cell lysis process should be improved to maximize the protein yield. The Flowthrough sample indicated that the retention using Ni-NTA worked properly, as only a barely detectable POI band remained there. In contrast, the bands resulting from unspecific proteins could mainly be found in the Flowthrough sample, exhibiting the removal of those from the final sample. In the Wash samples, especially in the second one, a clear protein band was detectable, which was a sign of protein loss while purification. However, a prominent band with a high intensity could be detected in the Elution sample, leading to the conclusion that the expression in M9 medium with subsequent purification was successful. After purification, the final yield was 1.5 mg/600 mL expression culture, determined by photometric measurement.

The analysis showed that the expression of His-TEV-FrsA_{TE} in M9 is feasible, but the yield is not sufficient for NMR measurements (minimum of 500 μ M in 500 μ L required), especially considering, that expression with labeled components generally decreases the yield. Thus, the expression as well as the purification had to be improved in advance of the first labeling attempts.

4.3.1.1. Optimization of Expression in M9 medium

To increase the yield of His-TEV-FrsA_{TE}, various intervention points while expression were addressed. The following chapter discusses each optimization approach and describes the improvements in yield. Each approach was conducted once due to the limited handling but in direct comparison to the other tested parameters (6.6.1.3).

The first adjustable parameters were the expression conditions: growth phase during harvest, incubation temperature, and duration, culture volume, as well as the amount of IPTG for induction.

Based on the assumption that the cell culture was not in the right growth phase for harvest and induction, cultures were grown to an OD₆₀₀ of 0.9-1.1 in 1 L TB medium in all following optimization attempts. Furthermore, the incubation temperature while expression was set to 20 °C. Simultaneously, the expression duration was elongated from 16 hours up to 17, 18, and 20 hours. With these combined optimizations, the amount of His-TEV-FrsA_{TE} could be increased nearly twofold. The expression at 20 °C for 18 h yielded 2.85 mg per 600 mL M9 expression culture, making these the most promising conditions.

In the next step, the amount of IPTG for induction was rendered to 0.4, 1, and 1.5 mM IPTG in 600 mL culture each. Additionally, it was examined, if downscaling to 200 mL could enhance the production of the POI due to the reduced cell culture size. However, neither the downscaling nor the increased concentration of IPTG could increase the yield. On the contrary, the yields were reduced considerably in certain experiments, hence it was decided not to modify these parameters in the further process of optimization.

After the initial expression in ^{15}N -M9 medium (4.3.1), *E. coli* LOBSTR was tested as a more promising expression strain. It derives from the *E. coli* BL21 (DE3) strain and promises an improved purification of low-expressing proteins. The strain was engineered to reduce the background contamination through the modification of genes, whose proteins would reduce the His-tag affinity to Nickel resins.^[218] The new strain therefore offered a further potential for optimization, hence the pHis8-TEV-FrsA_{TE} plasmid was transformed in chemically competent cells of the *E. coli* LOBSTR strain. The test expression was conducted in 300 mL M9 medium under the defined optimal conditions (20 °C, 18 h). The suggested amount for expression induction in the *E. coli* LOBSTR strain was 0.1 mM IPTG, hence this was applied for the initial trial. The test expression yielded 4.34 mg, which corresponds to approximately three times the yield of *E. coli* BL21. Therefore, all further expressions were performed using the *E. coli* LOBSTR pHis-TEV-FrsA_{TE} strain.

4.3.2. Purification of ^{15}N -FrsA_{TE}

After the expression improvements led to a sufficient yield and the first purification refinements were made, preliminary NMR measurements of the ^{15}N -FrsA_{TE} were performed. For this purpose, the nitrogen source in the M9 medium was replaced by ^{15}N -ammonium chloride, which led to the formation of ^{15}N -FrsA_{TE}.

The first 2D-transverse relaxation-optimized spectroscopy (TROSY) experiment revealed a promising result (**Chapter 4.3.3, Figure 49**): Clear signals, that originated from the labeled nitrogen within the TE domain, could be detected. However, the signals were broadened, impeding their assignment and identification. This effect usually occurs when the measured protein is not present as a monomer but as a dimer.

For the first 3D measurements, the sample was also further concentrated up to ~15 mg/mL (500 μM), which led to degradation (4.3.2.1.2, **Figure 47**). Since the prevention of the natural structure and functionality is of utmost importance to investigate the protein dynamics, the protein structure should be stabilized by using buffer additives throughout the purification

process. Another major issue affecting the protein stability was the storage conditions during shipment. Freezing or insufficient cooling also leads to degradation, resulting in an increased protein loss. To prevent any negative impact on the sample during purification, storage, and shipping, various optimization approaches were conducted, which are described in detail in the following chapters.

4.3.2.1. Optimization of Purification

4.3.2.1.1. Dimerization

A straightforward method to determine whether the purified FrsA TE domain was mainly dimerized, is analytical SEC-HPLC (6.6.11). In this method, the protein solution is passed over a column packed with a porous matrix. The molecules are separated by the differential exclusion from the pores. Molecules with a greater size, cannot enter the pores, which results in a smaller available column volume, leading to faster elution. In contrast, small molecules enter more pores of the stationary phase and therefore require more volume to elute from the column. To evaluate the size and identify the proteins, their mass is calculated as a function of their retention time by the regression of a calibration curve.^[219]

Calibration was performed with a standard mix composed of thyroglobulin (660 kDa), ovalbumin (43 kDa), and ribonuclease (13.7 kDa) (**Figure 46A**). Individual measurements allowed the assignment of the peaks to the proteins in the mixture, which allowed only the latter to be measured in further experiments.

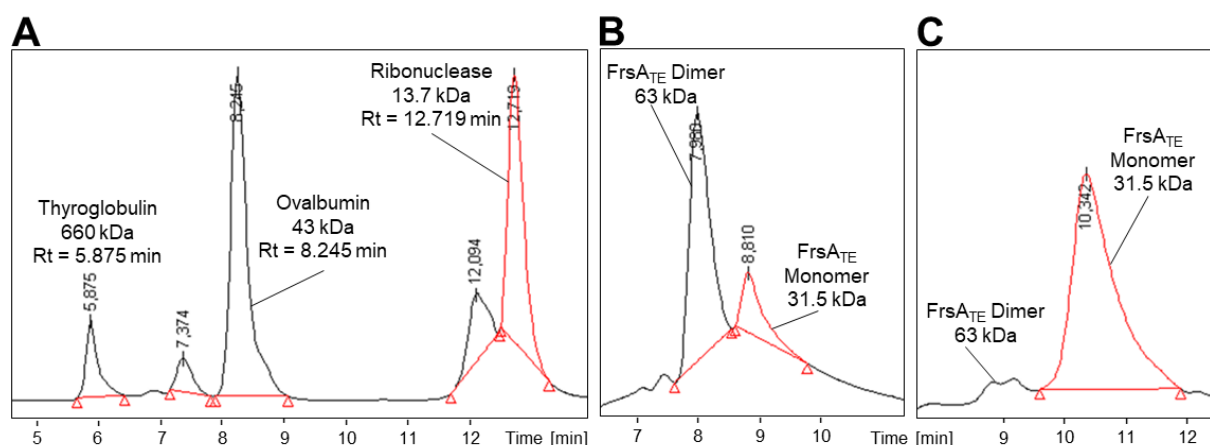


Figure 46: SEC-HPLC chromatograms of (A) calibration measurements with thyroglobulin (660 kDa), ovalbumin (43 kDa), and ribonuclease (13.7 kDa), (B) the measurement of FrsA_{TE} purified and stored using the standard methods, and (C) the FrsA_{TE} measurements after purification and storage using optimized conditions. Expected sizes: monomeric FrsA_{TE} = 31.5 kDa, dimeric FrsA_{TE} = 63 kDa.

In the first trial, FrsA_{TE} was purified under normal purification conditions (4 mL Wash buffer I; 4 mL Wash buffer II; 2.5 mL Elution buffer; final buffer = 25 mM HEPES, 50 mM NaCl, pH 7.5). SEC measurements showed two clear peaks in the expected retention time range. Conspicuously, the peak corresponding to the dimer showed a considerably higher intensity, which correlates with the results of the first NMR measurement. The peak with a higher retention time than ovalbumin could be assigned to the monomer of FrsA_{TE}, which seemed to be present to a much lesser extent (**Figure 46B**).

Two possible solutions considered to eliminate this ascertained aggregation were the variation of NaCl concentration (25-100 mM) in the buffer and/or the addition of DTT (5-20 mM) (**6.6.2.1**).

Using SEC-HPLC measurements, it was observed that the increased NaCl concentration did not affect aggregation. However, since the absence of NaCl had also no negative influence on the protein and NaCl could cause hindrance during the NMR measurements, it was completely removed from the final buffer. During the purification, the concentration of 25 mM NaCl was maintained for additional protein stability.

In contrast to NaCl, the addition of DTT showed a positive effect on the aggregation. Even small amounts of 5 mM DTT resulted in a significant shift in the monomer-dimer composition, as depicted in **Figure 46C**. A shift in the retention times of the peaks was noticed, but this effect was also visible when measuring the standard protein mixture in the modified buffer system. Thus, it could be concluded that DTT, as reducing agent, creates a reducing environment that prevents the formation of dimers. Since too high concentrations of DTT interfere with the resin of the Ni-NTA affinity purification and the positive effect could already be observed at the lowest DTT concentration (5 mM), it was refrained from the higher concentration. The addition of DTT was set to 5 mM final concentration in each purification buffer and in the final buffer.

4.3.2.1.2. Degradation

Another effect that could be observed during the first 2D-TROSY measurements was the presence of several weak signals with increasing intensity over time. These signals could be clearly distinguished from the signals originating from the protein by their shape. It was suggested that these were derived from degraded protein caused by instability or by proteases. To prevent the latter, protease inhibitors were added to each purification buffer after cell lysis. Before shipping, protease inhibitors were repeatedly added to the protein solution in the concentration specified by the supplier. This should prevent proteolytic digestion and contribute to an increased stability of the protein.

After the optimizations of purification with buffer additives, the NMR measurement was repeated to investigate how the modified conditions affected the protein quality. Therefore, FrsA_{TE} was expressed in 600 mL M9 medium without labeling. The first conducted measurement showed signals, which however were so weak that the protein was concentrated to 500 μ M, and the measurement was repeated. It was noticeable that the protein quality had decreased considerably.

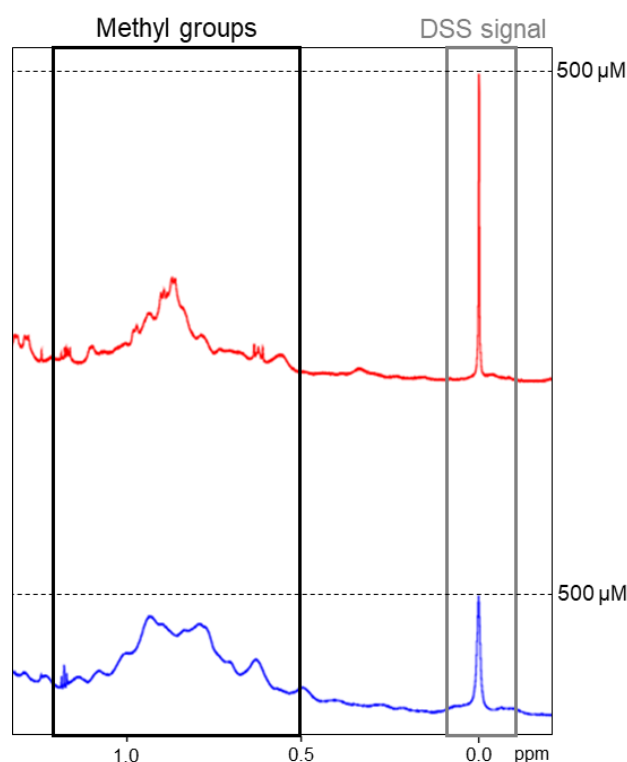


Figure 47: ¹H-NMR spectra with focus on the methyl group signals of FrsA_{TE} in concentrations of 100 μ M (red) and 450 μ M (blue). The dashed lines mark the expected signal strength of the target concentration of 500 μ M. The methyl group signals are framed in black and the DSS standard signal is framed in grey.

In **Figure 47**, the two ¹H-NMR spectra of the protein solution with different concentrations are comparatively depicted with focus on the methyl group signals. The second signal occurring at 0 ppm could be assigned to trimethylsilylpropansulfonate (DSS), which is used as a chemical shift standard in liquid-state NMR.^[220] Through appropriate referencing, the DSS signal was used to examine the concentration of the protein solution: If the intensity of the methyl group signals corresponds approximately to that of the DSS signal, it could be assumed that the concentration is around 500 μ M and complies to the requirements for further NMR measurements.

As already mentioned, the concentration of the first sample was quite low, which was also confirmed by the NMR measurement. The intensity of the methyl group signals is significantly lower than the DSS signal. However, it should be highlighted that the fine structure of the methyl

group signal looked distinct. In contrast, the second sample showed a clear deviation from the original signal form of the methyl groups. The signal was broader, and the fine structure was reduced. Nevertheless, the signal had a sufficient intensity, revealing that the desired concentration could be achieved.

The results exposed insufficient stability of the protein sample. The assumption was, that the degradation could be attributable to an unstable pH of the buffer. To monitor this, a minimal amount (1 mM final concentration) of imidazole was added to the final buffer right before the pH adjustment. These showed a pH-dependent chemical shift during NMR measurements, meaning that once the signal is identified, its position in the spectrum can be used to determine the pH value of the protein solution itself. Thus, a pH value deviation could be detected right at the beginning of the measurement, with an increasing shift over the measurement period.

To counteract this progress, a Tris buffer system was tested with various pH values adjusted with HCl and NaOH or glutamic acid and asparagine. Lower pH values are more favorable for NMR measurements, but previous tests revealed that the optimal pH for FrsA_{TE} is around 7.5. Thus, the pH values from 6.0 to 7.5 were tested in the buffer refinement experiments. Therefore, the unlabeled FrsA_{TE} was purified, rebuffed to the various buffers, transferred to NMR tubes, and stored for 10 days at 30 °C. Every day, the turbidity was evaluated, and at the time points after 0, 24, 72, and 168 h, NMR measurements were conducted to investigate the fine structure of the methyl group signals.

The result confirmed the assumption that the TE domain showed the highest stability at a pH value of 7.5. The NMR samples with a lower pH became turbid after 24 h, whereas the samples with a higher pH value > 7.0 had no detectable turbidity up to 168 h after preparation. Furthermore, 100 mM Tris buffer was able to maintain the adjusted pH over the entire time. How the pH value was adjusted did not seem to have any influence on the stability of the protein, which is why NaCl was completely removed from the final buffer and the pH was adjusted with glutamic acid and asparagine in the further process, as these do not interfere with the NMR method. The final buffer composition was thus set to 100 mM Tris, 5 mM DTT, 1 mM imidazole, protease inhibitors, and pH 7.5, which contributed to increased protein stability sufficient for NMR measurements.

4.3.2.1.3. Storage

Since the purified protein had to be sent to the Goethe University Frankfurt for NMR measurements, storage also played an important role in the protein stability investigation. Therefore, the FrsA TE domain was expressed in TB medium and purified by the standard

protocol. A part of the pre-cell lysis pellet was frozen to prepare the 'fresh' sample right before the SDS-PAGE analysis. After concentration to 6.5 mg/mL, the protein was equally split into multiple samples, which were then stored at different temperatures (RT, 4 °C, -20 °C, -80 °C) for 7 days. The samples stored in the -20 °C and -80 °C freezer, were flash-frozen with liquid nitrogen. For each temperature, three samples were prepared, of which two were mixed with glycerol (10% and 20%, final concentration). After one week, the frozen samples were thawed, and all samples were prepared according to the standard protocol and analyzed by SDS-PAGE (6.6.8).

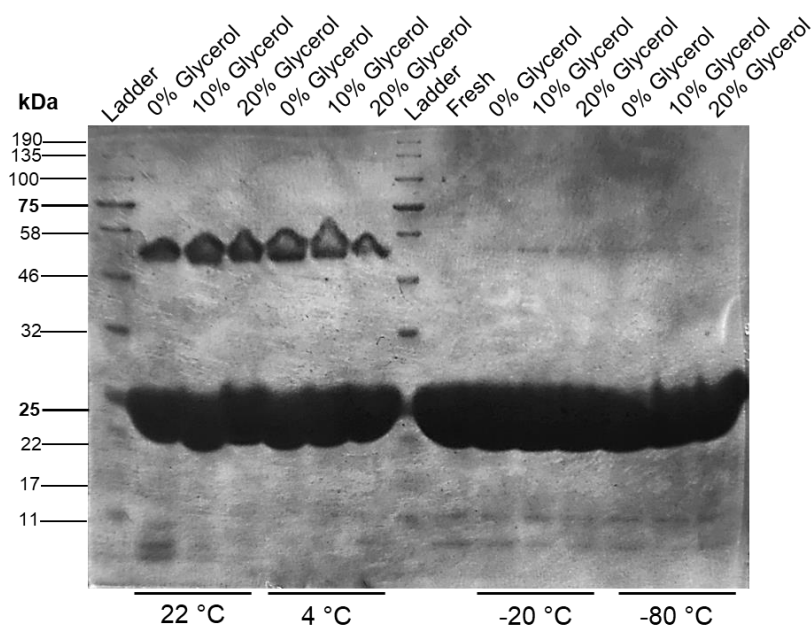


Figure 48: SDS-PAGE analysis of FrsA_{TE} samples under different storage conditions (temperature: 22, 4, -20, -80 °C; glycerol content: 0, 10, 20%). Expected size of FrsA_{TE} = 31.5 kDa. Applied volumes: protein samples = 10 µL, ladder = 4 µL.

The gel, depicted in **Figure 48** revealed that flash-freezing with subsequent storage at -20 °C or -80 °C seemed to be the gentlest way to store the TE domain of FrsA. It is noticeable that there was no detectable difference between the samples of different glycerol concentrations. In the samples stored at room temperature, clear bands could be seen at a height of approx. 50 kDa. They showed a deviating migration behavior, as the bands were deformed. These could indicate that the storage at higher temperatures led to the formation of protein aggregates. The bands could also be detected in the frozen stored samples but with a much lower intensity.

Based on these results, the freshly purified proteins were flash-frozen and shipped to Frankfurt on dry ice for the first measurements. However, it was discovered that the thawing and concentrating of the protein samples before measurement led to increased protein precipitation, resulting in a high protein loss.

The storage experiment was performed before the buffer optimization, which is why the buffer composed of 25 mM HEPES, 50 mM NaCl, and a pH of 7.5 was used. This explains the instability of the protein at higher temperatures: HEPES in such small quantities could not keep the pH stable over a long period, which is why the protein degraded. During the process of buffer optimization, the FrsA_{TE} samples at different buffer conditions were stored at 30 °C. The Tris buffer (100 mM Tris, pH adjusted to 7.5 with arginine and glutamic acid) achieved no visible protein degradation within the first 3-7 days. Even the NMR measurements did not disclose any degradation effect. Since the improved buffer conditions had already achieved sufficient protein stability, the protein was not frozen for storage and transport in order to prevent protein damage. Consequently, the protein was stored in the optimized buffer system at 4 °C and shipped overnight chilled on ice.

4.3.3. NMR measurement

To assess protein quality and labeling efficiency, ¹⁵N-BEST-TROSY NMR spectra were acquired for ¹⁵N-FrsA_{TE} (blue) and ¹⁵N-FrsA_{TE} (red) mixed with a two-fold molar excess of FR-Core. The results are depicted in **Figure 49** (6.6.14).

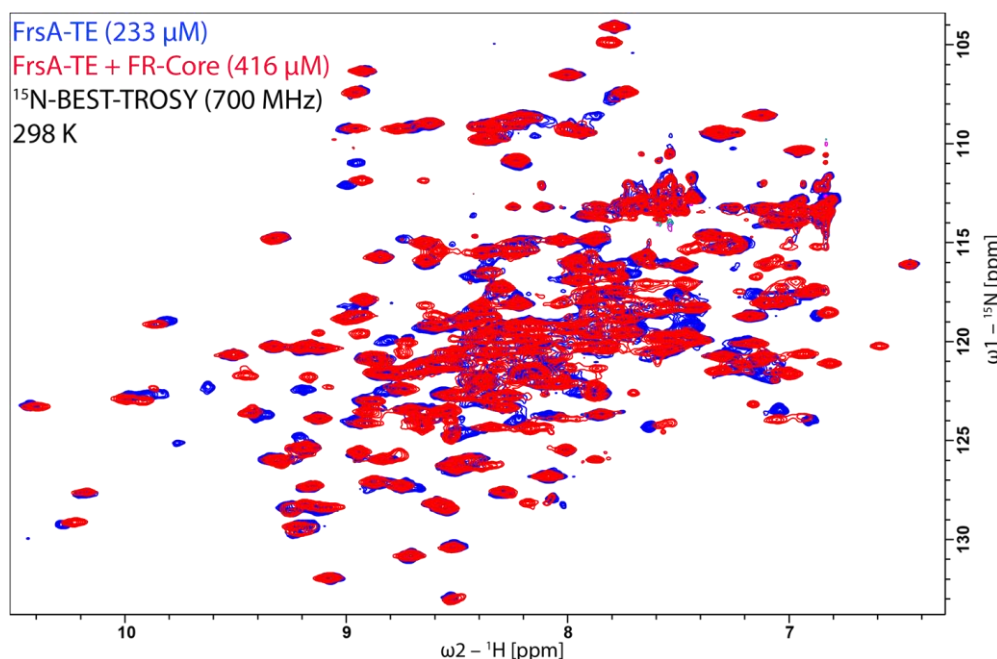


Figure 49: ¹⁵N-BEST-TROSY NMR spectra (700 MHz) of ¹⁵N-FrsA_{TE} (233 μM, blue) and ¹⁵N-FrsA_{TE} with a twofold molar excess of FR-Core (233 μM protein, 414 μM substrate, red).

Initial spectral analysis revealed the presence of clear signals, indicating successful protein expression and labeling. However, the observed signal broadening suggested the predominant formation of FrsA_{TE} dimers, which could potentially hinder subsequent resonance assignments

due to signal overlap. To address this, optimization of buffer conditions to disrupt dimerization was conducted (4.3.2.1.1). Concurrently, the emergence of several weak signals, particularly pronounced in samples subjected to prolonged storage, indicated proteolytic degradation. In purification and storage buffer optimization approaches, the application of protease inhibitors was tested to prevent this adverse effect.

The comparison between both measured spectra showed that the addition of FR-Core resulted in discernable chemical shifts for specific residues within FrsA_{TE}. These findings were indicative of either induced structural dynamics or direct protein-substrate interactions. To identify distinct amino acid residues involved in these processes, a comprehensive resonance assignment was conducted. This comprised the acquisition of three-dimensional NMR spectra from a uniformly labeled protein sample, complemented by two-dimensional NMR experiments on combinatorially labeled variants.

4.3.3.1. NMR assignment and elucidation of structural dynamics

Despite successful NMR signal detection, comprehensive resonance assignment was achieved for only one-third of the FrsA_{TE} amino acid sequence. Potential key structural regions, including the lid structure and the highly variable loop characteristic of TE domains, remained unassigned. These regions are hypothesized to play pivotal roles in protein dynamics and substrate interactions, making their structural elucidation essential for understanding the mechanism of action of FrsA_{TE}.

While the active site residues are well-characterized within the TE domain family, the absence of their assignment in the current experiments did not significantly hinder the overall interpretation.

The incomplete assignment of 3D NMR spectra in this study could be attributed to spectral overlaps, particularly pronounced in regions of high protein density, hindering the unambiguous identification of individual resonances. Furthermore, dynamic processes, such as conformational exchange and side chain flexibility, resulted in broadened or split resonance lines, further complicating the assignment process. Despite implementation of various heteronuclear experiments to enhance spectral resolution, limitations in experimental data, including suboptimal signal-to-noise ratios restricted the assignment to approximately one-third of the total protein sequence.

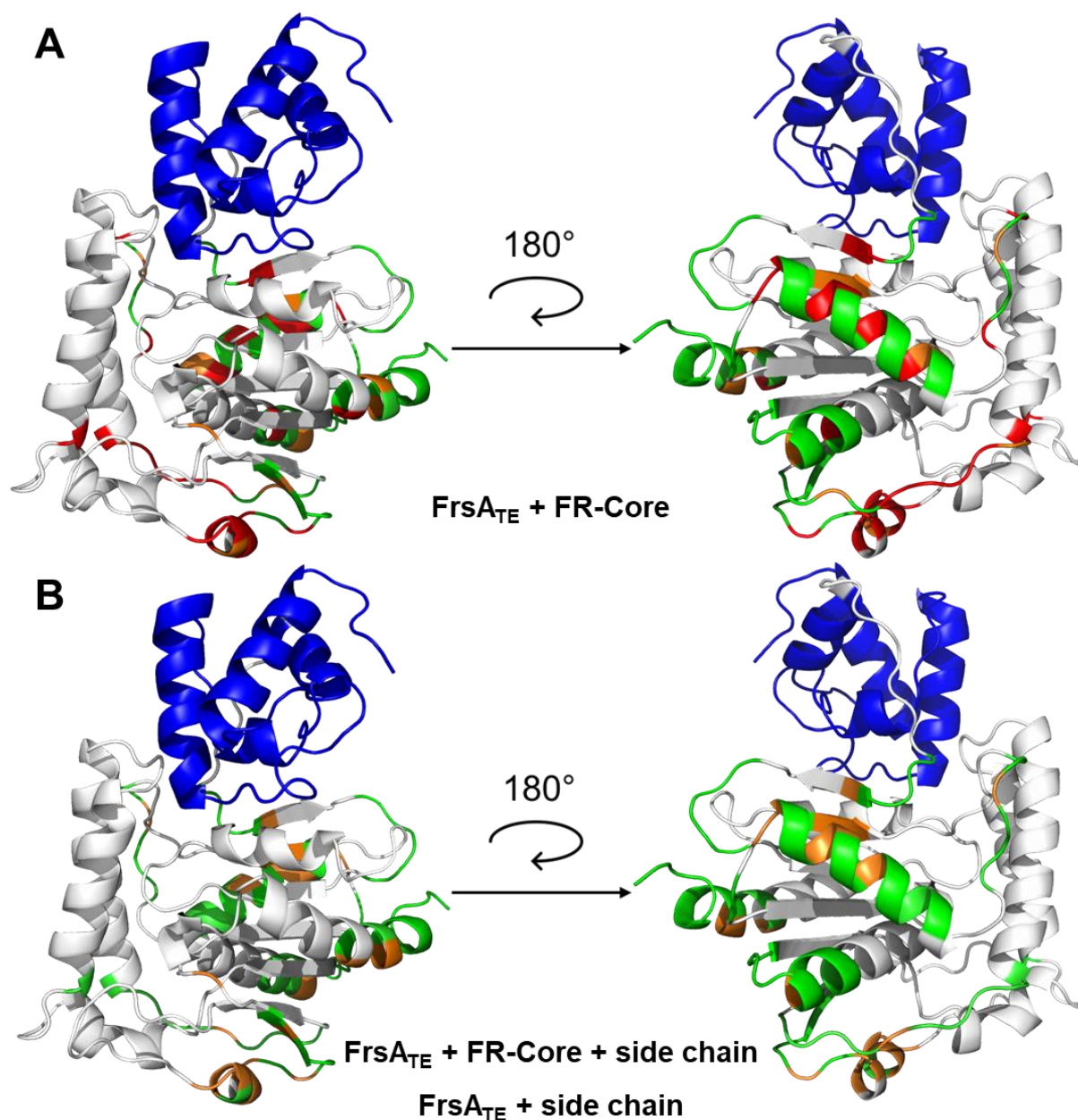


Figure 50: Visualization of the structure of FrsA_{TE} with color-coded residues that were identified by comprehensive resonance assignment, which showed shifts in the signals between the *apo*-form and **(A)** FrsA_{TE} + FR-Core and **(B)** FrsA_{TE} + sidechain (+ FR-Core). Color code: blue = Thiolation domain (not comprised in the protein provided for NMR measurements), white = regions that could not be assigned, green = residues showing no signal shifts, orange = residues showing weak signal shifts, red = residues showing distinct signal shifts.

Nevertheless, valuable insights were gained from NMR analysis. The addition of either FR-Core or side chain induced distinct chemical shifts of several signals, indicative of conformational changes while substrate binding (**Figure 51A**). Notably, FR-Core elicited pronounced shifts in specific residues, whereas the side chain caused only minor alterations. A comparative analysis revealed that the presence of FR-Core induced a substantial structural rearrangement, as evidenced by significant chemical shifts. Conversely, the addition of the side chain to a pre-formed

FrsA_{TE}-FR-Core complex appeared to revert protein structure to a state resembling that of the FrsA_{TE}-side chain complex alone. The attenuation of the chemical shifts through the addition of the side chain may be due to the fact that the chemical reaction catalyzed by FrsA_{TE} took place and conversion of the two substrates occurred, which are subsequently released and the enzyme reverts to its initial state (**Figure 50B**).

The observations were consistent with the proposed mechanism of FrsA_{TE} catalysis, where substrate presence and binding cause conformational changes to facilitate the entry of FR-Core into the active site, thus enabling the transesterification reaction. The inability to obtain co-crystals of FrsA_{TE} with its substrate FR-Core may be attributed to the enzyme's preferred *apo*-crystal formation, what could have prevented substrate binding. The NMR data presented herein provide compelling evidence for a dynamic interplay between FrsA_{TE} and its substrates, emphasizing the need for further structural studies to elucidate the molecular basis of catalysis.

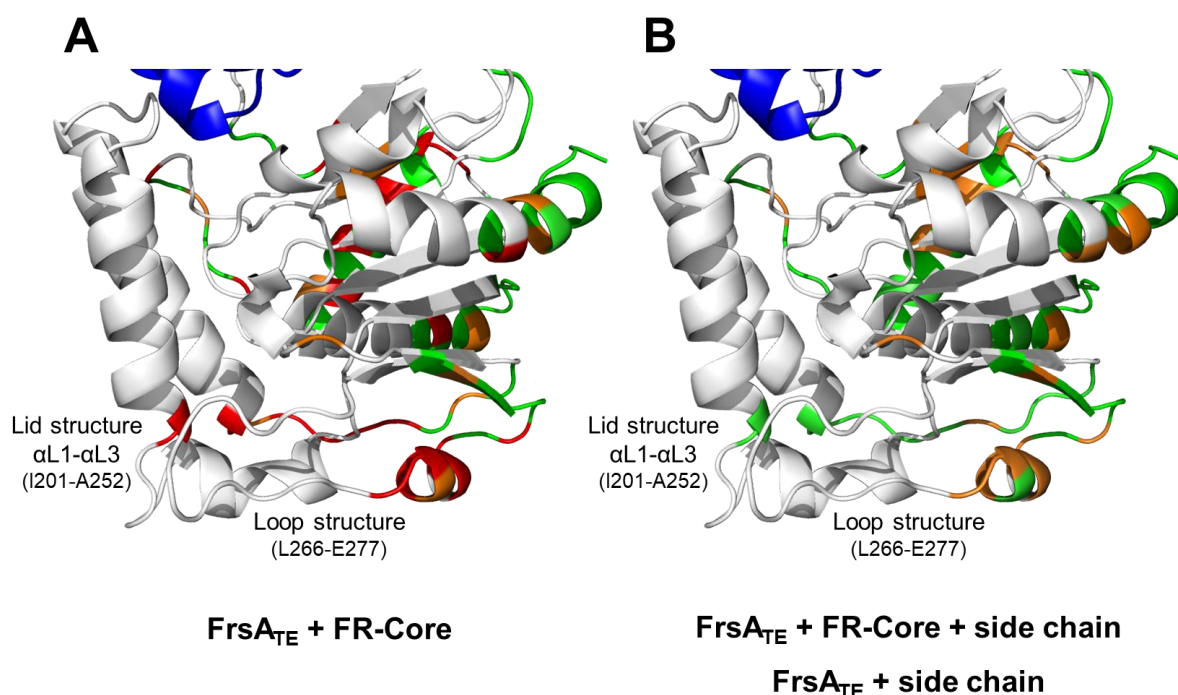


Figure 51: Visualization of the structure of FrsA_{TE} with color-coded residues that were identified by comprehensive resonance assignment, which showed shifts in the signals between the *apo*-form and **(A)** FrsA_{TE} +FR-Core and **(B)** FrsA_{TE} + sidechain (+ FR-Core) with focus on the lid (αL1-αL3) and the loop structure (between β6 and β7). Color code: blue = Thiolation domain (not comprised in the protein provided for NMR measurements), white = regions that could not be assigned, green = residues showing no signal shifts, orange = residues showing weak signal shifts, red = residues showing distinct signal shifts.

While a complete assignment of the lid structure proved elusive, specific residues at the N-terminus of αL2 (C212) and the C-terminus of αL3 (A249-G258) were successfully identified (**Figure 51**). NMR experiments conducted with FrsA_{TE} in the presence of FR-Core revealed

significant chemical shifts of these residues compared to the *apo*-form. This finding unequivocally indicates substantial alterations in the local chemical environment, suggesting that these regions undergo pronounced structural changes upon FR-Core binding.

It is reasonable to postulate that these conformational changes extend across the entire lid structure. However, the absence of a complete resonance assignment precludes a definitive conclusion. Intriguingly, upon the addition of the side chain substrate, these induced structural changes were completely reversed, returning the protein to a conformation closely resembling the *apo*-state.

Furthermore, while many assigned regions exhibited attenuated chemical shifts upon side chain addition, no discernible structural deviations from the *apo*-form were observed. This suggests that the transesterification reaction may have occurred, leading the enzyme to revert to an inactive/passive state (**Figure 51B**).

Similar observations were made for the variable loop, which borders the active site from below. Although only the helix α D (A278-Y284) at the C-terminus of the loop could be unambiguously assigned to NMR signals (E277-D293), the N-terminal region also appeared to undergo significant structural changes. This flexibility is indicative of a potential role in substrate interaction and coordination. However, due to the incomplete assignment, this hypothesis remains speculative. The observed structural dynamics in the absence of bound FR-Core suggest that the variable loop may possess inherent flexibility, independent of substrate binding.

In a protein region with high sequence coverage, an intriguing observation was made. The α B helix, which could be fully assigned, revealed that the amino acids oriented towards the core structure undergo significant chemical shifts upon substrate binding with FR-Core. Additionally, the opposing β sheets, β 1, and β 3, also exhibited shifts in their corresponding signals. Thus, it could be strongly suggested that these structural elements undergo conformational changes caused by the presence of FR-Core. While the exact nature of these structural changes could not be definitively determined from the acquired data, it was clear that the binding of FR-Core induces a structural reorganization, influencing the whole protein structure (**Figure 52A**). The localization of the shifting residues was inconsistent with the hypothesis that FR-Core binds and influences exclusively to the opposite side of the protein. However, a direct interaction with FR-Core could be excluded, due to the fact, that the residues are directed towards the interior of the protein core structure, thus making substrate binding unlikely, and that helix α B possess an unfavorable spatial orientation, which is not in close proximity to the active site.

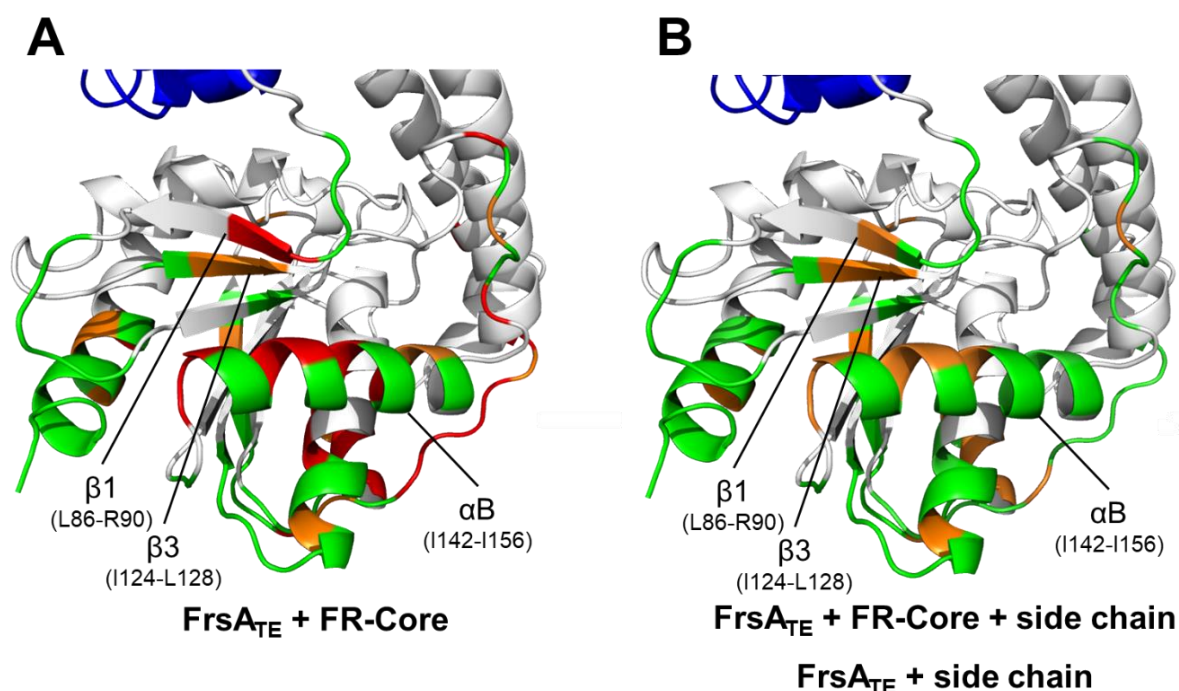


Figure 52: Visualization of the structure of FrsA_{TE} with color-coded residues that were identified by comprehensive resonance assignment, which showed shifts in the signals between the *apo*-form and **(A)** FrsA_{TE} +FR-Core and **(B)** FrsA_{TE} + sidechain (+ FR-Core) with focus on the 'back' of the protein (αB , $\beta 1$, $\beta 3$). Color-code: blue = Thiolation domain (not comprised in the protein provided for NMR measurements), white = regions that could not be assigned, green = residues showing no signal shifts, orange = residues showing weak signal shifts, red = residues showing distinct signal shifts.

The results were more consistent with the depicted model in which the lid structure opens, as shown in **Figure 34** of **Chapter 4.1.3.5**, leading to rearrangements of the structure on the 'backside' of the protein.

A loop located between the αB helix and the lid structure ($\alpha L3$), which was also almost completely assigned, contains residues exhibiting diverse and significant chemical shifts. These shifts do not provide a clear indication of how the structure behaves during substrate binding. The amino acids G134 and A139, oriented towards the $\alpha L3$ structure, could potentially approach or influence this structure, and *vice versa*. This hypothesis is supported by the observation that these specific residues do not exhibit any chemical shift changes in the FrsA_{TE} + FR-Core + side chain sample. However, the intervening amino acid L138, which is in close proximity to $\alpha L3$, does not undergo any conformational changes, even in the side chain sample (**Figure 52B**). These contradictory findings do not support the direct involvement of the loop in coordinating the lid structure.

In conclusion, the NMR data indicate that substrate binding induces structural changes not only in the immediate environment of the binding site but also in distant regions of the protein. While the precise nature of these dynamic effects remains to be elucidated, it is clear that they play a crucial role in the catalytic mechanism of FrsA_{TE}.

4.4. Investigation of FrsA substrate specificity

FR900359 is a selective Gq inhibitor that directly targets the Gαq subunit, effectively bypassing the structural diversity of GPCRs. This makes it an attractive tool for pharmacological studies, showing promising results in *in vivo* models of deviant Gq signaling. Besides FR, other naturally occurring FR derivatives harboring an altered side chain were discovered (**Chapter 2.2**). These FR analogs not only validated the critical role of the side chain in Gαq inhibition but also provided a foundation for structure-activity relationship studies. By systematically varying the FR side chain, one could gain deeper insights into the molecular determinants influencing the binding affinity and inhibitory potency of these compounds. The studies may highlight the potential of FR analogs as diverse chemical substances to further elucidate the complex molecular mechanisms underlying Gαq-mediated signaling.

Dr. Hermes and Dr. Wirtz established an *in vitro* assay wherein the side chain is biosynthetically assembled from precursors by purified FrsA_{CAT} and FrsH, and subsequently attached to FR-Core by FrsA_{TE}. Notably, they successfully generated the naturally occurring FR analogs, FR2 and FR5, *via* substrate exchange of the FrsA C starter domain. In this work, the native substrates of the FrsA C domain, propionyl-CoA, and the FrsA A domain, L-leucine, will be systematically replaced with structurally diverse acyl-CoAs and amino acid isomers, respectively. Upon validation of substrate acceptance and incorporation into the side chain, combinatorial approaches will be employed to generate novel side chain variants. The resulting modified FR analogs will be subjected to the TE domain assay to comprehensively assess FrsA_{TE} substrate specificity (**Figure 53A**).

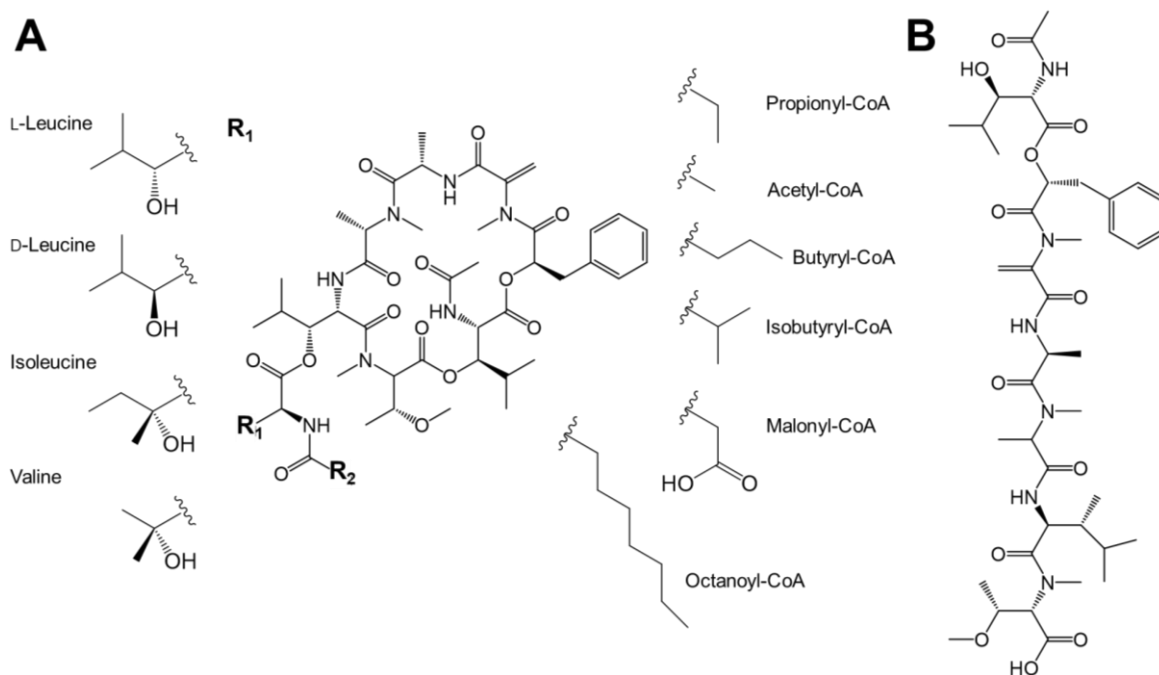


Figure 53: Graphical representation of **(A)** the generation of FR analogs by substrate exchanges of the established assays (used substrates are indicated as the residue R1 for amino acids and R2 for acyl-CoAs) and **(B)** linearized FR-Core, which is the expected product of a *C. vaccinii* mutant deficient in FrsA and FrsG8_{TE}.

Furthermore, the potential of generating linearized FR-Core (**Figure 53B**) as an alternative substrate for FrsA_{TE} will be investigated. Successful attachment of the side chain to linearized FR-Core would yield a linear FR molecule, with the potential to provide valuable insights into the catalytic mechanism of FrsA_{TE}. To enable the production of linearized FR-Core, a *C. vaccinii* mutant deficient in FrsA and FrsG8_{TE}, two key enzymes involved in side chain assembly, transesterification, and FR-Core cyclization, will be constructed. If this mutant strain accumulates linearized FR-Core, the compound will be isolated through upscaled fermentation and purification processes and characterized as a novel substrate for the TE domain assay.

4.4.1. Purification of enzymes FrsH, FrsA_{CAT}, and FrsA_{TE} for *in vitro* assay

The enzymes for the *in vitro* assays were expressed in *E. coli*, and isolated by cell lysis with subsequent purification using the affinity chromatography method (6.6.1-6.6.4). As Dr. Hermes and Dr. Wirtz provided the expression strains and optimized the purification^[147,163], the process will not be discussed in detail below.

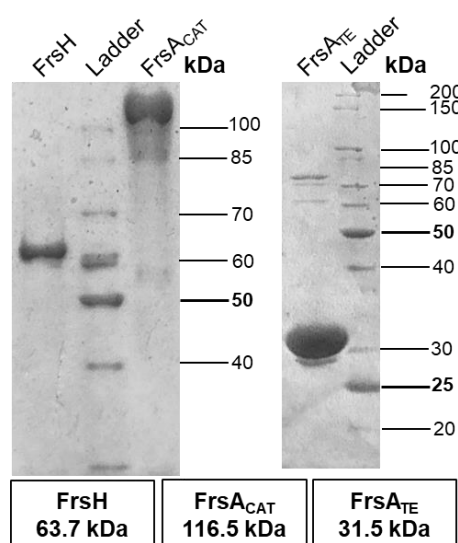


Figure 54: SDS-PAGE analysis of the protein samples FrsH (63.7 kDa), FrsA_{CAT} (116.5 kDa) and FrsA_{TE} (31.5 kDa) used for the side chain assembly and transesterification assay. Applied volumes: protein samples = 5 μ L, ladder = 3 μ L.

Since the purification of the assay enzymes has already been discussed in detail in **Chapters 4.1.2.4** and **4.3.1**, it should only be emphasized here that the enzymes required for the assay could be obtained in sufficient quantity and quality to perform the *in vitro* assays to investigate FrsA substrate specificity.

FrsH and FrsA_{CAT} samples were subjected to SDS-PAGE (6.6.8) using a 12% acrylamide gel to facilitate the separation of higher molecular weight proteins. Due to the limited gel's resolving power, the co-expressed FrsB (8.12 kDa) was not detectable. To verify that the absence of reaction product was not attributable to FrsB deficiency, a positive control was included in each assay using a freshly prepared enzyme batch. As FrsB is crucial for the correct functioning of the A domain, a control showing product formation confirmed the presence and activity of all involved proteins for each assay performed with this enzyme batch.

The purification of FrsA_{TE} was analyzed by SDS-PAGE using a 15% acrylamide gel to optimize the resolution of lower molecular weight proteins. While the FrsA_{TE} sample displayed multiple bands, indicating the presence of contaminating proteins, their low intensity suggested a negligible impact on subsequent assays.

4.4.2. *In vitro* assay optimizations

Prior to investigating the substrate promiscuity of the FrsA domains, the established *in vitro* assay was subjected to optimization to enhance product yield and facilitate experimental handling. Additionally, modifications were implemented to enable kinetic analysis of the FrsA_{TE}-catalyzed reaction. All optimization experiments were conducted by Bachelor student Georgina Kowalski (6.7.1-6.7.2).

Initial efforts focused on refining propionyl-CoA concentration. The original protocol employed a final concentration of 200 μ M propionyl-CoA; however, HPLC-MS analysis indicated a low substrate conversion efficiency. Reducing propionyl-CoA concentration to 100 μ M, while maintaining other assay parameters did not compromise side chain production to any extent (Figure 55A). Thereby, this modification was inherited for all the following assay approaches. Furthermore, it has to be highlighted, that the *in vitro* side chain assembly could be performed successfully and reproducibly, as the side chain could be distinctly detected by LC-MS measurement at the desired m/z .

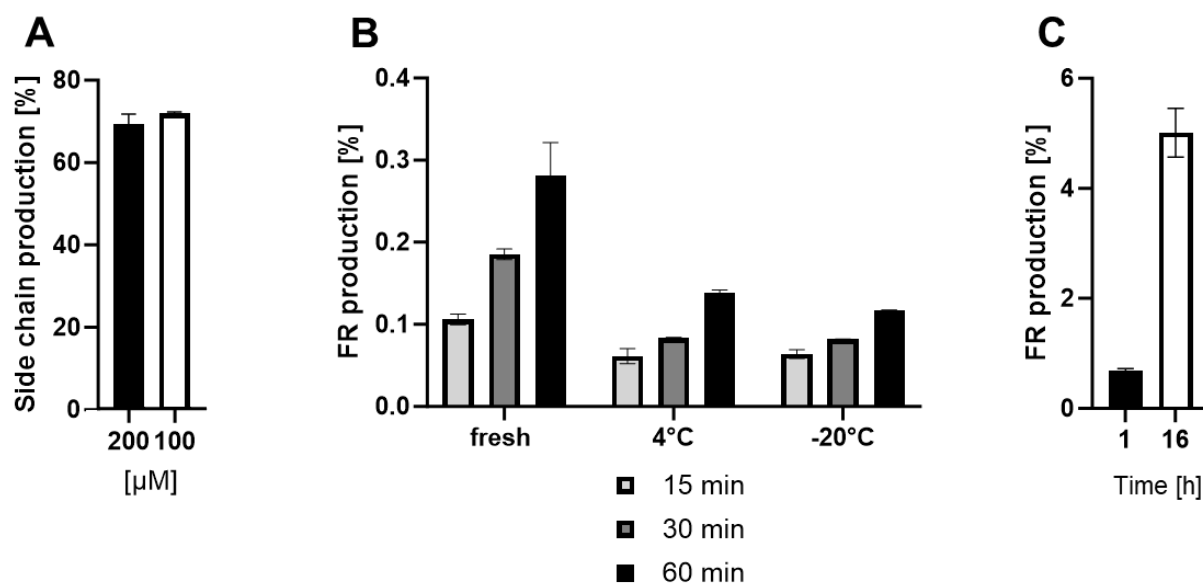


Figure 55: Graphical evaluation of the HPLC-MS measurements of (A) the side chain assay with altering propionyl-CoA concentration (calculated relative to the applied amounts of substrates [%], $n = 2$), (B) the average conversion of FR-Core to FR with altering protein storage conditions (fresh = no storage, 4 °C = stored overnight at 4 °C, and -20 °C = stored overnight at -20 °C) and incubation durations (15, 30, 60 min) ($n = 2$), and (C) the average conversion of FR-Core to FR [%] after 1 and 16 h at an FR-Core concentration of 100 μM ($n = 2$).

To optimize the assay workflow, the impacts of incubation time and enzyme storage on side chain yield were assessed. Reducing incubation time led to a significant decrease in FR formation, irrespective of enzyme storage conditions (4 °C, -20 °C) (**Figure 55B**). While enzyme storage resulted in a major reduction in FR yield compared to freshly prepared enzymes, no clear temperature-dependent effect was observed. To maintain optimal enzyme activity, fresh enzyme preparations were employed for all subsequent experiments. Previous studies by Dr. Wirtz demonstrated that extending incubation time beyond 60 minutes did not increase side chain formation; thus, a 60-minute incubation period was maintained for the side chain assembly assay.

To characterize the kinetics of the FrsA_{TE}-catalyzed reaction, time-course experiments were performed. Side chain assembly and subsequent TE domain assays were conducted with product extraction after 1 and 16 hours. HPLC-MS analysis revealed a sevenfold increase in FR production after 16 hours compared to the 1-hour time point (0.65% vs. 5.05%). While these results indicate a substantial increase in FR formation over time, the precise reaction completion time remains undetermined.

Accurate data interpretation was hindered by the relative determination of FR quantification based on initial FR-Core concentration. The maximum achievable FR yield, as represented by the 5.05% value in **Figure 55B** and **Figure 55C**, could not be further increased through optimization of reaction conditions or using the whole FrsA module. Moreover, the complex multi-step reaction

involving three enzymes precluded precise quantification and identification of rate-limiting steps. Consequently, it was refrained from kinetic analysis of FrsA_{TE} using the established *in vitro* assay.

4.4.3. *In vitro* synthesis of FR analogs

The established *in vitro* assay, comprising side chain assembly followed by TE domain-catalyzed attachment to FR-Core, provided a suitable fundament for generating FR analogs through substrate exchange. The native side chain, biosynthesized by FrsA_{CAT} from L-leucine and propionyl-CoA and subsequently hydroxylated by FrsH, served as a template for analog design. To create FR analogs, these native substrates were replaced with several amino acids and acyl-CoAs. The resulting side chain intermediates were then applied for FrsA_{TE}-mediated transesterification. Successful substrate acceptance and catalytic processing by both FrsA_{CAT} and FrsA_{TE} would lead to the formation of novel FR analogs (6.7.3).

The extracts of the reaction mixtures were analyzed by HPLC-MS to characterize the final products. Theoretical *m/z* values for each analog were calculated using ChemDraw. Extracted ion chromatograms (EIC) corresponding to the respective *m/z* values were integrated to quantify product abundance based on peak area (AUC). To ensure comparability of the product yields of different assay batches, the absolute abundance of each FR analog was normalized to the appropriate positive control. Positive and negative control reactions were included in each experimental set to validate assay performance. Both controls utilized native substrates (L-leucine and propionyl-CoA), whereby the enzymes for the negative control were heat-inactivated at 90 °C for 10 min prior to assay conduction.

This study was a collaborative effort between Bachelor student Georgina Kowalski and Master's student Alexander Kai Kania. Kowalski initiated the investigation by examining single substrate substitutions, while Kania subsequently explored combinatorial substrate exchange. The combined results are presented in **Figure 56**.

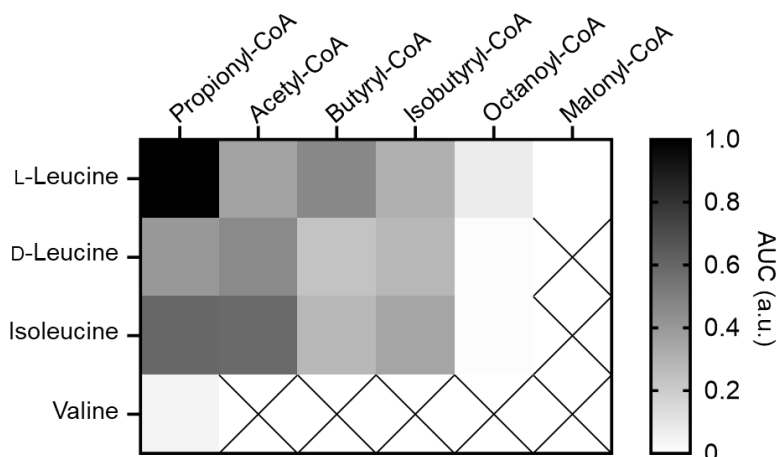


Figure 56: Heatmap evaluation of the substrate exchange experiments. Depicted is the relative abundance of each FR-analog as integrated area under the curve (AUC) of the sample peak at the m/z of the respective analog (a.u. = arbitrary unit). FR analogs were normalized *via* division by FR (established assay with L-leucine and propionyl-CoA) serving as positive control in each experiment. ($n = 3$)

Initial experiments focused on single substrate replacements (first row and column of the heatmap). It is striking that the replacement of L-leucine only led to the formation of an FR analog if the substituent was a leucine isomer, as exemplified by the absence of product formation when valine was applied. In contrast, most acyl-CoA analogs were accepted by the C domain, with malonyl-CoA being the only exception. This suggests that acyl chain length and conformation are critical determinants for substrate recognition, while the presence of a terminal carboxyl group, as in malonyl-CoA, seemed to interfere with side chain formation.

Since malonyl-CoA and valine did not show any formation of FR analogs even in the single exchange experiment, these substrates were excluded from subsequent combinatorial experiments. In the following **Figure 57**, the results from the heat map are presented in a column diagram, allowing a better quantitative evaluation. The results of the exchanges by malonyl-CoA and valine were excluded for clarity.

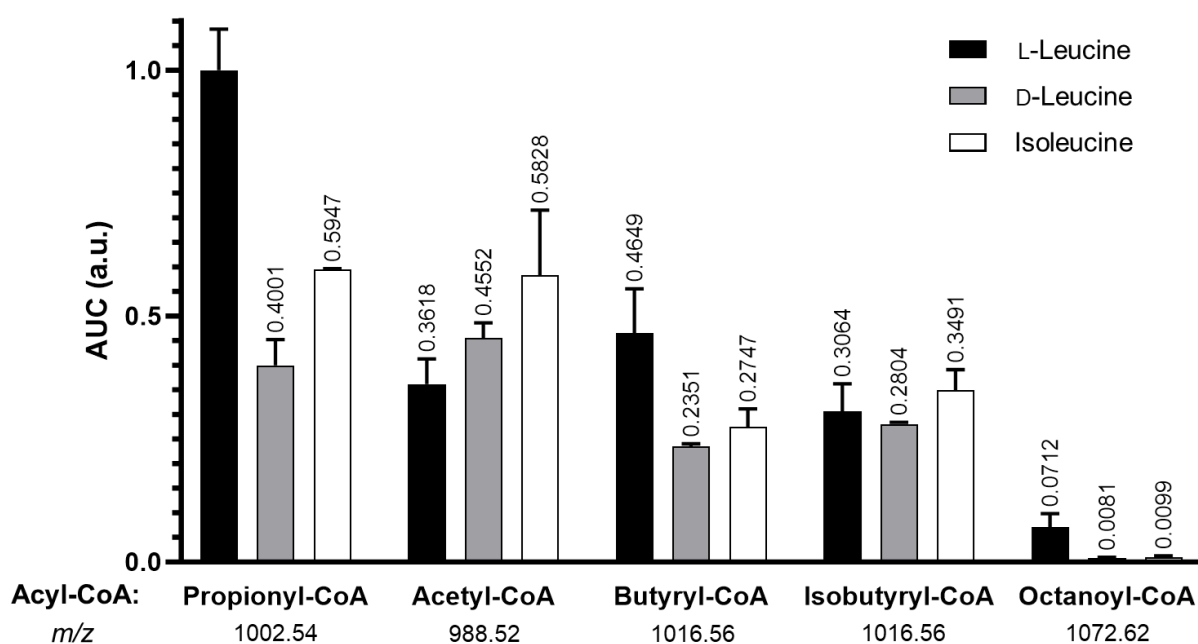


Figure 57: Column evaluation of the substrate exchange experiments. Depicted is the relative abundance of each FR-analog as integrated area under the curve (AUC) of the sample peak at the m/z of the respective analog (a.u. = arbitrary unit). FR analogs were normalized *via* division by FR (established assay with L-leucine and propionyl-CoA) serving as positive control in each experiment. The amino acid used for side chain assembly is marked by the color of the bar (L-leucine: black, D-leucine: grey, isoleucine: white). The used acyl-CoA and expected m/z of the product are annotated at the X-axis. Error bars were calculated as standard error of the mean (SEM). (n = 3)

The highest product yield was obtained with the native substrates L-leucine and propionyl-CoA, resulting in the formation of FR. Leucine isomers also yielded significant amounts of the corresponding FR analogs, with isoleucine exhibiting nearly 60% of the L-leucine product level. Using D-leucine as a substrate also achieved an abundance of 40%. In combination with propionyl-CoA, L-leucine is consequently the preferential substrate, followed by isoleucine. This pattern could also be observed in the combination of the leucine isomers with butyryl-CoA, whereby the yields were halved. However, a different trend was observed with acetyl-CoA, where isoleucine outperformed L-leucine. The most pronounced difference was observed with acetyl-CoA, where surprisingly, additionally of isoleucine yielded 0.22 arbitrary units (a.u.) more product than L-leucine. In combination with isobutyryl-CoA, all leucine isomers yielded nearly the same abundance of product. Furthermore, the abundances of the FR analogs with butyryl- and isobutyryl-CoA were all in the same range, except for the L-leucine-butyryl-CoA combination, leading to the formation of the known naturally FR analog FR5. When using octanoyl-CoA, the longest acyl chain tested, product formation was significantly reduced, likely due to steric hindrance. While a detectable amount of product was formed with L-leucine, the other amino acid isomers yielded negligible amounts of anticipated product.

Beyond the characterization of FR analogs, the MS data were analyzed for the presence of non-hydroxylated side chain intermediates. No evidence for such intermediates was found, indicating that β -hydroxylation by FrsH is a prerequisite for subsequent acylation and suggesting a broad substrate specificity of FrsH.

A consistent observation across all experiments was the presence of low-intensity signals corresponding to FR and FR2 isomers (**Figure 58**). Their occurrences persisted despite thorough experimental controls to eliminate contamination. While the absolute abundance of these contaminating peaks was minimal compared to the desired products, their presence raised concerns about data interpretation. Given the primary focus on identifying potential substrates for FR analog synthesis, the impact of these contaminations was considered negligible, especially for the distinct peaks of the newly synthesized FR analogs.

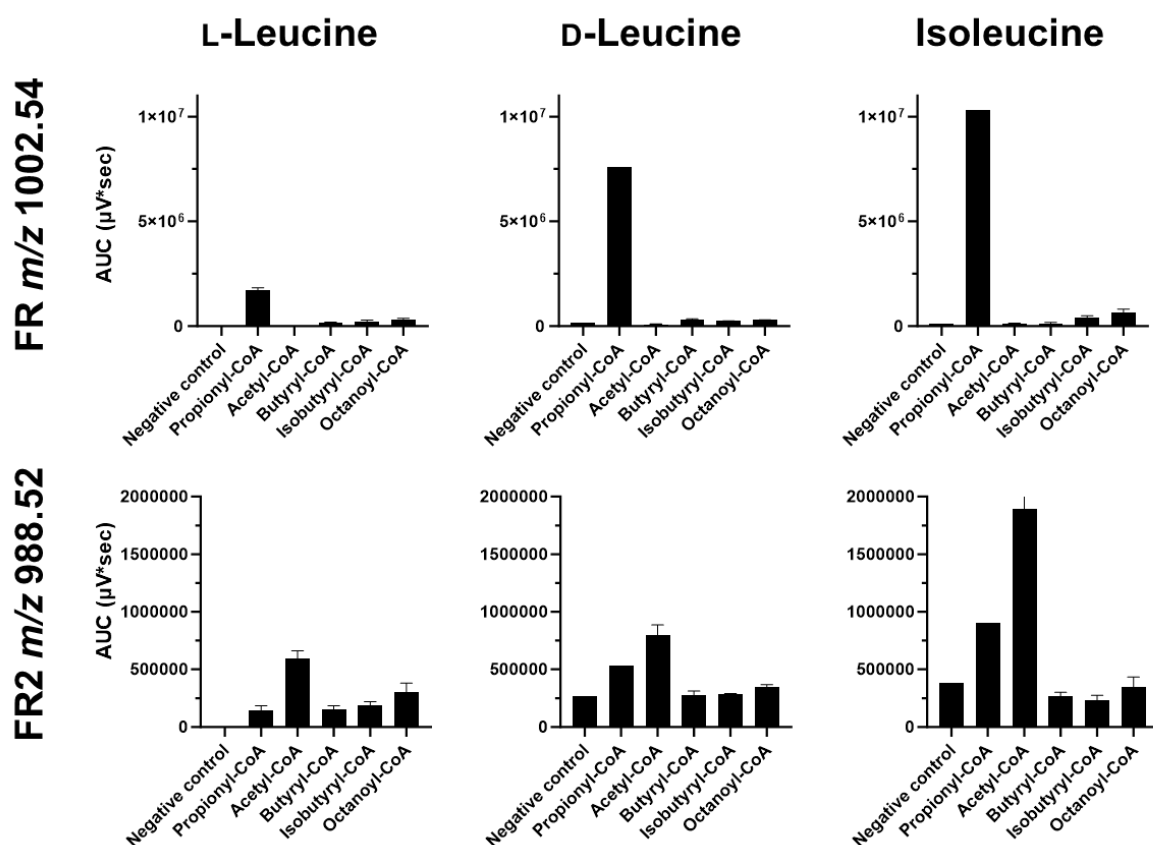


Figure 58: Evaluation of the absolute abundance of FR and FR2 isomers in FR analog samples. The abundance is measured as integrated area under the curve (AUC) of the signal peak at the respective m/z (shown on the left with characteristic FR analog) in $\mu V \cdot sec$. All propionyl-CoA combinations shown in this figure were conducted with L-leucine, to serve as positive control. The negative control was conducted with L-leucine + propionyl-CoA and heat-inactivated enzyme). FR analogues: $n = 3$. Positive and negative control: $n = 1$. Error bars are calculated as SEM.

The consistent detection of contaminant peaks at m/z 988.52 in nearly all samples, including prominent peaks corresponding to FR2 isomers in the negative control batches for D- and isoleucine combinations, presented a major challenge. The unexpected formation of these compounds in the absence of active enzymes was puzzling. Despite the inability to definitively identify the source of contamination, the clear differentiation between batches where FR2 formation was expected and those where it was not provided compelling evidence for the validity of the overall experimental design. However, the presence of these contaminant peaks did not compromise the interpretation of results for other combinations due to distinct m/z values.

To investigate the potential role of acyl-CoA stock contamination in the observed phenomenon, Alexander Kania conducted stability studies (**Figure 71**). HPLC-MS analysis of acyl-CoA stocks before and after incubation revealed low levels of acetyl-CoA contamination in butyryl-, isobutyryl-, and octanoyl-CoA stocks. However, these trace amounts were insufficient to account for the relatively high levels of FR and FR2 observed in the assay samples. Notably, propionyl-CoA, which exhibited the highest levels of contamination in the assays, was free of detectable acetyl-CoA impurities. Furthermore, the stability of all acyl-CoA stocks during the assay incubation period was confirmed by the absence of new peaks at the target m/z values.

An alternative hypothesis considered the possibility of acyl chain degradation following side chain assembly and attachment, potentially leading to the formation of FR and FR2. This explanation could account for the low abundance of FR analogs containing longer acyl chains and the concomitant increase in FR and FR2 levels.

4.4.4. Exchange of FR-Core as substrate for the TE domain

In previous chapters, the scope of side chain modifications compatible with FrsA_{TE}-mediated attachment to FR-Core was extensively investigated. However, it was also attempted to determine FrsA_{TE} substrate specificity with regard to the nucleophile, FR-Core. Given the central role of FrsA_{TE} in elucidating the mechanism of noncanonical transesterification and to validate its possible application as chemoenzymatic tool, the enzyme's tolerance for structural variations of FR-Core is worth investigating. A particularly intriguing possibility involves the utilization of linearized FR-Core (LFRC) as a substrate for the transesterification reaction.

For *in vivo* production of LFRC, a *C. vaccinii* Δ *vioA* Δ *frsA* Δ *frsG8*_{TE} strain should be generated. The deletion of *vioA* (encoding for the biosynthesis of violacein A) was intended to facilitate simplified isolation of the desired natural product, and the deletion of *frsA* served to prevent side chain assembly and attachment. The *C. vaccinii* Δ *vioA* Δ *frsA* strain was provided by Dr. René Richarz,

while the deletion of *frsG8_{TE}* was conducted by Alexander Kania within the scope of his master's thesis.

The TE domain of FrsG8 catalyzes the cyclization of the FR-Core precursor molecule and the associated release from the NRPS machinery. In theory, the deletion would therefore mean that the precursor can no longer be cyclized and remains linear. One problem that could occur is that LFRC is not released and remains bound to the FrsG8 T domain. However, if the *in vivo* production with subsequent isolation of LFRC is feasible, it should be used as substrate in the TE domain assay to investigate whether FrsA_{TE} is able to attach the side chain to the linearized nucleophile.

4.4.4.1. Construction of FrsG8_{TE} knock-out mutant

A strategy used for the investigation of FK228 (Romidepsin) biosynthesis,^[221] was adapted for the construction of *C. vaccinii* knock-out mutants by Dr. René Richarz (6.5.11). Therefore, the FRT-cassette, deriving from pPS858, a plasmid developed for generating *Pseudomonas* deletion mutants^[222], was inserted *via* Gibson assembly between the up- and downstream regions of *frsG8_{TE}* into the pEX18Tc vector. The resulting plasmid was then transformed into *E. coli* NEB10 β , used as a helper strain for triparental conjugation. The other components of this were the *E. coli* ET12567 pUB307 helper strain and *C. vaccinii* Δ *vioA* Δ *frsA*. A successful conjugation should result in the replacement of the *frsG8_{TE}* gene by the FRT-cassette in the *C. vaccinii* strain. After selection, the positive colonies of *C. vaccinii* Δ *vioA* Δ *frsA* Δ *frsG8_{TE}:FRT* were distinguishable by fluorescence. Promising colonies were used for amplification by Taq colony PCR, to analyze if the FRT-cassette was inserted at the right position in the genome. The applied primers targeted the up- and downstream regions of *frsG8_{TE}*.

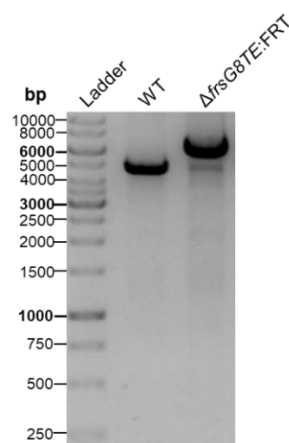


Figure 59: Agarose gel electrophoresis of the control PCR to verify the construction of the *C. vaccinii* Δ *vioA* Δ *frsA* Δ *frsG8_{TE}:FRT* mutant. Oligonucleotides amplified from the *frsG8_{TE}* gene region and separated on a 1% agarose gel at 120V. WT = gDNA of *C. vaccinii*. Predicted DNA fragment length: WT = 4,494 bp, Δ *frsG8_{TE}:FRT* = 5612 bp.

As the inserted FRT-cassette comprises more bp than the replaced *frsG8_{TE}* gene, the PCR products should show a size difference of approx. 1 kbp. The agarose gel electrophoresis of the PCR products revealed that the replacement of the gene of interest (GOI) was successful. The *C. vaccinii* wildtype showed a band right below the height of the 5000 bp ladder band, corresponding to the expected band of 4494 bp. The analysis of the PCR product of *C. vaccinii* Δ *vioA* Δ *frsA* Δ *frsG8_{TE}:FRT* resulted in a band at the height of the 6000 bp ladder band, which could be distinctly assigned to the desired FRT-cassette inserted into the *C. vaccinii* genome at the correct position (**Figure 59**).

After this conformation of the successfully conducted triparental conjugation, the FRT-cassette should be removed from the genome to avoid polar effects on downstream genes. This should take place by Flp-mediated site-specific recombination within the FRT-cassette, resulting in an 86 bp scar, where the *frsG8_{TE}* gene used to be. Therefore, a biparental conjugation with the *E. coli* DH10 β pTA-Mob + pMLBAD_{tet::Flp}/*SacB* strain was performed. However, even after several attempts, it was not possible to remove the FRT-cassette due to issues with conjugation and the subsequent selection. Thus, the *C. vaccinii* Δ *vioA* Δ *frsA* Δ *frsG8_{TE}:FRT* was used for fermentation approaches to investigate the production of linearized FR-Core.

4.4.4.2. Exchange of FR-Core by linearized FR-Core

To identify the final product of the FR biosynthesis by *C. vaccinii* Δ *vioA* Δ *frsA* Δ *frsG8_{TE}:FRT*, the knock-out strain was fermented in 50 mL LB cultures over 48 h. Since the TE domain of FrsG8 releases the linear precursor molecule of FR-Core from the NRPS machinery by cyclization, it was hypothesized, that the precursor would remain bound to the T domain in the absence of the TE domain. To address this, an additional alkaline hydrolysis step was incorporated into the extraction process. Following fermentation, cultures were subjected to two rounds of *n*-butanol extraction. The organic phase from the initial extraction was evaporated to yield the supernatant sample, while the aqueous phase underwent alkaline hydrolysis to release enzyme-bound peptides. Subsequent *n*-butanol extraction and evaporation of the organic phase produced the pellet sample. The same procedure was applied to the control strain, *C. vaccinii* Δ *vioA* Δ *frsA* (**6.8.2**).

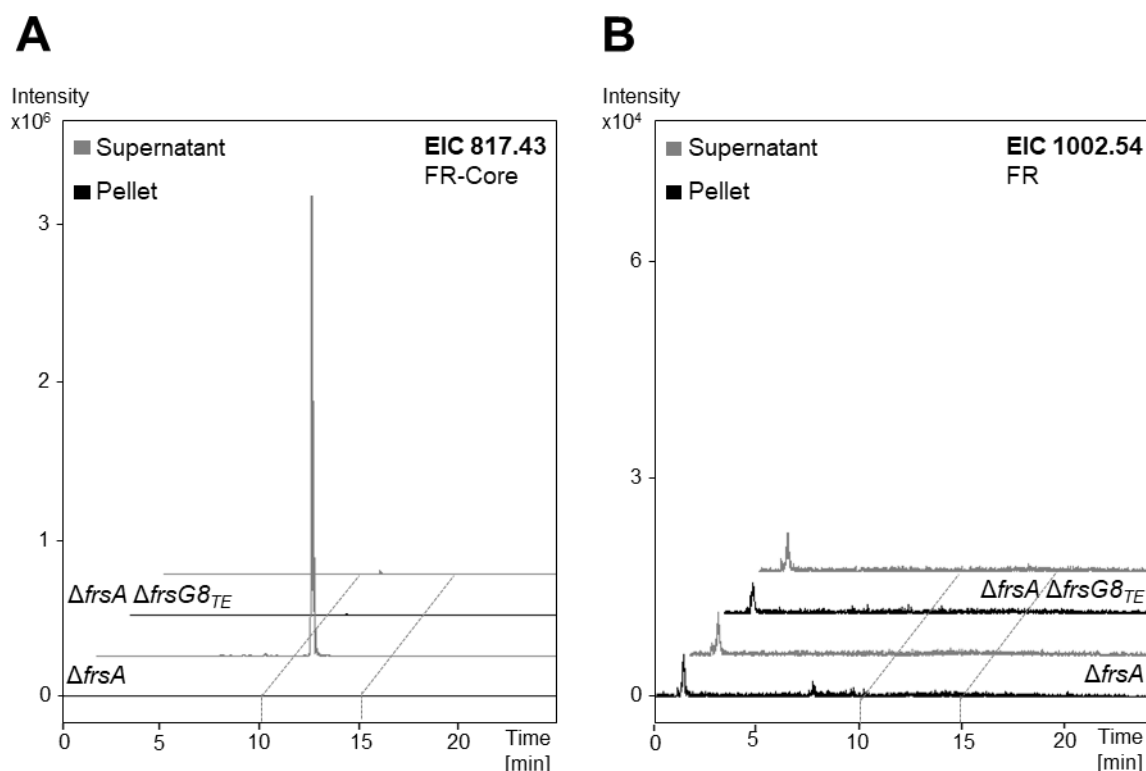


Figure 60: Extracted ion chromatograms for **(A)** FR-Core (m/z 817.43) and **(B)** FR (m/z 1002.54) of the HPLC-MS measurement of the butanolic extracts of *C. vaccinii* $\Delta vioA \Delta frsA$ ($\Delta frsA$) and *C. vaccinii* $\Delta vioA \Delta frsA \Delta frsG8_{TE}:FRT$ ($\Delta frsA \Delta frsG8_{TE}$). The supernatant chromatograms are depicted in grey and the pellet chromatograms in black.

HPLC-MS analysis of the extracted samples, focusing on the m/z values of the proton adducts of FR (1002.54; **Figure 60A**) and FR-Core (817.43; **Figure 60B**), revealed a distinct FR-Core peak in the supernatant of the *frsA*-deficient strain, consistent with previous findings by Dr. Hermes.^[21] This confirmed the efficiency of the initial extraction to isolate the vast majority of FR-Core. Notably, the FR-Core peak was absent in the $\Delta frsG8_{TE}:FRT$ mutant, indicating the abolishment of cyclization of the FR-Core precursor. Furthermore, the absence of an FR peak in both samples confirmed the complete inhibition of FR biosynthesis. Since it could thus be demonstrated that FR biosynthesis could be interrupted at the desired state, the new final product of the modulated biosynthesis was to be examined. Therefore, the HPLC-MS data were evaluated by searching for the m/z corresponding to the proton adduct of linearized FR-Core (835.43), which was calculated using ChemDraw. The extracted ion chromatograms are depicted in **Figure 61**.

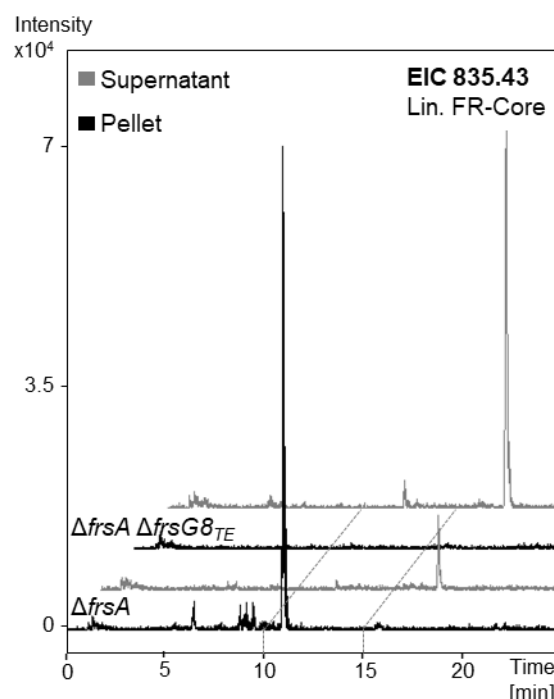


Figure 61: Extracted ion chromatograms for linearized FR-Core (m/z 835.43) of the HPLC-MS measurement of the butanolic extracts of *C. vaccinii* Δ *vioA* Δ *frsA* (Δ *frsA*) and *C. vaccinii* Δ *vioA* Δ *frsA* Δ *frsG8_{TE}*:FRT (Δ *frsA* Δ *frsG8_{TE}*). The supernatant chromatograms are depicted in grey and the pellet chromatograms in black.

The HPLC-MS analysis of the supernatant and the pellet sample of *C. vaccinii* Δ *vioA* Δ *frsA* revealed distinct peaks corresponding to the predicted m/z of LFRC. Notably, these peaks exhibited different retention times. The pellet sample peak, eluting at 11 minutes, aligned with the retention time of FR-Core, suggesting that this peak derived from the FR-Core ammonium adduct $[M+NH_4]^+$ that has a calculated mass of 834.46. Conversely, the peak in the pellet sample eluted at 17 minutes, likely due to the increased polar surface area of LFRC resulting in stronger interactions with the stationary phase during chromatography. Thus, it could be concluded that the peak with a retention time of 17 min may correspond to LFRC.

The presence of LFRC in the *C. vaccinii* Δ *vioA* Δ *frsA* pellet sample was hypothesized to be attributed to residual LFRC bound to the FrsG8 T domain. Through the alkaline hydrolysis with KOH, the stalled LFRC could have been detached from the T domain and might have led to a small amount of LFRC even in the *C. vaccinii* strain harboring a functional FrsG8_{TE}.

In contrast, the *C. vaccinii* Δ *vioA* Δ *frsA* Δ *frsG8_{TE}*:FRT mutant exhibited a significantly more intense potential LFRC peak at 17 minutes, confirming the successful disruption of the cyclization. However, the overall yield of this remained low, potentially due to the presence of the FRT cassette or the aggravated extraction by alkaline hydrolysis, what could lead to peptide degradation.

To address the low LFRC yield, a preparative-scale fermentation of about 6 L of the mutant strain was conducted, followed by flash chromatography purification (6.8.2). The gained amount of a fraction containing potential LFRC (2 mg) was insufficient for further purification *via* HPLC and was directly used as substrate for the TE domain assay.

To optimize side chain delivery for the TE domain assay, two approaches were performed: 1) a standard side chain assembly assay (6.7.2), and 2) direct addition of the side chain mimic, Pp-Hle-SNAC (6.7.2.1). Both assays were conducted in parallel under identical conditions. Given the challenges in quantifying side chain production in the assembly assay, the side chain mimic approach provided a more controlled method for evaluating LFRC substrate specificity. The assays were analyzed by HPLC-MS, whereby it was searched for the m/z of FR-Core (817.43), LFRC (835.45), FR (1002.54), and linearized FR (LFR, 1020.55).

Additionally, the TE domain assay with the side chain mimic was conducted twice, with different LFRC concentrations, due to the fact, that it was not possible to exactly determine the LFRC concentration of fraction 2. For the TE domain assay performed using the preceding side chain assembly assay, the entire remainder of fraction 2 was used to avoid insufficient amount of LFRC. For each case, a positive control and a negative control were included. The positive control was conducted with the native enzymes and substrates, as well as in the negative control, in which, by contrast, each involved enzyme was heat-inactivated.

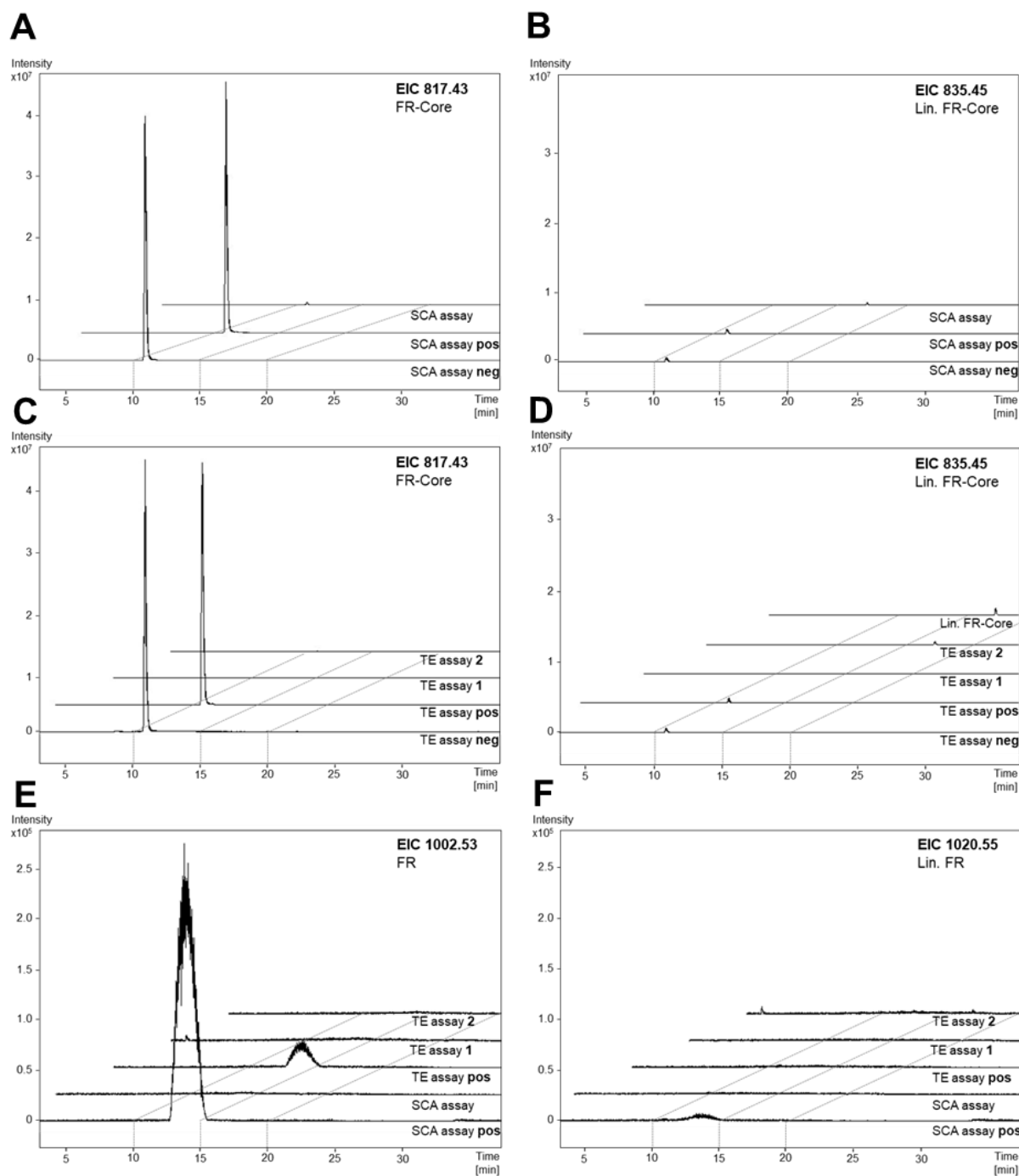


Figure 62: (A) Extracted ion chromatograms for FR-Core (m/z 817.43) of the HPLC-MS measurement of the TE domain assay with preceding sidechain assembly assay (SCA assay). (B) Extracted ion chromatograms for linearized FR-Core (m/z 835.45) of the HPLC-MS measurement of the TE domain assay with preceding sidechain assembly assay (SCA assay). (C) Extracted ion chromatograms for FR-Core (m/z 817.43) of the HPLC-MS measurement of the TE domain assay conducted with side chain mimic (TE assay). (D) Extracted ion chromatograms for linearized FR-Core (m/z 835.45) of the HPLC-MS measurement of the TE domain assay conducted with side chain mimic (TE assay). (E) Extracted ion chromatograms for FR (m/z 1002.53) of the HPLC-MS measurement of both assay approaches (TE domain assay with preceding sidechain assembly assay (SCA assay)/ TE domain assay conducted with side chain mimic). (F) Extracted ion chromatograms for linearized FR (m/z 1020.55) of the HPLC-MS measurement of both assay approaches (TE domain assay with preceding sidechain assembly assay (SCA assay)/ TE domain assay conducted with side chain mimic (TE assay)). Applied linearized FR-Core concentration: TE assay 1 = $< 50 \mu\text{M}$, TE assay 2/SCA assay = $\leq 100 \mu\text{M}$. TE Controls: pos = conducted with the native substrate FR-Core, neg = conducted with the native substrate FR-Core and heat-inactivated proteins.

To assess the presence of substrates and reaction products, extracted ion chromatograms (EICs) were generated for FR-Core (m/z 817.43) and LFRC (m/z 835.45) (**Figure 62A-D**). As expected, both positive and negative control samples exhibited a prominent FR-Core peak at the expected retention time. The absence of this peak in the assay samples indicated the absence of FR-Core contamination and precluded the unintended formation of FR or LFR.

The extracted ion chromatograms for the m/z of LFRC (835.43) revealed significantly lower LFRC levels than the applied amount of FR-Core. The corresponding peak with a retention time of 17 min was only detectable in the SCA assay and the TE assay 2 sample, in which the concentration of potential LFRC was deliberately adjusted considerably higher. A minor peak at 11 minutes, corresponding to the retention time of FR-Core, likely represented an ammonium adduct of FR-Core. **Figure 62D** shows a comparative HPLC-MS analysis of fraction 2 to confirm the presence of the LFRC peak.

EICs generated for the m/z of FR (**Figure 62E**) demonstrated successful assay execution, as indicated by the presence of a clear FR peak at 14 minutes in both positive controls. Notably, the positive control using the side chain assembly assay exhibited a higher FR yield compared to the TE domain assay with the side chain mimic, suggesting that the combined approach might be more efficient for substrate conversion. The absence of detectable FR peaks in assay samples was expected due to the lack of FR-Core.

The positive control of the side chain assay also showed a peak in the EIC of an m/z of 835.43 (**Figure 62F**), despite, the retention time of 14 min, indicating the assignment to FR with an ammonium adduct.

While no definitive LFR peak was observed in the assay samples, a slight increase in signal intensity at the expected retention time (17 minutes) in the TE domain assay with the side chain mimic could indicate low-level LFR formation. However, conclusive evidence for LFR production would require further analysis, such as HPLC-MS/MS.

To enhance LFRC production and facilitate downstream analysis, strategies such as increasing fermentation scale and optimizing purification protocols should be conducted. Furthermore, the insertion of an overexpression promotor and the removal of the FRT cassette could be used to increase the LFRC yield.^[24]

4.5. The Modifying Enzyme FrsH

Adenylation Domain-Guided Recruitment of Trans-Acting Nonheme Monooxygenases in Nonribosomal Peptide Biosynthesis

Daniel A. Wirtz, Niels Schneberger, **Sophie Klöppel**, René Richarz, Matthias Geyer, Gabriele M. König, Gregor Hagelüken, and Max Crüsemann

ACS Chem. Biol. 2023, 18/8, 1748–1759. doi:10.1021/acscchembio.3c00106

Introduction

Trans-acting β -hydroxylases of NRPS systems are known for introducing a functional group to the growing peptide chain. This free hydroxy group can itself contribute to the biological activity or provide an expanded chemical space on a natively inert amino acid, enabling further modifications (e.g. macrocyclization), which in turn enhance biological diversity and activity of the resulting natural products.^[21,163,223–226]

Trans-acting β -hydroxylases belong to different enzyme families like cytochrome P450-dependent monooxygenases (CYP450)^[152,153], FE(II)/ α -ketoglutarate-dependent dioxygenases^[227,228], and nonheme diiron monooxygenases (NHDMs)^[229]. Apart from the fact, that NHDMs were barely characterized to date, it should be noted that the molecular mechanisms of substrate recognition by *trans*-acting β -hydroxylases in general are scarcely understood. However, these enzymes hold great potential for *de novo* NRPS engineering.^[155]

The NHDM FrsH plays an important role in the nonribosomal peptide synthetase (NRPS) assembly line of FR900359 (FR) biosynthesis.^[21,163] FR is a potent and selective inhibitor of Gq proteins.^[54] Its strong target affinity can be attributed to the sidechain (2*S*,3*R*)-(N-propionyl)-3-hydroxyleucine (N-Pp-Hle), acting like an anchor in the FR Gq binding pocket. The side chain is assembled by a multi-enzymatic biosynthesis by the FrsA NRPS module, consisting of condensation (C), adenylation (A), thiolation (T), and thioesterase (TE) domain, the MbtH-like protein FrsB, and FrsH. The latter β -hydroxylates the thiolation-domain-bound L-leucine residues, whereby a specific interaction between FrsA_A and FrsH is assumed.

The aim of the study was to characterize the *trans*-acting NHDM FrsH structurally and biochemically and to provide insights into the substrate recognition mechanism associated with NRPS systems.

Summary

The crystallization attempts of the heterologous expressed and purified FrsH led to the formation of a single crystal, from which a complete 2.57 Å diffraction data set could be collected with which the structure of FrsH was solved. The structure showed a well-defined electron density of two intertwined FrsH monomers and could be modeled from residue 4-531. FrsH consists of two domains: the N-terminal domain (AA 4-200) forms a lid-like structure that covers the C-terminal domain (AA201-530), possessing an $\alpha\beta\alpha$ -sandwich structure and harboring the active site formed by a diiron cluster. The interaction of residues with the two metal ions could be identified (Fe1 – H306, H304, E376; Fe2 – D308, H309, D402). The length of the localized substrate tunnel correlates with the phosphopantetheinyl moiety of the T domain, carrying L-leucine. A higher electron density could be observed in the substrate channel between V492 and W437, however, bound molecules and the residues playing a role in substrate interaction could not be identified.

To investigate the interaction of FrsH and FrsA, analytical size exclusion chromatography in combination with multiangle light scattering (SEC-MALS), dynamic light scattering (DLS) experiments, thermal nano differential scanning fluorimetry (nanoDSF) experiments, and isothermal titration calorimetry (ITC) measurements were conducted. The results of each experiment revealed a clear dimeric character of FrsH and also a weak interaction of FrsH and FrsA, with a minor thermal stabilizing effect of FrsA.

Bioinformatical modeling using AlphaFold2 predicted the interaction interface between FrsA and FrsH. Additionally, several surface-exposed hydrophobic residues, as FrsA W573 and M555 of FrsA and FrsH S52 and M65 were identified. For the experimental validation of the interaction interface, mutations of the residues of interest on both enzymes were prepared and tested in an *in vitro* side chain assembly assay. Disruptions in the interaction interface reduce the affinity and slow down the hydroxylation process, which was quantitatively measurable *via* product formation. The FrsA mutations W573A and M555S and the FrsH mutations S52R and M65R showed a significant impairment of the enzyme affinity. Combinations of the mutants led to the formation of hydroxylated product of only 20%. Based on these results, the mutations showing the highest impact on complex formation were tested in different combinations in DLS experiments. Significant effects of the interface mutations on the complexation ability could be observed. Alignments of cognate and noncognate A domains show that these residues are not conserved within NHDM-recruiting A domains, leading to the assumption that the whole interaction interface plays an important role in protein-protein interaction. Nevertheless, the experimental data confirmed the crucial role of A domains for the recognition of NHDMs. Thus, the definite

mechanism still awaits elucidation, harboring great potential for targeted engineering of NRPS systems.

In vivo substitution experiments of FrsH were conducted with the NHDMs LybC and HynC from the biosynthetic pathways of the NRPS antibiotics lysobactin and hypeptin. Employing the *Chromobacterium vaccinii* Δ *frsH* deletion mutant and the newly developed *C. vaccinii* complementation system, it could show that HynC and LybC are able to replace FrsH in FR biosynthesis. This suggests that the members of the NHDM family share a compatible recognition mechanism and are thus exchangeable between the NRPS systems.

Author's contribution

The author of this thesis, cloned, expressed, and purified the FrsA mutants M555S, W573A, D585A, R597A, and F635A. The plasmids for the FrsH mutants S52R, M65R, F70R, and R449E were cloned by Niels Schneberger and transformed, expressed, and purified by the author. The experimental validation of the complex interface was carried out by the author, who also created the illustration (Figure 6B) and was involved in the preparation of this part of the manuscript.

Furthermore, the author provided the heterologous expressed and purified enzymes FrsH and FrsA, as well as their mutants for the SEC-MALS, DLS, and nanoDSF experiments, conducted by Niels Schneberger.

4.6. The Modifying Enzyme FrsC

A Specialized Dehydrogenase Provides L-Phenyllactate for FR900359 Biosynthesis

Sophie Klöppel, René Richarz, Daniel A. Wirtz, Natalia Vasenda, Gabriele M. König, and Max Crüsemann

ChemBioChem 2022, 23/10. doi: 10.1002/cbic.202100569

Introduction

Phenyllactate (PLA; 2-Hydroxy-3-phenylpropanoic acid) is a naturally occurring organic acid, which contains a chiral carbon atom, resulting in two enantiomers: L- and D-PLA.^[51,230] The latter is a nonproteinogenic building block of the cyclic depsipeptide FR900359 and of utmost importance for its strong and selective Gq inhibitory activity. For example, the replacement of D-PLA by phenylalanine in the structurally closely related YM-254890 derivatives revealed a 25 up to 107-fold loss of Gq inhibition activity, highlighting its structural relevance.^[156,231]

Numerous D-PLA-containing cyclic depsipeptides have already been described, lacking an explanation of the origin of this special building block.^[232–239] The main producers are lactic acid bacteria (LAB)^[160,240,241], however, also non-LAB^[230,242] with the ability to produce PLA were characterized recently. PLA is a by-product of the phenylalanine metabolism, since phenylalanine undergoes a transaminase-catalyzed conversion to phenylpyruvic acid (PPA) to avoid intercellular accumulation. PPA is then further reduced to the PLA enantiomers by L-lactate and D-lactate dehydrogenases (L-LDH, EC 1.1.1.27, D-LDH, EC 1.1.1.28).^[160,161]

Bioinformatic analysis of the *frs* biosynthetic gene cluster classified FrsC as NAD(P)H-dependent dehydrogenase, as it shows high sequence similarities to malate dehydrogenases (MDH; EC 1.1.1.37) and lactate dehydrogenases (LDH; 1.1.1.27/28).^[15] This enzyme classes catalyze the NAD(P)⁺/NAD(P)H-dependent oxidoreduction of 2-keto and 2-hydroxy carboxylic acids.

This led to the assumption, that FrsC reduces PPA to L-PLA, which is then incorporated and on-line modified by FrsE3, the third module of the FR NRPS system. In this paper, this hypothesis should be validated by an in-depth biochemical investigation of FrsC. Furthermore, the relation of FrsC to dehydrogenases from primary metabolism was bioinformatically analyzed to provide insights into the evolution of the enzyme, enabling the acceptance of the aromatic PPA with high specificity.

Summary

For the *in vitro* investigation of the putative PPA reductase, the *frsC* gene from *Chromobacterium vaccinii* was cloned into *E. coli*. FrsC was heterologously expressed and purified *via* affinity chromatography. The isolated enzyme was then used in NAD(P)H-dependent activity assays. As NAD(P)H is the cofactor for the reaction catalyzed by FrsC, the consumption could be measured photometrically, allowing a profound examination of the enzymatic activity. Thus, it could be determined that NADPH is the preferred cofactor, and the optimal *in vitro* reaction parameters are at 30 °C and a pH value of 6. Based on these optimal assay conditions, the substrate specificity was analyzed, revealing that FrsC nearly exclusively accepts PPA as substrate. Only hydroxy-phenylpyruvic acid was also accepted to a minor extent, indicating a high substrate specificity with low tolerance towards alterations at the aromatic moiety. Through the variation of substrate concentration of 0.1 up to 7 mM under optimal conditions, the kinetic parameters could be defined to a K_M of 1.37 ± 0.41 mM, a V_{max} of 15.93 ± 1.34 $\mu\text{mol min}^{-1} \text{mg}^{-1}$, and a k_{cat} of 9.40 ± 0.79 s^{-1} by applying the Michaelis-Menten equation and a Hanes-Woolf plot.

To identify the configuration of the FrsC reaction product, the *in vitro* assay was conducted under optimal conditions and measured using HPLC with (2-hydroxypropyl)- β -cyclodextrin as chiral selector added to the mobile phase. Through the measurement of L- and D-PLA as standards, the retention time of the product peak could be assigned to L-PLA.

To validate this result, it was subsequently investigated, which PLA configuration is mainly accepted by the A domain FrsE3. The performed mass spectrometry-based γ - $^{18}\text{O}_4$ -ATP exchange assay^[146] with the heterologously expressed and purified FrsE3 showed a rather low substrate specificity. However, L-PLA (52%) achieved the highest exchange of ATP, followed by L-phenylalanine (48%) and D-PLA (42%). Since the structurally similar substrates DL-*p*-OH-phenyllactate (OH-PLA) and *p*-fluoro-D-phenyllactate (F-PLA), were not accepted as substrates, it can be assumed that structural modifications on the phenylring negatively affect the A domain substrate binding.

In vivo experiments conducted with a *C. vaccinii frsC* deletion mutant^[21] should lead to the complementation and restoration of FR production by feeding of D- and L-PLA. The HPLC-MS measurements of the butanolic extracts of the fermentation cultures revealed that the FR production could be restored to $3.6 \pm 0.14\%$ by the addition of L-PLA, whereas feeding of D-PLA to the mutant only yielded $0.6 \pm 0.09\%$ FR production compared to the WT.

The obtained data confirmed the hypothesis that FrsC reduces PPA to L-PLA which is then activated by the FrsE3 A domain, incorporated into the peptide chain, and finally epimerized on-line to D-PLA by the FrsE3 E domain.

The phylogenetic tree constructed with FrsC homologues (72% identity) and 903 other L-LDH/L-MDH sequences, revealed that FrsC clusters with cytosolic-like MDHs. Furthermore, the high specificity for aromatic substrates hints at the evolution of a former MDH, enabling a new function. The comparatively low catalytic activity and high substrate specificity of FrsC support the theory, that it evolved from a primary to a specialized metabolism enzyme.^[243] The alignment of FrsC with selected MDHs and LDHs highlighted an unusual structure of the catalytic loop.^[15] It possesses an insertion of five residues and has a relatively high number of alanine residues, making the acceptance of bulkier substrates plausible. The 'specificity residue' at position 102 is in the case of FrsC a serine, unlike in the other LDHs and MDHs.^[244,245] If this serine plays a role in substrate recognition must be investigated by further experimental analysis like mutational studies.

Author's contribution

All strains used in the paper were provided by Natalia Vasenda and René Richarz. The author of this thesis performed and evaluated all experimental approaches. Excluding the bioinformatical section, which was contributed by Dr. René Richarz, the author created all tables and figures and wrote the manuscript in cooperation with Dr. Max Crüsemann.

5. Conclusion and Outlook

FR900359 is a potent and selective inhibitor of Gq proteins, emphasizing its value as a pharmacological tool and potential therapeutic agent. To date, two bacterial producers of FR have been identified: the leaf endosymbiont "*Candidatus Burkholderia crenata*" and the soil bacterium *Chromobacterium vaccinii*. Remarkably, both organisms harbor architecturally identical *frs* BGCs encoding an octamodular NRPS machinery. The monomodular FrsA is responsible for synthesizing the side chain, a critical determinant of FR's inhibitory activity. Notably, this module possesses a noncanonical TE domain capable of catalyzing intermolecular transesterification, yet the underlying mechanism remains elusive. Given the pivotal role of the FrsA TE domain-catalyzed reaction and the distinctive nature of its function, this study focused on elucidating the transesterification mechanism and its potential applications.

5.1. Construct generation of FrsA_{TE}

The primary objective of this study was the structure elucidation of FrsA_{TE} to gain detailed insights into its noncanonical function. Initial crystallization trials of FrsA_{TE} performed by Dr. Hermes yielded crystals, however, the diffraction quality proved insufficient for high-resolution structure determination (4.1). Given the limitations of crystallization optimization, strategies for construct development were pursued, since it was hypothesized that flexible regions within the protein structure hindered crystal formation.

To identify suitable constructs, several approaches were explored. Controlled proteolysis was used to remove flexible, protease-accessible regions (4.1.1.1). Additionally, bioinformatic analysis, utilizing tools like antiSMASH, I-TASSER, and AlphaFold2.0, was conducted to identify conserved structural elements and potential truncation sites (4.1.2). Furthermore, the crystallization of the didomain construct FrsA_{T+TE} was attempted, leveraging the stabilizing effects of domain-domain interactions (4.1.3).

Although controlled proteolysis generated the new construct FrsA_{TE}Ed, the experimental effort was considered disproportionate to the obtained results (4.1.1.1). In most approaches, the input POI was digested into small, non-crystallizable peptides. At this point, it should be emphasized that the addition of the substrate FR-Core led to a reduced proteolytic digest of the POI, indicating a stabilizing effect. The identification of promising products was hardly feasible using SDS-PAGE analysis and subsequent sequence coverage determination by HPLC-MS/MS measurements. The FrsA_{TE}Ed could be gained through the digest with elastase (1:1000 (protease:protein), +FR-Core, 37 °C for 2 h), and revealed the removal of the 27 amino acids at the N-terminus (including the

His-tag). After successful cloning, expression (*E. coli* LOBSTR pHis8-TEV-FrsA_{TE}Ed), and purification (4.1.1.2), the new construct failed to produce diffracting crystals (4.1.1.3).

Bioinformatics-guided optimization proved to be a more efficient approach. Structural predictions using I-TASSER and comparison with FrsG8_{TE} revealed flexible regions at the FrsA_{TE} termini, and an incorrect annotation of the *frsA_{TE}* domain was identified (4.1.2.1). The linker region between the T and TE domain of FrsA, which seemed to be highly flexible, was considered to be a significant impediment to crystallization. Bioinformatic analysis offers a targeted approach to construct identification, however, its limitations lie in accurately assessing the impact of distal structural elements. The constructs FrsA_{TE}short and FrsA_{TE}long, generated through bioinformatic analysis, while promising, both yielded insufficient amounts for crystallization approaches (*E. coli* LOBSTR pHis8-TEV-FrsA_{TE}short/long; 4.1.2.3). However, the obtained quantities were sufficient to examine their activity using an *in vitro* assay, thus confirming that the functional core of the proteins was still complete and active (4.1.2.4). The truncated linker, while likely dispensable for protein function, significantly contributed to protein solubility. To address this, optimization of expression and purification was undertaken using fusion proteins (4.1.2.5).

To improve protein yield and solubility for crystallization, the fusion proteins MBP (4.1.2.5.1) and mVenus (4.1.2.5.2) were employed. MBP functions as a molecular chaperone, enhancing the solubility of the POI. Conversely, mVenus, a fluorescent protein, can be utilized to optimize expression conditions, as its fluorescence enables rapid screening of various expression parameters. Both proteins were inserted into the expression vector pHis8-TEV-Linker²-POI.

Intriguingly, the results highlighted the critical role of the dual linker sequence 'GGGSGGGS' between the TEV protease cleavage site and the POI. This linker facilitated efficient cleavage by the TEV protease, enabling the release of the target protein suitable for crystallization. In contrast, constructs with a single or absent linker sequence exhibited significantly reduced or absent cleavage, as evidenced by SDS-PAGE analysis.

While MBP significantly increased protein yield, the subsequent removal of the fusion tag through TEV protease cleavage resulted in high protein loss, impeding downstream crystallization experiments. Although mVenus demonstrated potential for optimizing expression conditions, its utility in this study was however limited due to the concurrent structure elucidation of the FrsA_{T+TE} construct (see below). Nevertheless, mVenus remains a valuable tool for future protein expression optimization endeavors, particularly when coupled with appropriate SDS-PAGE analysis using ladders specific for green fluorescent proteins.

These findings underscore the importance of meticulous protein engineering for successful crystallization. The optimization of linker sequences and the judicious choice of fusion tags are crucial factors influencing protein expression, solubility, and purification.

The didomain construct FrsA_{T+TE} was cloned into the pHis8TEV vector and expressed using the *E. coli* LOBSTR strain, yielding suitable amounts for downstream processing. After minor optimizations, purification using Ni-NTA affinity chromatography, subsequent reverse affinity chromatography after His-tag removal (40-50% loss), and SEC-FPLC (20-30% loss) led to a high-quality protein sample (4.1.3.3), forming numerous crystals (after 12 hours in the condition containing 20% (w/v) PEG 3350, 0.2 M Na₂SO₄ and 0.1 M BIS-Tris at pH 7.5). The diffraction was sufficient to record data sets, enabling structural elucidation of the *apo*-form of FrsA_{T+TE} (4.1.3.4). This is likely attributable to the inherent stability provided by domain-domain interactions, a common feature of multidomain proteins like NRPS. The interaction between the T and TE domains, crucial for substrate transfer, likely stabilized the overall structure, which facilitated crystallization. The smaller size of the didomain construct did also not exceed the entropic barrier to crystal formation.

In conclusion, a combination of bioinformatics, protein engineering, and crystallization optimization was essential for obtaining a high-resolution structure of the target protein. The structural information gained provides a foundation for further functional and mechanistic studies of FrsA_{TE}.

5.2. Structural elucidation of FrsA_{T+TE}

The FrsA_{T+TE} crystals diffracted with a resolution of 2.1 Å, leading to the elucidation of the protein structure. While the co-factor FR-Core was essential for crystal formation, it was absent from the final structure, resulting in an *apo*-form representation (4.1.3.5).

The T domain exhibited a canonical four-helix topology with a distinctive loop region. The active site serine residue (S35) at the N-terminus of α 2 was identified, which is essential for carrying the phosphopantetheine moiety, substrate binding, and transport throughout the NRPS machinery. However, the Ppant moiety could not be detected in the FrsA_{T+TE} structure, which could be due to the omission of incubation with a corresponding PPTase.

The TE domain displayed the characteristic α/β hydrolase fold, including seven parallel β -sheets, surrounded by five α -helices, and the lid structure formed by three α -helices between β 5 and β 6. The catalytic triad is composed of S170, D197, and H300 positioned within the active site. The

structural arrangement of these residues is highly conserved and consistent with the catalytic mechanism of TE domains.

Further structural comparison of FrsA_{T+TE} with ten other NRPS thioesterase domains revealed a high degree of structural conservation in the domain's core structure. However, the lid region exhibited significant variability, suggesting a role in substrate specificity and catalytic mechanism.

It is known that the lid structures show high mobility and an interaction between the substrate as well as the T domain can be considered. Since only the *apo*-form could be investigated, this hypothesis could not be confirmed. It was assumed that FrsA_{TE} crystallized in the 'closed' conformation, where the lid acts as solvent shield, and the active site is barely accessible. Docking studies suggested a potential substrate binding site, but the limitations of this approach, including the static nature of the protein structure, prevented definitive conclusions.

Comparative analysis with other substrate-bound thioesterase domains, particularly NocB_{TE} and Vlm2_{TE}, revealed that substrates occupy a deep position within the active site, proximal to the catalytic serine. This observation supports the hypothesis that FR-Core, acting as the nucleophile in the FrsA_{TE}-catalyzed reaction, likely binds within the active center. However, the unique nature of the catalyzed intermolecular transesterification and the absence of suitable nucleophiles in the analyzed *holo*-protein structures hindered direct comparison and mechanistic insights.

Structural alignment analysis of FrsA_{T+TE} with other structurally characterized T-TE domain complexes, including EntF_{T+TE} and SulM_{TE}, provided insights into potential protein-protein and Ppant-bound substrate interaction. The T domain interaction interface was mapped to specific residues on both the T and TE domains. The Ppant arm of the EntF_{T+TE} was modeled into the FrsA_{T+TE} structure, suggesting its potential binding mode and interactions with key residues within the substrate tunnel. The identification of an oxyanion hole, a conserved feature in α/β -hydrolases, gave further insights into the catalytic mechanism of the FrsA TE domain.

A loop located beneath the active site exhibited significant variability among TE domains. This region, potentially involved in substrate binding and recognition, warrants further investigation. While no direct substrate interaction was observed in the analyzed structures, its proximity to the substrate binding pocket suggests a potential role in the non-canonical reaction catalyzed by FrsA_{TE}.

By comparative studies with TE domains having similar structures, well-founded hypotheses regarding the functional mechanism of FrsA_{TE} could be formulated. Given that FrsA_{TE} catalyzes an intermolecular transesterification only observed in two other TE domains (Sln9_{TE}, YtfA_{TE}) with unknown mechanisms, a comprehensive understanding of the underlying mechanism could not

be achieved through comparisons alone. The unique property of utilizing FR-Core as a nucleophile is particularly noteworthy. Consequently, the interaction between FrsA_{TE} and FR-Core was the primary focus of this investigation.

Since co-crystallization and crystal soaking attempts did not yield the desired *holo*-form, the precise substrate binding site, and substrate organization into the active site could not be determined. However, based on the elucidated structure, several residues were identified as potential candidates for substrate interactions. These findings provide a foundation for further investigations. For instance, site-directed mutagenesis could be used to introduce point mutations into these identified residues. By assessing the reaction product(s), it would be possible to determine whether the mutated residue is involved in substrate binding. The method required for this approach has already been established in this study and applied to FrsH and FrsA_{CAT} (4.5).

To circumvent the limitations of the *apo*-form crystal structure, a strategy to generate a substrate-bound T-TE domain complex was devised (4.1.4). This involved a modified one-pot assay where the FrsA protein was separated into the functional modules FrsA_{C+A} and FrsA_{T+TE} and supported by the MLP FrsB and the modifying enzyme FrsH. While the approach showed promising results in protein purification and assay execution, the goal of producing a substrate-bound complex was not achieved. Since this outcome suggests a lack of interaction between the didomains FrsA_{C+A} and FrsA_{T+TE}, experiments could be conducted to connect the modules using a linker unit with an inserted cleavage site (e.g. TEV). This way, the didomains could be separated before crystallization to release the side chain linked FrsA_{T+TE} complex. Further, mutating the catalytic serine or another residue of the catalytic triad should also be considered to prevent the reaction from occurring after substrate addition in the crystallization process, thereby avoiding the formation of the *apo*-form.

Another explanation for the absence of the *holo*-form enzyme of the co-crystallization approaches could be the lacking phosphopantetheinylation of the T domain. Without the Ppant arm, the T domain is unable to transfer the side chain synthesized by FrsA_{CAT} to the active site of the TE domain. It is reasonable to assume that the presence of this flexible arm influences substrate binding of the TE domain. However, the 3D NMR data, which exclusively analyzed FrsA_{TE} in absence of the T domain, contradict this hypothesis. Despite the absence of the Ppant arm, these results demonstrate a significant structural alteration caused by the addition of the nucleophilic substrate FR-Core. Consequently, it can be inferred that an interaction between the substrate and the protein can occur even without the Ppant-bound T domain. Nonetheless, given that the didomain represents a distinct construct that may induce different interactions between the involved molecules, it is advisable to mimic the native conditions of the reaction as closely as possible. To this end, incubation with a PPTase would be beneficial, as it would catalyze the

phosphopantetheinylation of the catalytic S35 of the T domain, thereby facilitating the formation of the catalytically essential Ppant arm and supporting proper substrate binding and *holo*-crystal formation of the FrsA TE domain.

As with many TE domains, also the FrsA_{TE} primarily forms dimers, enabling the use of cryo-electron microscopy for structure elucidation. Despite successful sample preparation and data acquisition, the small sizes of the recorded particles indicated monomeric proteins, hindering 2D classification and thus the generation of high-resolution structural data (4.2). One explanation for this result could be that the protein used for these studies had been stored at -80 °C for an extended period, possibly leading to degradation or dissociation of dimers. An experiment with freshly purified protein might provide the desired datasets. The urgently needed dimerization could also be supported by buffer adaptations (pH, salt concentration) or employing methods such as metal ions, molecular glue, host-guest systems, or G-quadruplexes.^[246] However, it is important to note that these methods may influence substrate binding or the protein mechanism to an unknown extent.

Within the scope of this work, the constructs His-MBP-TEV-Linker²-FrsA_{TE}Short/long were produced (4.1.2.5.1), which, due to the use of MBP, have a size of over 70 kDa and thus exceed the critical size threshold for cryoEM. Consequently, these could also be used for further cryo-EM investigations. Cloning the FrsA_{T+TE} didomain into the expression vector pHis-MBP-TEV-Linker² would result in a product of approximately 80 kDa, promising a higher chance of achieving sufficient resolution with cryo-EM. However, there is also the risk that the MBP may have a negative influence, making it probably the most straightforward option to conduct initial tests with freshly purified protein and, if necessary, to support dimerization to avoid any interfering effects.

5.3. Optimization approaches for FrsA_{TE} expression in labeled M9 medium

Initial attempts to express FrsA_{TE} in M9 medium using the spin-down approach yielded insufficient protein quantities for NMR analysis (4.3.1). Extensive optimization, including variations in expression strain, duration, temperature, and IPTG concentration was necessary to establish optimal conditions for protein production. The optimized protocol, utilizing *E. coli* LOBSTR induced with 0.1 mM IPTG, and incubated for 18 h at 20 °C, provided sufficient ¹⁵N-labeled protein for initial NMR experiments (4.3.1.1). The ¹⁵N-BEST-TROSY NMR spectra revealed successful protein labeling and indicated potential protein dynamics upon substrate addition, as evidenced by chemical shifts of several peaks (4.3.3). However, the spectra also highlighted challenges such as dimerization, proteolytic degradation, and concentration-dependent

instability. Through comprehensive optimization of purification and storage buffers including additive, pH adjustments, and storage conditions (purification buffer additives: 25 mM NaCl, 5 mM DTT, 1x protease inhibitors; storage buffer: 100 mM Tris, 1 mM imidazole, 5 mM DTT, 1x protease inhibitors, pH=7.5; storage/transport: 4 °C), the protein was stabilized, enabling the production of sufficient quantities of ^{13}C - ^{15}N -FrsA_{TE} for 3D NMR measurements (4.3.2.1). These experiments were expected to facilitate complete resonance assignment, allowing the identification of residues involved in conformational changes and protein-substrate interactions (4.3.3.1).

5.4. Investigation of protein dynamics of FrsA_{TE}

Despite successful NMR signal detection, the comprehensive resonance assignment of FrsA_{TE} remained incomplete. Due to several weak and superimposed signals, only 1/3 of the FrsA_{TE} residues could be assigned (4.3.3.1), excluding key structures hypothesized to be involved in protein dynamics and substrate interaction. Their structural characterization is however essential for a complete understanding of the FrsA_{TE} mechanism, while the active site residues within the TE domain are well-conserved and their specific assignment in this study was not critical for overall interpretation.

NMR analysis revealed several distinct chemical shifts upon substrate binding. These findings suggest substantial conformational changes induced by FR-Core, in contrast to the more subtle effects of the side chain. The observed attenuation of chemical shifts upon simultaneous addition of both ligands may indicate substrate turnover and enzyme reversion to the initial state.

The NMR data support a dynamic model for FrsA_{TE} catalysis, where substrate binding induces conformational changes, facilitating substrate entry in the active site and subsequent reaction. Unfortunately, the lid structure (αL1 - αL2) could not be fully assigned, however, the residues at its termini, undergo significant conformational changes upon FR-Core presence, suggesting substrate coordination or binding. The variable loop between $\beta 6$ and $\beta 7$ could also not be fully assigned except for some residues at the C-terminus, which exhibited distinct shifts in structure. These support the assumption of involvement of substrate interaction, organization, or structural rearrangement. However, the absence of complete resonance assignments for both regions precludes definitive conclusions about their precise roles.

Regions at the 'back' of the protein (αB and surrounding β -sheets) showed a high sequence coverage through resonance assignment. Furthermore, these regions undergo significant chemical environment changes, indicating conformational alterations. These findings suggest a more extensive structural reorganization than initially anticipated. The loop region adjacent to the lid

also exhibited dynamic behavior, although its precise involvement in substrate binding or structure rearrangement remains unclear.

The observations did not lead to the identification of the residues directly interacting with or organizing FR-Core. Additionally, the NMR approach failed to give a deeper insight into the lid structure movement, while substrate acceptance and reaction catalyzation. It could only be ascertained that the structural rearrangement of FrsA_{TE} appears to involve the entire enzyme. However, the reaction mechanism could not be fully elucidated.

Certainly, the potential of NMR measurements has not yet been fully exhausted. Since the protein can be reliably and fully labeled in sufficient quantities, measurements could be repeated to achieve clearer signal separation. In addition, it is also feasible to add labeled FR-Core to determine a distinct interaction of specific residues.

Furthermore, the sequence of FrsA_{TE} contains several unique amino acid pairs, occurring only once in the entire structure. In this work, three combinatorial labeling approaches were conducted, showing high yields and significantly advancing the assignment. It is conceivable to produce additional combinatorial samples, in which various residues and unique pairs are labeled, further supporting the assignment and thus contributing to the identification of key structures.

5.5. Exploration of substrate specificity of the FrsA domains and obtainment of new FR analogs

The discovery of naturally occurring FR analogs gave rise to investigations of the structural determinants of their potent and selective inhibitory activity on Gq proteins. The results underscored the critical role of the FR side chain, that is biosynthesized by the FrsA module. Notably, recent structural analyses revealed the critical interaction between the side chain and the Gβq subunit, resulting in the formation of a 'silent' trimeric complex that completely abrogates downstream signaling.^[55] By systematically varying the structure of the side chain, the aim was to gain valuable insights into the structure-activity relationships influencing the interaction between FR analogs and the Gq protein. Recent findings not only underscore the importance of the FrsA catalyzed side chain for FR's inhibitory potency but also provide a robust foundation for the rational design of novel Gq inhibitors with improved pharmacological properties, promising therapeutic applications in diseases characterized by aberrant Gq activity (**2.3.2**).

For further investigations of the structure-activity relation of FR analogs, the substrate specificity of the individual FrsA domains was determined by systematically substituting the natural substrates L-leucine and propionyl-CoA. These modified side chains were synthesized *in vitro*

using the established side chain assembly assay, and further attached to FR-Core by subsequently performing the transesterification assay (4.4.1).

After optimization trials of the established assays, the optimal conditions of the side chain assembly assay (10 μ M methyl viologen, 1 mM NADH, 10 μ M FrsH, 10 μ M FrsA_{CAT}, 1 mM ATP, 1 mM L-leucine, 100 μ M propionyl-CoA, 100 μ M DTT; incubation for 1 h at 20 °C) and the TE domain assay (100 μ M FR-Core, 5 μ M FrsA_{TE}; incubation for 16 h at 22 °C) were set and used for the substrate exchange approaches. While combining these assays proved challenging for quantitative kinetic analysis due to the multi-step nature of the reaction, it provided a suitable basis to assess qualitative FR analog formation (4.4.2).

Previous experiments of Dr. Hermes already demonstrated the substrate promiscuity of FrsA_C, by the exchange of propionyl-CoA with butyryl-CoA and acetyl-CoA. In this work, these results could be reproduced, and it was further possible to replace the native substrate with isobutyryl-CoA, and octanoyl-CoA. Only malonyl-CoA completely impeded the product formation and was thus excluded from the combinatorial substrate exchange experiments (4.4.3).

The FrsA A domain only showed an acceptance towards the L-leucine isomers, D-leucine and isoleucine. Feeding of valine did not lead to the formation of an FR analog.

The amino acids and acyl-CoAs that had previously tested positive for product formation were then combined in the *in vitro* assays, which was expected to lead to the formation of FR analogs harboring completely new side chains. These were indeed detectable by HPLC-MS measurements, except for the combinations of octanoyl-CoA with the amino acid isomers. Only the octanoyl-L-leucine FR analog could be detected in minor amounts. The twelve newly generated analogs were in a range from 23 up to 59% FR-analog formation relative to the native substrate combination.

The absence of FR analogs with unhydroxylated side chains demonstrated the low substrate specificity of FrsH, whereby the formation of FR analogs also emphasized that the TE domain has a high acceptance towards the modified side chain. However, whether the absence of a product depends on the substrate specificity of the C or A domain and the side chain was thus not formed, or if the TE domain did not accept the modified side chain and therefore no analog was detected, could not be determined due to the lack of high-resolution measurements. Thorough structure verification of the synthesized FR analogs is imperative before initiating investigations into their G protein inhibitory properties. Consequently, comprehensive characterization employing HPLC-MS/MS and structure elucidation *via* NMR spectroscopy after upscaling of the assay and preparative isolation of the analogs will be essential. This necessity was underscored by the unexpected and pervasive presence of FR and FR2 isomers within FR analog samples.

To investigate whether the flexible substrate specificity also extends the nucleophilic substrate of the TE domain, a knockout mutant was generated for *in vivo* production of linearized FR-Core (LFRC) (4.4.4.1). The aim was to produce a FrsA and FrsG8_{TE} deficient *C. vaccinii* mutant, which accordingly lacks side chain formation and cyclization of the FR-Core precursor molecule. The *E. coli* NEB10 β pEX18Tc: Δ frsG8_{TE}-FRT helper strain could be successfully obtained by Gibson assembly and was applied in a triparental conjugation with *C. vaccinii* Δ vioA Δ frsA, where frsG8_{TE} was replaced by the FRT-cassette in the extrachromosomal plasmid. The FRT-cassette should be removed by biparental conjugation through Flp-based recombination, however, this was not achieved within the scope of this study. Fermentations of the *C. vaccinii* Δ vioA Δ frsA Δ frsG8_{TE}:FRT with subsequent extraction and HPLC-MS measurement revealed a complete abolishment of FR and FR-Core biosynthesis. Although a peak corresponding to the theoretical *m/z* of LFRC was observed, showing a higher retention time, the quantity was insufficient for isolation and purification beyond a crude extract. When used as a substrate in the coupled *in vitro* assays (4.4.4.2), the expected linear FR product peak could not be detected at the calculated *m/z*. Thus, the specificity of FrsA_{TE} in catalyzing the transesterification of the side chain and LFRC could not be ascertained within the scope of this study. HPLC-MS/MS measurements are required to clearly identify LFRC and the possible reaction product.

If the product of the modified biosynthesis does not correspond to the desired LFRC, the strain *C. vaccinii* Δ vioA Δ frsA Δ frsG8_{TE}:FRT could be further processed by removing the FRT cassette *via* biparental conjugation. As the frsG8_{TE} gene was already replaced by the cassette, biosynthesis was interrupted, as evidenced by the absence of detectable FR-Core. However, the exact influence of the FRT cassette on FR biosynthesis remains unclear and may lead to impairments.

The utilization of alkaline hydrolysis to detach the peptide intermediate from the FrsG8_T could also have caused undesired side effects. Given the potential instability of the linear peptide, LFRC might be produced during biosynthesis but subsequently degraded during extraction.

Alternatively, hydrolysis could be utilized to modify FR-Core. Thus, the cyclic depsipeptide could be cleaved at the two adjacent ester bonds by enzymatic or chemical hydrolysis. This could lead to the formation of the linearized FR core, but also to the loss of the *N*-Ac-Hle building block. However, unless the peptide is completely dissolved, the hydrolysis product could be used as a nucleophilic substrate for the FrsA TE domain.

Should further investigations of the FrsA_{TE} structure and dynamics lead to the clarification of which residues are crucially involved in protein-substrate interaction, more potential substrates for FrsA_{TE} could be identified *via* this approach and be tested in *in vitro* assays. These approaches will provide detailed insights into the unusual mode of action and evaluate its potential for chemoenzymatic syntheses, such as generating altered, bioactive FR analogs with novel effects.

6. Material and Methods

6.1. Chemical and reagents

All chemicals used in this work were purchased from Carl Roth (Karlsruhe, Germany) and Merck KGaA (Darmstadt, Germany). Enzymes were procured from New England Biolabs (Frankfurt am Main, Germany) and Promega (Mannheim, Germany).

6.2. Vectors and plasmids

Table 1: Used vectors and plasmids in this work

Vector/Plasmid	Purpose/Specification	Resistance (concentration)	Origin
pET28a(+)	Heterologous expression C- and N-terminal 6x His-tag T7 promotor <i>lacI</i> operator	Kanamycin (50 µg/mL)	Merck KGaA (Darmstadt, Germany)
pCDFDuet-1(Apra)	Heterologous expression MCS1: N-terminal 6x His-tag MCS2: N-terminal Strep-tag T7 promotor <i>lacI</i> operator	Apramycin (50 µg/mL)	Prof. Craig Townsend, modified by Dr. Nils Böhringer
pHis8-TEV	Heterologous expression N-terminal 8x His-tag T7 promotor <i>lacI</i> operator TEV cleavage site	Kanamycin (50 µg/mL)	Dr. René Richarz
pHis8-MBP-L-TEV	Heterologous expression N-terminal 8x His-tag T7 promotor <i>lacI</i> operator maltose binding protein (fusion protein)	Kanamycin (50 µg/mL)	This work
pHis8-mVenus-L-TEV	Heterologous expression N-terminal 8x His-tag T7 promotor <i>lacI</i> operator mVenus (fusion protein, GFP ⁺)	Kanamycin (50 µg/mL)	This work
pEX18Tc	Knock-out vector lac promotor	Tetracyclin (10 µg/mL)	Dr. René Richarz

	lac operator		
pPS858	Contains GmR-GFP FRT cassette, Ap ^R , Gm ^R , gfp ⁺ ,	Gentamycin (30 µg/mL)	Dr. René Richarz
pUB307	Mobilization helper plasmid RP1 derivative ^[247]	Chloramphenicol (25 µg/mL)	Dr. René Richarz
pTA-Mob	Conjugation helper plasmid Pc, HIS3, URA promotor	Gentamycin (20 µg/mL)	Dr. René Richarz
pMLBAD_tet::Flp/SacB	Vector optimized for inducible gene expression in Burkholderia araBAD, cat, Pc promotor	Tetracyclin (10 µg/mL)	Dr. René Richarz

6.3. Media, Buffers, and Solutions

Table 2: Composition of media for cultivation

Medium	Composition	
Lysogenic-Broth (LB)	10 g/L	Tryptone
	5 g/L	Yeast extract
	10 g/L	NaCl
		pH 7.5
LB-agar	10 g/L	Tryptone
	5 g/L	Yeast extract
	5 g/L	NaCl
	15 g/L	Agar-agar
		pH 7.5
Terrific-Broth (TB)	12 g/L	Tryptone
	24 g/L	Yeast extract
	5 g/L	Glycerol
	100 mL/L	TB salts (10x)
		pH 7.5
TB salts (10x)	0.17 M	KH ₂ PO ₄
	0.72 M	K ₂ HPO ₄
M9 minimal medium	60.28 mM	K ₂ HPO ₄
	33.07 mM	KH ₂ PO ₄
	8.56 mM	NaCl
	18.69 mM	NH ₄ Cl
	2.00 mM	MgSO ₄
	10.00 µM	FeCl ₃

	11.10 mM	Glucose
	13.60 μ M	CaCl ₂
	6.96 μ M	ZnSO ₄
	11.83 μ M	MnSO ₄
	188.43 μ M	Thiamine
	406.14 μ M	Nicotinic acid
	4.09 μ M	Biotin
		pH 7.0
Super optimal broth with catabolite repression (SOC medium)	20 g/L	Yeast extract
	5 g/L	Tryptone
	10 mM	NaCl
	2.5 mM	KCl
	20 mM	Glucose

Table 3: Composition of buffers

Buffer	Composition	
Lysis buffer	50 mM	NaH ₂ PO ₄
	300 mM	NaCl
	10 mM	Imidazole
		pH 8.0
Wash I buffer	50 mM	NaH ₂ PO ₄
	300 mM	NaCl
	20 mM	Imidazole
		pH 8.0
Wash II buffer	50 mM	NaH ₂ PO ₄
	300 mM	NaCl
	35 mM	Imidazole
		pH 8.0
Elution buffer	50 mM	NaH ₂ PO ₄
	300 mM	NaCl
	250 mM	Imidazole
		pH 8.0
Separating buffer	1 M	Tris-HCl
		pH 8.8
Stacking buffer	0.375 M	Tris-HCl
		pH 6.8
SDS Running Buffer (10x)	250 mM	Tris-HCl
	2 M	Glycine

	1%	SDS
		pH 9.0
A domain assay buffer	20 mM	Tris
	5% (v/v)	Glycerol
		pH 7.5
Sidechain assembly assay buffer	50 mM	Tris
	10 mM	MgCl ₂
	150 mM	NaCl
		pH 7.5
TE domain buffer	25 mM	HEPES
	50 mM	NaCl
		pH 7.5
NMR buffer	100 mM	Tris
	5 mM	DTT
	1 mM	Imidazole
		pH 7.5
TAE buffer	1 mM	EDTA
	40 mM	Glacial acetic acid
	20 mM	Tris
		pH 8.0
ISO buffer	10 mM	MgCl ₂
	200 µM	dNTPs
	10 mM	DTT
	5% (w/v)	PEG-8000
	1mM	NAD
	100 mM	Tris-HCl
		pH 7.5
Transformation buffer	10 mM	HEPES
	15 mM	CaCl ₂
	250 mM	KCl
	55 mM	MnCl ₂

Table 4: Composition of solutions

Solution	Composition	
Staining solution	1 g	Coomassie Blue
	100 mL	Glacial acetic acid
	400 mL	Methanol
	500 mL	ddH ₂ O

Destaining solution	200 mL	Methanol
	100 mL	Glacial acetic acid
	700 mL	ddH ₂ O

Table 5: Antibiotic stocks (x1000)

Reagent	Concentration [mg/mL]	Solvent
Kanamycin	50	ddH ₂ O
Apramycin	50	ddH ₂ O
Carbenicillin	50	50% EtOH / 50% ddH ₂ O
Tetracyclin	10	70% EtOH / 30% ddH ₂ O
Chloramphenicol	25	100% EtOH
Gentamycin	30	ddH ₂ O
Ampicillin	100	ddH ₂ O

6.4. Microbiological methods

6.4.1. Cultivation of bacteria

In this work, strains of the Gram-negative bacteria *E. coli* and *C. vaccinii* MWU205 were used. The strains and their application are listed in **Table 6**.

Table 6: Bacterial strain used in this work.

Strain	Genotype	Application	Origin
<i>E. coli</i> α-select silver	F ⁻ <i>deoR endA1 recA1 relA1 gyrA96 hsdR17</i> (r _K ⁻ , mk ⁺) <i>supE44 thi-1 phoA Δ(lacZYA argF)U169 Φ80lacZΔM15λ⁻ Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- φ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL</i> (Str ^R) <i>rph spoT1 Δ(mrr-hsdRMS-mcrBC)</i>	Cloning, plasmid storage	Bioline
<i>E. coli</i> NEB 10β	<i>fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHI ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5</i>	Cloning, plasmid storage	NEB® inc.
<i>E. coli</i> BL21 (DE3)	BL21(DE3) <i>ΔprpRBCD::T7prom-sfp</i> , T7prom- <i>prpE</i> [180]	Heterologous expression	NEB® inc.
<i>E. coli</i> BAP1	BL21(DE3)-RIL <i>fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHI ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5 arnA</i> (H359S, H361S, H592S, H593S) <i>slyD</i> (1-150)	Heterologous expression	Prof. Bradley Moore
<i>E. coli</i> LOBSTR	F ⁻ <i>mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ (ara-leu)7697 galU galK λ- rpsL</i> (Str ^R) <i>nupG</i>	Heterologous expression	Kerafast/ Thomas. U. Schwartz
<i>E. coli</i> DH10β	methylation defective strain <i>dam-13::Tn9, dcm-6, hsdkf</i>	Cloning	Thermo Scientific™
<i>E. coli</i> ET12567	Wild type (type strain)	Helper strain conjugation	
<i>Chromobacterium vaccinii</i> MWU205 DSM25250	<i>ΔvioA ΔfrsA</i>	Donor DNA	DSMZ
<i>C. vaccinii</i> MWU205 <i>ΔvioA ΔfrsA</i>		Mutagenesis	Dr. René Richarz, AG Crüsemann

The *E. coli* strains were cultivated in LB, TB, and M9 medium at 37 °C and 220 rpm. *E. coli* strains with tetracycline resistance were cultivated at 30 °C as well as 200 rpm and *C. vaccinii* MWU205 and the *C. vaccinii* knock-out mutation strains. Precultures were inoculated with cell material transferred from a cryoculture into the medium using a toothpick and incubated for 16 up to 24 h.

6.4.2. Optical density measurement

The growth of the bacterial strains was monitored by measuring the increase in optical density of the cultures at 600 nm (OD₆₀₀) with the Eppendorf BioPhotometer and the FastGene NanoSpec Photometer. For each measurement, 1 mL of the culture was sampled under sterile conditions and measured in the devices using a cuvette.

6.4.3. Strain maintenance

Cryogenic cultures were prepared for long-term storage of the newly generated strains. For this purpose, 1 mL of a preculture was transferred in a cryotube and mixed with 500 mL of 50% glycerol, until homogeneity was achieved. The cryogenic cultures were then stored at -80 °C. While using, it was always ensured that they were kept in a chilled container to avoid thawing.

6.4.4. Preparation of chemically competent cells

To enable *E. coli* cells to receive plasmid DNA through heat-shock transformation, they must first be made chemically competent. In this work two protocols were used to gain chemically competent cells: the CaCl₂ method and the preparation of super-competent cells.

6.4.4.1. Enabling chemical competence with calcium chloride

One milliliter of a preculture (6.4.1) was added to 99 mL LB medium in an Erlenmeyer flask and incubated at 37 °C and 220 rpm until an OD of 0.4-0.5 was reached. The culture was transferred into centrifuge tubes, stored on ice for 30 min, and centrifuged at 4 °C and 4000 rpm for 10 min. The supernatant was discarded, the pellet was resuspended in 40 mL 0.1 M CaCl₂ and again stored on ice for 30 min. After a further centrifugation step, the supernatant was removed, and the cells were resuspended in 5 mL 0.1 M CaCl₂ with 15% glycerol. The mix was then aliquoted into 100 µL portions, flash-frozen with liquid nitrogen, and stored at -80 °C until used for transformation.

6.4.4.2. Preparation of super-competent cells

A preculture of the desired *E. coli* strain was prepared as described in 6.4.1. 100 mL SOC medium in an Erlenmeyer flask was inoculated with 4% preculture and then cultivated at 18 °C and 200 rpm for 16-20 h until an OD600 of 0.4-0.6 was achieved. After 10 min storage on ice, the cells were transferred in a sterile, chilled centrifuge tube and centrifugated at 4 °C and 4000 rpm for 10 min. The supernatant was removed sterile, and the cells were resuspended in 32 mL chilled transformation buffer. After further 10 min on ice, the cells were centrifugated again at 4 °C and 4000 rpm for 10 min. The supernatant was removed, and the cell pellet was resolved in 7.44 mL chilled transformation buffer until no cell lumps remained. Sterile, warm DMSO was added to a final concentration of 7%, and the cells were stored on ice until 100 µL aliquots were prepared in chilled 1.5 mL reaction tubes, which were flash-frozen with liquid nitrogen and stored at -80 °C until usage.

6.4.5. Transformation

To introduce external DNA, like plasmids, into bacterial cells, the method of the chemical (heat-shock) transformation was used in this work. Therefore, the cells were briefly exposed to high temperatures and then cooled down quickly, leading to an increase in the permeability of the cell membranes, allowing the DNA to permeate the cell.

After thawing one aliquot of chemically competent cells (6.4.4) on ice, 1 µL plasmid DNA or 5 µL of a Gibson assembly mix was added, mixed gently, and stored on ice for 15-30 min. Then, the cells were heated at 42 °C for 45-90 s and subsequently chilled on ice for 2 min. One milliliter LB medium was added, and the cells were shaken at 37 °C and 200 rpm for 1 h for regeneration. After the incubation the cells were centrifuged at 4000 rpm for 1 min, the supernatant was removed by inverting the tube carefully and the cell pellet was resolved in the remaining supernatant (~100 µL) and plated on selective agar. The plate was incubated at 37 °C overnight.

6.5. Molecular biological methods

6.5.1. Primer design

For the amplification and investigation of the *frs* gene cluster, primers were designed using the program SnapGene 2.3.2 (Eurofins MWG Operon, Ebersberg, Germany). It must be ensured that the annealing region is 18-27 bp long, has a GC content of 40-60%, begins and ends with a G/C, and has a melting temperature between 45-72 °C (depending on the polymerase used). For the insertion of a gene of interest (GOI) into a vector through ligation, restriction enzyme (RE)

cleavage sites were attached to the annealing part. To maintain proper functionality of the RE a TATA motif was added to the RE recognition motif. It is also necessary to be aware that the inserted GOI is within the reading frame and that a C-terminal stop codon is included if an N-terminal His-tag is used. To combine several fragments using Gibson assembly, primers were designed that contain overlapping regions of 20-40 bp from the fragments to be combined.

The lyophilized primers were resolved in ddH₂O to a concentration of 100 µM. For the PCR approaches, the primer solution was diluted again 1:10 to a final concentration of 10 µM and aliquoted.

Table 7: Primers for sequential cloning of *frs* genes. Restriction sites are underlined, and the annealing sequence is bold.

Resulting construct	Primer name	Sequence (5'-3')
pHis8TEV:FrsA _{TE} short	Cv_FrsATEs_for_BamHI	tat <u>ggatcc</u> cccttgcttctgcattcatc
	Cv_FrsATEs_rev_HindIII	tat <u>aagctt</u> ttaagacaacttcgactcgag
pHis8TEV:FrsA _{TE} long	Cv_FrsATEl_for_BamHI	tat <u>ggatcc</u> cccttgcttctgcattcatc
	Cv_FrsATEl_rev_HindIII	tat <u>aagctt</u> ttattgcttgacagcggtgac
pET28a:FrsA _{C+A}	Cv_FrsACA_for_BamHI	tata <u>ggatcc</u> aacagtgaatcgccaatcca
	Cv_FrsACA_rev_HindIII	tata <u>aagctt</u> aatgcgagccgccaac

Table 8: Primers for Gibson assembly of *frs* genes. Gene annealing sequence is bold, inserted linker is underlined.

Resulting construct	Primer name	Sequence (5'-3')
pET28a:FrsA _{T+TE}	Cv_pET_FrsATTE_for	actggtggacagcaaagggtcgcatc agcgccgcgcaa
	Cv_pET_FrsATTE_rev	tggtgctcgagtgcggccgcaagctt attgcttgacagcggtgacagaca ac
pHis8TEV:FrsA _{T+TE}	Cv_pHT_FrsATTE_for	tgagaatcttattttc aggaggacagcgccgcgcaa
	Cv_pHT_FrsATTE_rev	tcgagtgcggccgcat tattgcttgacagcggtgacag
pET-His-MBP-TEV-FrsA _{TE} short	pHMT_FrsATEs_for	atctgagaacctgtacttcca atcccccttgcttgcattcatcc
	pHMT_FrsATEs_rev	tttgctgtccaccagtcagctagct taagacaacttcgactcgagaa
pET-His-MBP-TEV-FrsA _{TE} long	pHMT_FrsATEl_for	atctgagaacctgtacttcca atcccccttgcttgcattcatcc
	pHMT_FrsATEl_rev	tttgctgtccaccagtcagctagct tattgcttgacagcggtga
pET-His-MBP-TEV-L-FrsA _{TE} short	pHMTL_FrsATEs_for	cctgtacttccaat ccggaggaggatctggaggaggatctcccttgcttgc ttcatcccgaggcgagg
	pHMTL_FrsATEs_rev	atccgcgacccatttgctgtccaccagtcagctagcttattgcttgacagcggt ctaagacaact

pET-His-MBP- TEV-L- FrsA _{TE} long	pHMTL_FrsATEL_for pHMTL_FrsATEL_rev	cctgtacttccaat <u>ccggaggaggatctggaggaggatct</u> cccttgttctgca ttcatcccaggggcg cggatccgcgacccatttgctgtccaccagtcagctagct ttattgcttgaca gcggtgacagaca
pET-His-mVenus- TEV-L- FrsA _{TE} short	pHmVTL_FrsATEs_for pHmVTL_FrsATEs_rev	ccgcgcggcagccatat tggtgagcaagggcgaggagct gattggaagtacaggtt ctccttgtagctcgtccatgcc
pET-His-mVenus- TEV-L- FrsA _{TE} long	pHmVTL_FrsATEL_for pHmVTL_FrsATEL_rev	ccgcgcggcagccatat tggtgagcaagggcgaggagct gattggaagtacaggtt ctccttgtagctcgtccatgcc
pET-MBP-TEV	pET_MBP_TEV_for pET_MBP_TEV_rev	gctagcatgactggtggacagcaaa ggattggaagtacaggttctcagat
mVenus	mVenus_for mVenus_rev	ccgcgcggcagccatatggtgagcaagggcgaggagct gattggaagtacaggttctccttgtagctcgtccatgcc

Table 9: Primers for the construction of *C. vaccinii* MWU205 deletion mutants. Gene annealing sequence is bold, the overhang to FRT cassette is underlined, and the overhang to plasmid is roman.

Primer name	Sequence (5'-3')	Function
ΔfrsG8TE_up_for	gattacgaattcgagctcggtagccggggatcctctagag agaggacggc gaagcgc	Upstream region of <i>frsG8TE</i>
ΔfrsG8TE_up_rev	ataggaacttcggaataggtacttcaagatcccaattcgtag tgtaatgc agaacagcggc	
ΔfrsG8TE_dn_for	gaaagtataggaacttcagagcgcgttttgaagctaattcgat cgaaacgca agctggccacgc	Downstream region of <i>frsG8TE</i>
ΔfrsG8TE_dn_rev	cacgacgttgtaaaacgacggccagtgccaagcttgcat gcaagctgaa ggcgatttcctgt	
FRT_for	cgaattagcttcaaaagcgtctga	Amplification
FRT_rev	cgaattggggatcttgaagttcct	FRT cassette
pEX18Tc_control_for	aaaccgcctctccccgcgc	Identification of correct Gibson assembly
pEX18Tc_control_rev	agctcgccattcgccattcaggc	
Gap_5_for	cccaatcatctggtattg	Colony screening of <i>C.</i> <i>vaccinii</i> Δ <i>vioA</i> Δ <i>frsA</i> Δ <i>frsG8TE_FRT</i>
PCR_FrsH_rev	ggtccaaccagatttcc	
Seq_KO-G8TE_for	cgaagtgggtgttcgcaag	Amplification of <i>frsG8TE</i> gene region for sequencing
Seq_KO-G8TE_rev	gctgatcctccggaatctcc	

6.5.2. Polymerase chain reaction

The polymerase chain reaction is a molecular biology method to amplify specific DNA sequences using designed primers for the GOI. Different polymerases were used depending on the desired quality for the PCR products: For the cloning with subsequent heterologous expression, the Q5 polymerase, and for the monitoring of the cloning or Gibson assembly success the *Taq* polymerase is used. These two approaches follow specific protocols which are described in the following chapters.

6.5.2.1. Q5 PCR

The Q5 polymerase chain reaction is performed to gain a PCR product of high quality for cloning and subsequent heterologous expression. The used Q5 high fidelity DNA polymerase (New England Biolabs), is a *pfu* polymerase, having a proofreading activity, leading to a fast amplification with an ultra-low error rate. The composition of a standard 50 μ L reaction mix is shown in **Table 10**.

Table 10: Composition of a Q5 PCR reaction mix with a final volume of 50 μ L.

Volume [μ L]	Component
23	H ₂ O
10	5 x Q5 reaction buffer
10	Q5 high GC enhancer
2.5	Forward primer (10 μ M)
2.5	Reverse primer (10 μ M)
1	dNTPs (10mM)
0.5	Template
0.5	Q5 DNA polymerase

The genomic DNA of *Chromobacterium vaccinii* MWU205 was used as a template with a concentration of ~ 10 ng/ μ L. The components were pipetted in the order shown in the table, mixed thoroughly, and placed in the PCR cycler. The program can be found in **Table 11**.

Table 11: PCR program of a Q5 PCR reaction.

Step	Temperature [°C]	Time [s]
Initial denaturation	98	30
Denaturation	98	10
Annealing	50-70	20
Polymerization	72	20-30 s/kb
Final polymerisation	72	120
Chilling	4	∞

The annealing temperature was determined in advance *via* gradient PCR (6.5.2.3) or calculated with the NEB T_m calculator (<https://tmcalculator.neb.com/#!/main>). The polymerization duration depended on the length of the desired fragment.

6.5.2.2. GoTaq PCR

The PCR performed with the *GoTaq* DNA polymerase was used to investigate the transformation and Gibson assembly results. This polymerase gives robust amplification but does not harbor a proofreading activity, which is why it was only used for progress monitoring. The standard *GoTaq* reaction mix can be found in the following table.

Table 12: Composition of a *GoTaq* PCR reaction mix with a final volume of ~21 µL.

Volume [µL]	Component
10	H ₂ O
5	5 x <i>GoTaq</i> Puffer
2.5	MgCl ₂ (10mM)
1.25	DMSO
0.5	Primer for (10 µM)
0.5	Primer rev (10 µM)
0.5	dNTPs [10mM]
0.125	<i>Taq</i> Polymerase
1	Template

For the colony screening, 11 μL H_2O was provided in an PCR reaction tube. A small amount of colony was transferred into the tube with a toothpick and resuspended properly. The remaining components of the reaction mix were assembled, and the appropriate volume of this master mix was added to the colony mix, which served as a template.

To investigate the product of the Gibson assembly, 1 μL of the Gibson assembly mix was used as a template.

All components were combined and mixed in a PCR reaction tube, which was then placed into a PCR cycler. The program was set as shown in the following table.

Table 13: PCR program of a GoTaq PCR reaction.

Step	Temperature [°C]	Time [s]	
Initial denaturation/ Cell lysis	95	30/360	
Denaturation	95	30	30 cycles
Annealing	45-68	45	
Polymerization	68	1 min/kb	
Final polymerisation	68	360	
Chilling	4	∞	

For the colony screening, 360 s were used for cell lysis as an initial step to make the DNA accessible, while 30 s were used to investigate the Gibson assembly mix.

The annealing temperature was determined in advance *via* gradient PCR (6.5.2.3) or calculated with the NEB T_m calculator (<https://tmcalculator.neb.com/#!/main>). The polymerization duration depended on the length of the desired fragment.

6.5.2.3. Gradient PCR

To find the ideal annealing temperature for the used primers a gradient PCR was performed. Therefore, the PCR mixture was prepared as described in Chapters 6.5.2.1 and 6.5.2.2, depending on the primer application. The mix was then split into 10 μL per tube (x12) and placed into the gradient cycler. The program was set appropriate to the protocols for the used DNA polymerase. The cycler is suitable for applying a gradient for the annealing temperature, whereby each reaction tube has a specific annealing temperature, which can then be analyzed using agarose gel electrophoresis to determine the optimal temperature for primer annealing.

6.5.3. Agarose gel electrophoresis

An agarose gel electrophoresis is a technique to separate and analyze DNA fragments according to their size. The agarose gel forms a matrix through which small DNA fragments migrate faster and larger fragments migrate more slowly, resulting in separation based on fragment size. Due to the applied voltage during electrophoresis, the DNA migrates through the gel according to its negative charge to the cathode.

For the gel preparation, 1% agarose (w/v) was added to 1x TAE buffer and heated in a microwave until the agarose is completely dissolved. The agarose had to be cool before adding 0.75 µL Midori Green (Nippon Genetics, Germany) per 25 mL agarose solution, for DNA staining. The mix is then gently mixed, to avoid bubbles, poured into the cast, and mounted with a comb to provide the formation of the chambers for sample application. After hardening, the running chamber was filled with 1x TAE buffer, and the comb was removed. The Q5 PCR products had to be mixed with 5 x purple gel loading dye (NEB) or 6x TriTrack DNA-loading dye (ThermoFisher). The GoTaq PCR products were already in a suitable buffer for the application on the agarose gel. For the identification of the band size, 3-4 µL GeneRuler DNA Ladder Mix (ThermoFisher) was added to every gel. To perform the electrophoresis an electric tension was set to 120 V for 25-30 min. For the analysis, the gel is removed from the chamber and the bands were made visible under UV light or blue light and documented.

6.5.4. DNA isolation

6.5.4.1. Isolation of genomic DNA

The genomic DNA of *C. vaccinii* was isolated using the “Wizard Genomic DNA Purification Kit” from Promega. The isolation was performed according to the manufacturer's protocol for Gram-negative bacteria.

6.5.4.2. Purification of PCR products

After amplification, the DNA had to be isolated from the PCR reaction mix to make it suitable for ligation or Gibson assembly, to prevent potential interferences. For that purpose, the FastGene Gel/PCR Extraction Kit (Nippon Genetics) were used, according to the producer's manual. In the final step the DNA was eluted with 20-50 µL autoclaved ddH₂O at RT.

6.5.4.3. DNA isolation from agarose gel

If the optimization of the PCR conditions did not prevent the formation of unspecific bands in addition to the desired band, the PCR product had to be purified using agarose gel separation before further use. Therefore, the PCR reaction was completely applied to an agarose gel with large loading pockets, and the agarose gel electrophoresis was performed as described in 6.5.3. The desired band was then excised from the gel with a scalpel and transferred to a fresh reaction tube. The DNA was then isolated from this gel piece using the “FastGene Gel/PCR Extraction Kit” (Nippon Genetics), following the producer’s manual. The DNA was eluted with 20-25 µL autoclaved ddH₂O at RT.

6.5.4.4. Plasmid isolation

The plasmids required for the cloning process were obtained from liquid *E. coli* cultures harboring the desired plasmids. Thus, 5 mL precultures with appropriate antibiotics were prepared from the existing cryocultures and 16-24 h incubated. For the isolation of the plasmids, the “FastGene Plasmid Mini Kit” (Nippon Genetics) was used. Therefore, 4 mL of these precultures were transferred in a new reaction tube and the purification procedure was conducted as described in the operating instructions for high-copy DNA plasmids. For the elution step, 20-50 µL autoclaved ddH₂O was used.

6.5.5. DNA concentration determination

For the quantification of gDNA, the DNA fragments generated *via* PCR and purified plasmid DNA, photometric measurements were performed at 260 nm. The devices used for this were the Eppendorf BioPhotometer kinetic with µCuvette and the FastGene NanoSpec Photometer.

6.5.6. Restriction digest

To insert the GOI into a plasmid *via* ligation or Gibson assembly, suitable overhangs have to be generated *via* restriction enzymes. These type II restriction endonucleases were provided by New England Biolabs and led to the formation of so-called sticky ends, which enable simple combinations of the desired fragments. The composition of a standard digestion mix is shown in **Table 14**.

Table 14: Composition of a standard restriction digestion mix.

Volume [μL]		Component
analytic	preparative	
8	25	DNA (50-100 ng/μL)
1	3	10x CutSmart buffer (NEB)
0.5	1	Restriction enzyme 1
0.5	1	Restriction enzyme 2

After the preparation of the digestion mix, it was incubated at 37 °C for 2 h. If the concentration of the digested product was too low at the end, the incubation time was extended up to 16 h at 4 °C to increase the yield. The analytical digest was then directly analyzed *via* agarose gel electrophoresis (6.5.3). For the preparative approach, the restriction enzymes had to be heat-inactivated (information on correct heat inactivation is provided by the manufacturer). To prevent the restriction enzymes from causing any further interference during the following cloning process, the digestion was purified using the “FastGene Gel/PCR Extraction Kit” (Nippon Genetics) following the instructions for the purification of PCR products. 20 μL autoclaved ddH₂O was used to elute the DNA.

To avoid the re-ligation of the plasmid, shrimp alkaline phosphatase was added to the plasmid digest after 1.5 h of incubation. This results in the dephosphorylation of the 5'-ends of the linearized plasmid.

6.5.7. Ligation

To combine the provided complementary DNA ends of the plasmid and inserts, a T4 ligase (Promega) was used. The standard ligation approach can be obtained from **Table 15**.

Table 15: Composition of a ligation approach (10 μL).

Volume [μL]	Component
8	DNA (Plasmid 1:3 Insert)
1	Reaction buffer (Promega)
1	T4 ligase

The ligation mix was then incubated for 24-72 h at 4 °C and transformed into chemically competent *E. coli* cells.

6.5.8. Gibson assembly

Gibson assembly is a simple and efficient molecular method to combine multiple DNA fragments and insert them into a vector. Therefore, the primer to amplify the fragments needs to be designed, having overlapping sequences of 18-25 bp. The T5 exonucleases in the Gibson reaction mix then provide the overhangs, the complementary ends can aggregate and are subsequently linked by the *Taq* DNA ligase, also contained in the mix.

Table 16: Composition of the Gibson assembly master mix.

Volume [μL]	Component
320	5x ISO buffer
160	Taq DNA ligase (10 U/μL)
20	Phusion polymerase (2 U/μL)
0.64	T5 exonuclease (10 U/μL)
1200	ddH ₂ O

The components were combined on ice and well mixed. Afterwards aliquots on 15 μL were prepared and stored at -20 °C until needed.

To perform a Gibson assembly, one aliquot was thawed on ice and 5 μL of DNA fragments (Plasmid 1:3 Insert) were added. The insert fragments were amplified *via* PCR and purified as described in 6.5.4.2. The plasmid was linearized with restriction enzymes and purified (6.5.6). The reaction took place at 50 °C for 60 min.

6.5.9. Sanger sequencing

To investigate the correctness of the assembled and mutated plasmids, as well as the knockout mutants, DNA samples were sent to Eurofins Genomics Germany GmbH (Ebersberg, Germany) for Sanger sequencing. The plasmids pET28a(+) and pHis8TEV were sequenced using the standard primers T7 and T7 term. For other approaches, the primers were sent separately and can be found in **Table 7**. The results were supplied as FASTA in .txt-format or as chromatogram in .pd4-format and were analyzed with SnapGene 2.3.2 (Eurofins MWG Operon, Ebersberg, Germany).

6.5.10. Generation of *E. coli* expression strains

For *in vitro* investigation of the function of the modules and domain, the *frs* genes were inserted in expression vectors and transformed into an appropriate *E. coli* strain. First, primers had to be designed to amplify the genes of interest *via* Q5 PCR (6.5.2.1). The fragments were then purified as described in 6.5.4.2. A preculture was prepared from a cryogenic culture of an *E. coli* strain harboring the expression plasmid, which encodes for an His-Tag (6-8 histidines) for protein purification. The plasmid was reisolated (6.5.4.4) and digested with restriction enzymes for linearization (6.5.6). For an assembly using a ligation approach, the inserts had to be digested with the same restriction enzymes, to gain complementary sticky ends. For an assembly *via* Gibson approach this step was not necessary. For the combination of the generated vector and insert fragments the ligation (6.5.7) or Gibson assembly (6.5.8) approach was performed. The assembled plasmid was purified if necessary, and heat-shock transformed in an *E. coli* strain optimized for cloning and plasmid storage (6.5.10). After confirming the successful transformation by colony PCR (6.5.2.2), and the validation of the sequence by Sanger sequencing (6.5.9), the plasmid was reisolated and transformed in an *E. coli* expression strain using the heat shock method (6.5.10).

6.5.11. Construction of knockout mutants

6.5.11.1. Preparation of the knockout vector

The knockout vector consisted of four DNA fragments, which were combined by Gibson assembly. The vector used was pEX18Tc (Tet^R,*sacB*), which was linearized through the digestion with *SacI* and *SphI* (NEB; 6.5.6). The FRT cassette (1.8 kbp; *Gem^R Gfp⁺*) was amplified using Q5 PCR using the isolated pPS585 with a concentration of 10-25 ng/μL as template (primer in Table 9). Due to unspecific PCR products, the desired fragment had to be purified *via* agarose gel electrophoresis. The upstream (up) and downstream (dn) regions of the gene to be knocked out were also amplified *via* Q5 PCR. It is important to ensure that the size of the flanking regions is adapted to the size of the GOI. For each kbp of the GOI, the upstream and downstream regions have to be at least 500 bp long. In addition, the primers of the regions have to harbor overlapping sequences with the FRT cassette and the digested pEX18Tc plasmid (6.5.6). If the fragments could be gained in sufficient quantity and purity, they were applied in a plasmid to insert a ratio of 1:3, into one aliquot of a Gibson assembly mix. After the Gibson assembly success was verified, the assembled plasmid was introduced into chemically competent cells of *E. coli* NEB10β using heat-shock transformation. The resulting colonies were first checked for fluorescence and tested by colony

screening to determine whether they contained the desired plasmid, which was then reisolated of the positive clones and analyzed *via* Sanger sequencing.

6.5.11.2. Multistep conjugation

To replace the GOI with the FRT cassette, *via* homologous recombination, the plasmid has to be introduced into *C. vaccinii*. Therefore a triparental conjugation with the strains *E. coli* NEB10 β pEX18Tc: Δ *frsG8_{TE}*:FRT, the helper strain *E. coli* ET12567 pUB307, and *C. vaccinii* MWU205 were performed. To gain linearized FR-Core the knockout mutant *C. vaccinii* MWU205 Δ *vioA* Δ *frsA* was used. All strains were grown in 25 mL LB medium with appropriate antibiotics at 30 °C, 220 rpm, or overnight at 16 °C, 200 rpm to an OD600 of 0.4-0.6. Then 2 mL were harvested and washed two times with 1 mL LB medium to remove the antibiotics by centrifugation at 4000 rpm for 2 min. 0.6 mL of the *C. vaccinii* strain and 0.2 mL of each helper strains were combined, mixed gently and plated as a 'puddle' on LB agar without antibiotics. The plate was incubated at 30 °C for 20-24 h. The resulting cell layer was scraped off and resuspended in 1 mL LB medium without salt and 100 μ L of this suspension was plated on selective LB agar without salt (Amp200, Gem30, Suc 15%). The plates were again incubated at 22-25 °C until the first colonies became visible. If the insertion of the FRT cassette in the right place was confirmed using colony PCR and sequencing, the biparental conjugation was performed with the *C. vaccinii* MWU205 Δ *vioA* Δ *frsA* Δ *frsG8_{TE}*:FRT and the helper strain *E. coli* DH10 β pTA-Mob + pMLBAD_{tet}::Flp/SacB. So the strains were grown to an OD600 of 0.4-0.6 at 30 °C in LB medium with appropriate antibiotics. 2 mL of each strain were harvested and washed twice with 1 mL LB (4000 rpm, 2 min). 0.6 mL of the *C. vaccinii* strain was combined with 0.2 mL helper strain, plated as 'puddle' on an LB agar plate without antibiotics, and incubated at 30 °C for 20-24 h. The cell layer was then scraped off and resuspended in 1 mL LB medium and grown at 30 °C until colonies became visible. These colonies were then screened with appropriate primers to investigate if the FRT cassette was removed by Flp-based recombination resulting in an 86 bp scar where the GOI used to be. To remove the pMLBAD_{tet}::Flp/SacB from the *C. vaccinii* MWU205 Δ *vioA* Δ *frsA* Δ *frsG8_{TE}* strain, 300 mL of an overnight preculture, were washed twice with 1 mL LB without salt (4000 rpm, 2 min) and the pellet was resolved again in 1 mL. After preparing a 1:100000 dilution, of which 100 μ L were plated on LB agar without salts, the plate was incubated at 30 °C for 2-5 days. The resulting colonies were checked for antibiotic susceptibility. The colonies that only grow on LB agar plates with Amp200 were cultivated, the gDNA was isolated and the sequence was checked by Sanger sequencing. If the desired 86 bp scar could be verified to be in the right place in the genome, the *C. vaccinii* knockout strain was used for fermentation to investigate the biosynthetic product of the modified *frs* gene cluster.

6.6. Protein analysis

6.6.1. Protein overexpression

6.6.1.1. Overexpression in TB medium

Protein overexpression was conducted in 100-1000 mL TB medium. In advance, upscaling experiments were performed to investigate the relation of culture size to protein yield, as these are rarely linearly related to each other. So, depending on the needed amount of proteins the volumes of media were prepared in a baffled flask, to provide a sufficient oxygen supply during cultivation. The appropriate antibiotics were added, and the media was inoculated to 1% with an overnight culture, which was cultivated at 37 °C and 200 rpm for at least 16 h. The expression cultures were incubated at 37 °C and 200 rpm until they reached an optical density at 600 nm (OD₆₀₀) of 0.9-1.2. Subsequently, the cultures were induced with Isopropyl- β -thiogalactoside (IPTG) to a final concentration of 0.4 mM. IPTG binds the lactose repressor instead of lactose, leading to a derepression, whereby the expression of GOI is activated. This takes place at 16 °C, 180 rpm for 16 h.

6.6.1.2. Overexpression in M9 medium with a spin-down approach

The protein analysis using NMR must be expressed in M9 medium so that later on several components can be exchanged for labeled substances to gain a labeled POI. To avoid the slow growth, which normally occurs while cultivation in M9, a spin-down approach was applied. Therefore, 1 L TB cultures were prepared as described earlier (inoculated to 1%, cultivation at 37 °C and 200 rpm). When an OD₆₀₀ of 0.9-1.2 was reached the cells were harvested (15 min, 4000 rpm, 10 °C) under sterile conditions. The cell pellets were washed twice with 10 mL M9 medium (10 min, 4000 rpm, 10 °C), resuspended in 10 mL, and added to 600 mL prepared M9 medium in a baffled flask. The cultures were incubated for 1 h at 37 °C and 200 rpm to recover from the harvesting procedure. The OD₆₀₀ was measured to ensure, that the cells continued growing. If the OD₆₀₀ increased slightly after 1 h of incubation the expression was induced by adding 0.4-1 mM IPTG. The expression parameters were varied to find the best conditions. The optimal conditions were set to 20 °C for 18 h at 180 rpm.

6.6.1.3. M9 expression optimization

To increase protein yield while expression in minimal media, various parameters were tested to determine the optimal conditions.

Different *E. coli* expression strains, like BL21 and LOBSTR, were used to minimize the background consisting of unspecific proteins contaminating the final product and complicating the purification. The incubation temperature was varied from 16-20 °C. The incubation duration was also altered from 16-20 h. The addition of IPTG was adjusted from 0.1-1 mM, depending on the expression strain. To achieve results that were as comparable as possible, all approaches of the modified parameter were tested simultaneously, using the same preculture and the same medium batch for all approaches. Care was also taken during the subsequent purification process to ensure that no deviations occurred between the approaches.

6.6.1.4. Expression in labeled M9 medium

To generate a labeled FrsA TE domain for 2D and 3D NMR measurements, it was expressed in M9 with labeled components. So, ^{15}N -ammonium chloride was used as an altered nitrogen source, resulting in a ^{15}N labeled FrsA_{TE}. To obtain the ^{13}C - ^{15}N -TE domain, glucose was additionally exchanged by ^{13}C -glucose.

To identify specific residues, hence supporting the assignment, three further samples of FrsA_{TE} were produced, in which the vitamin mix was supplemented with labeled and unlabeled amino acids. The composition is shown in **Table 17**.

Table 17: Amido acid composition of the combinatorial labeling approaches

Sample	Labeled amino acids
1	^{15}N -Met, ^{15}N -Tyr, ^{15}N -Lys, ^{13}C - ^{15}N -His, Phenylpyruvate
2	^{13}C - ^{15}N -Met, ^{15}N -Tyr, ^{15}N -Phe, ^{15}N -Arg
3	^{13}C - ^{15}N -Met, ^{15}N -Phe, ^{15}N -Lys, ^{15}N -Cys, Hydroxyphenylpyruvate

Each proteinogenic amino acid not mentioned in the table was added unlabeled. Depending on the stability of the amino acid component, it was added before or after autoclaving. Furthermore, phenylpyruvate and hydroxyphenylpyruvate were used as precursors to avoid tyrosine-phenylalanine scrambling.

The addition of labeled amino acids should allow the identification of arginine, cysteine, histidine, lysine, methionine, tyrosine, and phenylalanine residues in the 2D- ^{15}N -HSQC.

6.6.2. Protein purification *via* nickel-nitrilotriacetic acid (Ni-NTA) affinity chromatography

For immediate use of the expression cultures, the cells were harvested by centrifugation (10000 rpm, 15 min, 4 °C), the supernatant was discarded, and the cell pellet was resuspended in 5 ml per g pellet with chilled lysis buffer **Table 3**, until no cell clumps remained. For later use, the cell pellets were frozen undissolved at -20 °C and stored for several weeks.

To avoid protein degradation the samples were kept at ice for the whole purification process. The cells were lysed by sonification (10 x 10 impulses at 20% impulse magnitude). Before the next centrifugation step (10000 rpm, 15 min, 4 °C) 100 µL of this lysate was taken as an induced sample for SDS-PAGE analysis. The supernatant was transferred in a new falcon tube and mixed with 0.5-4 mL nickel-nitrilotriacetic acid (Ni-NTA), which was washed 3 times with lysis buffer (7500 rpm, 30 s, RT) in advance. The mixture is incubated for 1-2 h on ice at 200 rpm and subsequently purified over a polypropylene column from Quiagen with 4 mL Wash I buffer, and 4 mL Wash II buffer. The desired protein was then eluted by the addition of 2.5 mL elution buffer. After proper mixing, 30 µL of every purification sample was taken, to be analyzed *via* SDS-PAGE **(6.6.8)**.

6.6.2.1. Purification optimization

To optimize the purification conditions, the single steps of the purification were analyzed *via* SDS-PAGE. In order to increase the cell lysis efficiency, the ideal stirring during ultrasonic lysis was ensured. Furthermore, care was taken in later purifications, that the cell pellet after expression culture harvest, undergoes at least one freezing and thawing cycle prior to ultrasonic lysis, as this also leads to cell disruption. A combination of both lysis methods led to an increase in POI yield. The amounts of the used Wash I and Wash II buffers were adjusted to gain a high purity of the end product of the purified protein. Additionally, NaCl, DTT, and protease inhibitors were added to the purification buffer to prevent digestion, dimerization, and degradation of the protein of interest. NaCl supports the protein stability and was tested in 25-150 mM final concentrations. DTT was applied in concentrations of 5 up to 20 mM. It should provide a reducing environment, which prevents dimerization of the protein. Protease inhibitors (PIs) avoid the digestion of the protein during the purification process. The PIs were added in a standard concentration (1x) according to the manufacturer's instructions. The optimal amounts of the additives were investigated and determined through several optimization approaches and were set to 5 mM DTT, 25 mM NaCl, and 1x PIs.

6.6.3. Desalting and buffer system exchange

If the purified protein should be used for an *in vitro* assay subsequently, it had to be rebuffered into the target assay buffer to remove imidazole to avoid potential protein degradation. Therefore, a PD-10 column (GE Healthcare) was used as stated in the gravity-based protocol. Another desalting possibility is described in the next section.

6.6.4. Concentration measurement and concentration

Protein concentration was determined using the method established by Gill and Hippel.^[248] In this approach, the molar extinction coefficient at $\lambda = 280$ nm was calculated based on the amount of tryptophanes, tyrosines and disulfide bonds within the respective protein. Absorbance measurements were conducted using a μ Cuvette and the Eppendorf Biophotometer kinetic instrument and later with a FastGene NanoSpec Photometer. Employing the calculated extinction coefficient and a blank measurement with an appropriate buffer in advance, the photometer determined the protein concentration in mg/mL.

If the concentration was insufficient for the planned *in vitro* assay or crystallization approaches, the proteins were concentrated using a Vivaspin® 500 filter (Sartorius) with a molecular weight cut-off of 10 kDa or 30 kDa, ensuring that the cut-off is less than half the molecular weight of the desired protein. After the protein solution was added to the Vivaspin® tube, it was repeatedly centrifuged (7500 rpm, 5 min, 4 °C) until the desired concentration was reached. In between, the protein was rinsed from the filter by repeated pipetting up and down to avoid clogging.

The Vivaspin® tube is also suitable for desalting large amounts of protein during the concentration step. Therefore, the concentrated protein is diluted in the desired buffer at least 3 times with 4 mL of the target buffer and centrifuged (7500 rpm, 5 min, 4 °C) until it attains the final concentration.

6.6.5. Protein storage

To store the proteins for later use Glycerol was added to a final concentration of 15% to the protein solutions and mixed thoroughly. The mixture was aliquoted in small amounts (250-500 μ L), flash-frozen in liquid nitrogen, and immediately stored at -80 °C. Before usage, the aliquots were thawed at 20 °C and 300 rpm and the proteins were reactivated by the addition of 5 mM DTT. In advance, it should be investigated if the proteins withstand the procedure without loss of natural folding and activity.

6.6.6. His-Tag removal utilizing TEV-protease

The protein purification was performed using Ni-NTA affinity chromatography. Crucial for this method is the His-tag with which the protein has to be expressed. However, this tag has to be removed for the crystallization experiments to avoid interferences during protein folding. Therefore, the gene was cloned into a pHis8TEV vector, leading to the synthesis of the protein with a cleavage site for the Tobacco Etch Virus (TEV) protease between the His-tag and the protein of interest (POI). So, the TEV protease can cut off the His-tag, releasing the natural folded protein.

The TEV protease was expressed, purified, and desalted separately, as described in the upper sections (6.6.6). Then the protease was mixed with the His-TEV-POI in a mg/mL ratio of 1:5 and incubated for 16 h at 4 °C. Subsequently the POI was purified *via* reverse-Ni-NTA-purification. Therefore 0.5-2 mL Ni-NTA were added to the digest and incubated for 1 h on ice and 200 rpm. The suspension was then filtered over a polypropylene column and washed with 4-8 mL Wash I buffer and 4 mL Wash II buffer and the undigested was eluted with 2.5 mL elution buffer. The POI without His-tag should be present in the flowthrough sample. Samples were taken during purification to analyze the purification process and to adjust the parameters.

6.6.7. Size exclusion-Fast protein liquid chromatography

To gain a sufficient quality for the crystallization process the isolated and digested POI was further purified *via* Size Exclusion Fast Protein Liquid Chromatography (SEC-FPLC). The separation principle is based on the highly porous matrix of the column, forming pores of different sizes. Small molecules can easily enter the pores and are thus retained more than molecules with a higher molecular weight, which can enter all pores. Thus, the proteins can be separated depending on their apparent size. The used column was the Superdex 200 increase 10/300 GL (Cytiva), which was connected to the Bio-Rad NGC system (F10 pump, Mixer module, sample inject valve module, column switching valve module (10 mL), single UV/conductivity detector module, BioFrac fraction collector, outlet valve module).

Before the purification runs, the system was equilibrated with filtered and degassed FPLC buffer (Table 3). The protein was concentrated to 10-20 mg/mL and centrifugated (13000 rpm, 10 min, 4 °C), to avoid loading degraded protein to the system. 200-500 µL of the protein mixture was injected over a 1 mL syringe with an interposed 0.22 µm filter, preventing undesirable solid particles from being brought into the system. In addition, it is important to ensure that no bubbles get onto the column, to retain the separation efficiency of the column. The flow was set to 1 ml/min and the fraction volume was adjusted to 0.3 mL. To process the whole batch, several runs were needed. All fractions, containing the POI, were analyzed *via* SDS-PAGE, pooled, and then

concentrated to a proper concentration for crystallization. FPLC run analysis was performed with the ChromLab Software.

6.6.8. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

For the analysis of protein identity and purity, discontinuous sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed. Therefore, the gel was prepared as shown in **Table 1**. The separating gel was prepared, poured into the gel cassette, and overlaid with isopropyl to avoid dehydration and bubble formation. After polymerization of the separating gel, the isopropyl was removed and the stacking gel was layered on the separating gel. A comb was inserted into the unpolymerized stacking gel, forming the sample wells. Wrapped into wet towels and an airtight plastic bag, the gel can be stored at 4 °C for several weeks.

Table 18: Composition SDS-gels

Component	12% 17-65 kDa		15% 15-45 kDa	
	Separating gel	Stacking gel	Separating gel	Stacking gel
Bis-Tris acrylamide	4 mL	0.51 mL	2.5 mL	0.34 mL
H ₂ O	3.4 mL	2.04 mL	1.15 mL	1.36 mL
Buffer	2.5 mL	0.375 mL	1.25 mL	0.25 mL
10% SDS	100 µL	30 µL	50 µL	20 µL
10% APS	100 µL	30 µL	50 µL	20 µL
TEMED	4 µL	3 µL	5 µL	2 µL

The obtained samples for SDS-PAGE analysis had a volume of 30 µL and were added with NuPAGE sample buffer (4x) and 4 µL NuPAGE reducing agent and heated at 70 °C for 10 min to cause denaturation.

The gel cassette was clipped into the vertical XCell SureLock Mini-Cell Electrophoresis System (ThermoFisher). The running chamber was filled with approx. 500 mL Running buffer (**Table 3**), the comb was removed, and the wells were loaded with 3-15 µL sample. To analyze the molecular weight of the proteins, a protein standard (Unstained Protein Standard (10-200 kDa); Color Prestained Protein Standard (10-150 kDa)) was used on each gel. The lid was attached and connected to a power supply. For the discontinuous method, 90 V was applied for the first 30 min. Afterwards, the voltage was increased up to 140 V for 90 min. When gel electrophoresis was

completed, the gel cassette was detached, and the gel was carefully removed and shaken in ddH₂O at 200 rpm for 10 min. For visualization, the ddH₂O was exchanged for staining solution (**Table 4**) and the gel was shaken for another 30 min. The staining solution was discarded, and the gel was washed repeatedly with ddH₂O until the staining solution was completely removed. Subsequently, the gel was immersed in destaining solution (**Table 4**) and shaken overnight between two layers of cellulose paper to absorb the Coomassie blue. The gels were documented with a camera under visible light.

6.6.9. Proteolytic digest

The POI was digested with the proteases α -chymotrypsin, trypsin, elastase, and papain, solved in TE domain buffer. The tested ratios were 1:10, 1:100 and 1:1000 (protease:protein). To achieve a stabilizing effect on the important core structures, the substrate FR-Core was added in some approaches in the same molarity as FrsA_{TE}. The approaches were incubated for 1/2/24 h at 30/37 °C, and then analyzed *via* SDS-PAGE.

6.6.10. Protein sequencing

After finding a promising band in the SDS-PAGE analysis of the proteolytic digestion of FrsA_{TE} it was necessary to determine the sequence coverage, of the resulting protein fragment. Therefore, the proteolysis product was applied on a preparative SDS-gel. The preparation followed the same procedure as described in **6.6.8**, except for the utilization of a comb forming bigger chambers for the application of the whole proteolytic approach. Furthermore, it was ensured to provide contaminations, which included the production and use of new solutions. After segregating the protein fragments, the bands of interest were excised using a scalpel under sterile conditions, transferred to a 2 mL reaction tube, and freeze-dried. The gel pieces were then sent to the Mass Spectrometry Core Facility Analytical Proteomics of the University of Bonn. There, the protein fragments were further processed by performing an in-gel digestion using trypsin. The obtained fragments are then measured by ESI-LC/MS and the results are aligned with the amino acid sequence of FrsA_{TE} to analyze the sequence coverage.

6.6.11. Analytical size-exclusion chromatography

Another method to investigate the protein size is size-exclusion chromatography. Compared to the SDS-PAGE, this method has the advantage that the proteins do not have to be denaturated through the addition of SDS, negating covalent interactions. Due to experimental data, a homodimeric

structure was suggested for some purified proteins. *Via* SEC it was possible to determine structure and the dimerization and to identify the optimal buffer condition to gain the protein mostly present as a monomer.

For this purpose, the HPLC system was connected to a TSKgel G3000SWXL (7.8 x 300 mm; Tosoh Bioscience). The system had a diode array detector with which the proteins were detected at 280 nm. The flow was set to 1 mL/min and the injection volume was 10 μ L. Before the analysis of the protein, a protein standard mix was measured, consisting of 0.5 mg/mL thyroglobulin (660 kDa), 1.0 mg/mL ovalbumin (43 kDa), and 1.5 mg/mL ribonuclease a (13.7 kDa). Based on this data, the peaks of the sample can be assigned to the monomer or the dimer, caused by the linear regression, which was calculated using GraphPad Prism.

To determine the size of the protein of interest, it was purified as described in **Chapter 6.6.2-6.6.4** and buffered into an SEC interaction buffer. The concentration was set to 1 mg/mL and 20 μ L was injected into the chromatography system.

6.6.12. Crystallography and X-ray analysis

FrsA_{T+TE} was adjusted to a concentration of 1 mM in TE domain buffer (25 mM HEPES, 50 mM NaCl, pH 7.5) while FR-Core was solubilized in methanol at a concentration of 10 mM. By mixing, a solution with a final concentration of 0.83 mM (FrsA_{T+TE}) and 1.66 mM (FR-Core) was achieved. Crystallization assays were set up in a 2-drop 96-MRC sitting drop plate. Using a Crystal Gryphon (Art Robbins Instruments, LLC, USA) 200 nL of protein and 200 nL of mother liquor of the crystallization screens condition were pipetted onto the plate. Crystals were grown at 20 °C in a Rock Imager 1000 (Formulatrix, US) and imaged automatically. Crystal growth was observed after 12 hours in the condition containing 20% (w/v) PEG 3350, 0.2 M Na₂SO₄ and 0.1 M BIS-Tris at pH 7.5. Crystals were harvested after 21 days, soaked in mother liquor supplemented with 35% glycerol, flash frozen, and stored in liquid nitrogen.

Diffraction data were collected at the DESY (Deutsches Elektronen Synchrotron, Hamburg) at beamline P13 ($\lambda = 0.976251$ Å) using an EIGER 16M detector.^[249] The dataset was processed with EDNAproc^[250] which itself also invokes XDS^[251], AIMLESS^[252], POINTLESS^[253] and CCP4 programs^[254]. The structure was solved by molecular replacement, using the Phaser module of Phenix^[255,256] using an AlphaFold^[257] model as a search structure. Phenix.refine^[258] was used to refine the structure, COOT^[259] to manually build and MolProbity^[260] for validating the structure.

Table 19: X-ray crystallography data collection and refinement statistics

Space group	P 6 5
Cell dimensions (Å)	
<i>a</i> , <i>b</i> , <i>c</i>	178.7 178.7 54.1
α , β , γ	90, 90, 120
Resolution range (Å)	44.68 - 2.03 (2.08 - 2.03)
Wilson B-factor (Å ²)	49.76
CC*	1 (0.662)
<i>I</i> / σ (<i>I</i>)	12.81 (0.70)
Completeness (%)	99.91 (99.69)
Redundancy	20.5 (20.5)
Unique reflections	64200 (4578)
<i>R</i> _{work}	0.20 (0.42)
<i>R</i> _{free}	0.23 (0.45)
<i>R</i> _{merge}	0.13 (4.5)
Average B factor (Å ²)	63.21
Protein	63.01
Ligands	87.11
Solvent	63.35
Non-hydrogen atoms	5378
Protein	5149
Ligands	34
solvent	195
RMS (bonds) (Å)	0.011
RMS (angles) (°)	1.12
Ramachandran favored (%)	98.14
Ramachandran allowed (%)	1.86
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.00
Clashscore	3.67

For the soaking approaches, the FrsA_{T+TE} crystals were first grown as described earlier, using the LigandFriendly (Molecular Dimension) screen. The obtained crystals were then fished and incubated in a solution containing 50 mM FR-Core, 15% (v/v) mother liquor, and 35% (v/v) glycerol for 30 s. Afterwards, the crystals were flash-frozen and stored in liquid nitrogen.

6.6.13. CryoEM: sample preparation, data collection and processing

FrsA_{T+TE} was prepared at a concentration of 37 mg/mL and 50 μ L were subsequently loaded on a Superdex 200 3.2/300 on a Agilent HPLC 1260 infinity II, equilibrated with TE domain buffer (25 mM HEPES, 50 mM NaCl, pH 7.5). The sample was manually collected according to the elution profile, the concentration was determined by UV absorption at 280 nm (NanoDrop 2000, Thermofisher Scientific), and a 5-fold excess of FR-Core was added. Afterwards, the sample was diluted to 0.5 mg/mL of FrsA_{T+TE} and used for Cryo-EM experiments.

Grid preparation was performed using a Vitrobot (Thermofisher Scientific) at 4 °C, 95% humidity, -5 blotting force, and 5 seconds blotting time, on Quantifoil QF 1.2/1.3 Cu200 grids, subsequently stored in liquid nitrogen. Data collection was done with a Talos Arctica Cryo-TEM (Thermofisher Scientific) operated at 200 kV, equipped with a BioQuantum K3 camera. Images were taken with -3 μ m defocus, 62 e/ $\text{\AA}^2$ dose, 4.00 s exposure over 75 frames. Processing was done with CryoSparc (Punjani et al. 2017). After a patch motion correction and patch contrast transfer function (CTF), exposures were curated and 7327 of them were used downstream for an automatic blob picker job. This resulted in 9.8 million particles which were used in several rounds of 2D classification. The quality of the 2D classes was not sufficient for further processing.

6.6.14. NMR structural analysis

The samples were prepared by Christian Bonifer from the working group of Prof. Dr. Glaubitz (Institute for Biophysical Chemistry, the Centre for Biomolecular Magnetic Resonance of the Goethe University of Frankfurt), who also evaluated the acquired data. The NMR measurements were conducted by Dr. Frank Löhner from the working group of Prof. Dr. Volker Dötsch (Institute for Biophysical Chemistry, the Centre for Biomolecular Magnetic Resonance of the Goethe University of Frankfurt). The NMR spectra were recorded on Bruker spectrometers (600: 3D HN(CO)CA/3D HN(CO)CACB; 700MHz: 2D ^{15}N -BEST-TROSY/3D HNCA; 800 MHz: 3D HNCACB/ 3DHNCO/3D NH(CA)CO/Combinatorial labeling measurements (2D ^{15}N -BEST-TROSY + HNCO)) equipped with a $^1\text{H}\{^{13}\text{C}/^{15}\text{N}\}$ TCI CryoProbe at 303 K. For the initial 2D NMR experiments a fingerprint [$^1\text{H},^{15}\text{N}$]-BEST-TROSY (Transverse Relaxation Optimized Spectroscopy) was recorded to assess the ^{15}N -labeling efficiency and to identify potential changes in the chemical shifts within the backbone amides caused by substrate binding. The two-dimensional fingerprint experiment provided a rapid overview of the spectral quality and served as a preliminary scrutiny for protein integrity. For a complete backbone resonance assignment, three-dimensional experiments were performed. This included the recording of 3D HNCA-, HN(CO)CA-, HNCACB-, and HN(CO)CACB-spectra from a ^{13}C - ^{15}N -FrsA_{TE} sample. Since large proteins (>150 residues) often display broad and overlapping signals, caused by slow tumbling, HNCO- and HN(CA)CO-spectra were recorded additionally. To

enhance the efficiency and accuracy of the assignment process, combinatorial labeling strategies were implemented. This involved selective labeling of specific amino acids (Chapter XXX) and a 2D CO-filtered ^{15}N , ^1H -BEST-TROSY experiment. The software CCPN (Collaborative Computing Project for NMR) was used for the manual assignment. To support the manual process and provide additional validation, an automated assignment approach was conducted by NMRtist. This is a cloud computing service, for automated protein NMR spectra analysis. By employing a combination of these techniques, a comprehensive and reliable assignment of backbone and side chain resonances was achieved. This assignment was then used to identify the residues that undergo changes in chemical shifts during substrate binding, allowing assumptions about the interaction between FrsA_{TE} and its substrates. Therefore, $[\text{}^1\text{H}, \text{}^{15}\text{N}]$ -BEST-TROSY of ^{15}N -FrsA_{TE} + FR-Core, ^{15}N -FrsA_{TE} + side chain, and ^{15}N -FrsA_{TE} + FR-Core + sidechain were recorded.

6.7. *In vitro* assays

6.7.1. Side chain assembly assay

The sidechain assembly assay was performed with the purified proteins FrsA_{CAT}, FrsB, and FrsH to gain the sidechain *N*-Pp-Hle or a modified sidechain as substrate for a subsequent TE domain assay. The reactions were conducted in a one-pot approach of 500 μL sidechain assembly assay buffer (SCA-Buffer). First, 10 μM FrsH was mixed with 10 μM methyl viologen, and 1 mM NADH, and incubated on ice for 5 min. Then 10 μM FrsA_{CAT}, 1 mM ATP, 1 mM L-Leu, 100 μM propionyl-CoA, and 100 μM DTT were added, and the assay was conducted at 22 $^{\circ}\text{C}$ for 10-60 min. To analyze the hydroxylation or the successful formation of the sidechain, the reaction was stopped by adding 50 μL 100% TCA (w/v) and incubation at 30 min on ice. The precipitated proteins were then centrifuged (5000 rpm, 5 min, RT) and the pellets were washed twice with assay buffer to remove the TCA. The thioester bond should be removed by the addition of 200 μL 0.1 M KOH. The mixture was shaken at 300 rpm at 70 $^{\circ}\text{C}$ for 20 min and lyophilized subsequently. The extract was dissolved in 100 μL MeOH and centrifuged for 10 min at 13000 rpm to avoid entrainment of particles through the transfer in the HPLC vials. The measurement was performed on an LC-MS (Waters e2695 Separation Module, coupled to an ESI, Acquity QDa, and 2998 PDA detector) with an EC 250/4.6 NUCLEODUR 10-5 HILIC-column (Macherey-Nagel). After the injection of 5 μL sample, 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) was set over 20 min runtime with a flow of 0.7 mL/min. The program Empower3 was used for data acquisition evaluation.

6.7.2. TE domain assay

To examine the thioesterase activity an assay was established based on the assay developed by Ray *et al.*, which was used to investigate the TE domain activity of the salinamide biosynthesis.^[149] The required components are the involved enzymes, FrsA_{CAT} and FrsA_{TE}, the sidechain, *N*-Pp-Hle, and the nucleophile, the cyclized peptide backbone FR-Core. The enzymes were purified as described in 6.6.2. FR-Core was extracted and purified from a *C. vaccinii* Δ voA Δ frsA strain (6.8.1/6.8.2). The sidechain could be synthesized *in vitro*, through the sidechain assembly assay, or a synthetically produced sidechain was used.^[147]

The sidechain assembly assay was performed as described in 6.7.1. Instead of stopping the assay after 60 min of incubation, the TE domain assay was started by the addition of 5 μ M FrsA_{TE} and 100 μ M FR-Core and then incubated at 22 °C for 5-16 h. The reaction was stopped by the addition of 750 μ L dichloromethane. The mixture was inverted carefully and after the separation of the phases, the organic phase was transferred to a new tube. The extraction step was repeated 3 times. The organic phase was evaporated and dissolved in methanol for LC-MS analysis. For the negative control, the enzymes FrsA_{CAT}, FrsH, and FrsA_{TE} were heat-inactivated at 80 °C for 20 min before the addition of the substrates.

6.7.2.1. TE domain assay with synthetic side chain

To avoid the multi-enzymatic *in vitro* reaction, which is necessary for the formation of the sidechain, it was synthesized. In the natural biosynthesis of FR, the sidechain is assembled, while it is bound through a thioester on the thiolation domain of FrsA. *N*-acetylcysteamine is capable of mimicking this thiolation domain-bound phosphopantetheine moiety. So, the *N*-propionyl-3-hydroxyleucine-*N*-acetylcysteamine was synthesized by Dr. Küppers and Dr. Hermes, for a simplified TE domain assay.^[147]

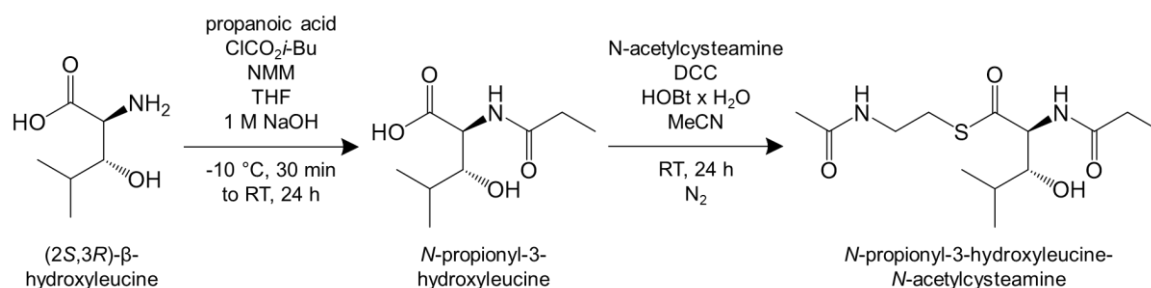


Figure 63: Synthesis of the side chain mimic *N*-Pp-Hle-SNAC thioester. (CICO₂*t*-Bu = isobutyl chloroformate, NMM = 4-methylmorpholine, THF = tetrahydrofuran, DCC = *N,N'*-Dicyclohexylcarbodiimide, HOBT = 1-Hydroxybenzotriazole, MeCN = acetonitrile).^[147]

The assay was performed in 500 μ L C domain buffer with a final concentration of 5 μ M FrsA_{TE}. The sidechain and FR-Core are dissolved in MeOH to a concentration of 10 μ M and added to a final concentration of 100 μ M each. The assay mix was incubated at 22 °C for 5-16 h and extracted as described in **Chapter 6.7.2**.

6.7.3. Substrate exchanges in side chain assembly assay

For the investigation of substrate specificity of FrsA, the side chain assembly as well as the TE domain assay were modified by exchanging the substrates. **Figure 53** shows by which the respective substrates were replaced.

Table 20: Listing of the substrates exchanged in the modified assays.

FrsA domain	Natural substrate	Exchanged by
A	L-Leucine	D-Leucine, Isoleucine, Valine
C	N-Propionyl-CoA	Acetyl-CoA, Butyryl-CoA, Isobutyryl-CoA, Malonyl-CoA, Octanoyl-CoA
TE	FR	linearized FR-Core
	N-propionyl-OH-leucine	modified side chain

When exchanging the substrates during the side chain assembly assay, it was ensured that the final concentration of the component still corresponds to that of the standard assay protocol. In the TE domain assay, however, it was not possible to determine the concentration of the modified side chain and the linearized FR core. Therefore, the maximum possible concentration was used in each case.

6.7.4. Generation of substrate-bound FrsA TE domain

To generate a substrate-bound FrsA TE domain, a modified combination of the side chain assembly assay and the TE domain assay was performed. Therefore, the FrsA module was separated into the FrsA_{C+A} and FrsA_{T+TE}. The used FrsA_{T+TE} construct was already produced for the crystallization attempts. FrsA_{C+A} was specifically prepared for this approach by a standard cloning process *via* ligation (6.5.7). It was coexpressed with FrsB, and freshly purified for the assay as well as the enzyme FrsH, all harboring an His-tag. FrsA_{T+TE} was purified one day before the assay, to perform a TEV protease-induced removal of the His-tag overnight (6.6.6). Right before the assay

conduction, FrsA_{T+TE} was purified *via* reverse affinity chromatography and rebuffed in SCA assay buffer. The one-pot assay was set up in a volume of 2 mL, nevertheless, following the standard SCA assay protocol: 10 μ M FrsH was mixed with 10 μ M methyl viologen, and 1 mM NADH, and incubated on ice for 5 min. Then 10 μ M FrsA_{C+A}, 10 μ M FrsA_{T+TE}, 1 mM ATP, 1 mM L-Leu, 100 μ M propionyl-CoA, and 100 μ M DTT were added, and the assay was incubated for 60 min at 22 °C. Afterwards, 1 mL Ni-NTA was added to the assay and the mix was incubated on ice at 200 rpm. Through the purification using a polypropylene column, the enzymes with His-Tag should be removed from the assay mix, resulting in the remaining FrsA_{T+TE}. To verify if the side chain was attached to the T domain, a subsequent TE domain assay was performed by adding 100 μ M FR-Core to the assay. It was incubated for 16 h at 22 °C. The reaction was stopped by the addition of 750 μ L dichloromethane. The mixture was inverted carefully and after the separation of the phases, the organic phase was transferred to a new tube. The extraction step was repeated 3 times. The organic phase was evaporated and dissolved in methanol for LC-MS analysis. For the negative control, the enzymes FrsA_{C+A}, FrsA_{T+TE}, and FrsH were heat-inactivated at 80 °C for 20 min before the addition of the substrates.

6.8. *In vivo* investigations of *C. vaccinii*

6.8.1. Fermentation and extraction

For the fermentation of *C. vaccinii* MWU205 and the *C. vaccinii* knockout mutants 40-4000 ml LB medium with 50 μ g/mL carbenicillin was inoculated with preculture to 1%. The cultivation took place at 25 °C and 200 rpm for 48 h. The extraction took place, through the addition of 40-4000 mL *n*-butanol and further 24 h at 25 °C and 200 rpm. The extraction mixture was centrifuged (4000 rpm, 20 min, RT), and the solvent phase was transferred in a round bottom flask and evaporated (40 °C, 25 mbar). For HPLC-MS measurement the crude extract was resolved in methanol to a concentration of 1 mg/mL. To isolate the FR and its derivatives the crude extract was processed as described in 6.8.1.

6.8.2. Isolation of FR, FR-Core, and linearized FR-Core

For the isolation of FR, FR Core, and other FR derivatives the crude material of the butanoic extract, was fractionated through a Grace Reveleris X2 flash chromatography system with Reveleris C18 flash column (220 g, pore-size of 40 μ m). The flow was set to 65 mL/min and a gradient was applied, which is shown in **Table 21**.

Table 21: Flash chromatography gradient for FR, FR-Core, and LFRC isolation.

Time [min]	Solvent composition
13	H ₂ O 50/50 MeOH
14	H ₂ O 30/70 MeOH
26	H ₂ O 25/75 MeOH
14	H ₂ O 20/80 MeOH
26	H ₂ O 15/85 MeOH
11	H ₂ O 0/100 MeOH

The device was equipped with an evaporative light scattering (ELSD) and UV/Vis detection and the measured signals led to the conclusion that FR Core eluted from the column after 70 min.

The fraction of interest was then processed further *via* HPLC with a semi-preparative Macherey-Nagel Nucleodur C18 column (250 x 8 mm, 5 µm). The flow was set to 2 mL/min with an isocratic elution of H₂O 19/81 MeOH solvent composition. The purification was performed with a Waters HPLC system, consisting of a 600E pump, a 996 PDA detector, and a 717 plus autosampler, or on a Waters Breeze HPLC system equipped with a 1525µ dual pump, a 2998 photodiode array detector, and a Rheodyne 7725i injection system.

6.9. Bioinformatic methods

6.9.1. AntiSMASH

AntiSMASH is an open-source online tool, allowing the identification, annotation, and analysis of secondary metabolite biosynthesis gene clusters in bacterial genomes.^[261] DNA files can be uploaded in .gbk, .embl, or .fasta format and antiSMASH unifies a large number of *in silico* tools, enabling an in-depth analysis of the BGC. The results can be evaluated on the browser surface or downloaded in .gbk-format. antiSMASH is powered by several open-source tools: NCBI BLAST+, HMMer 3, Muscle 3, Glimmer 3, FastTree, TreeGraph 2, Indigo-depict, PySVG and JQuery SVG.

6.9.2. I-TASSER

The Iterative Threading ASSEmbly Refinement (I-TASSER) server is an online platform for protein structure and function prediction based on their amino acid sequences. To generate a prediction

the amino acid sequence can be submitted in single letter code. Then the server tries to retrieve template proteins of similar fold or secondary structure from PDB library by LOMETS, a locally installed meta-threading approach. In the second step, the protein fragments showing a similar fold are excised and reassembled into a full-length model. Missing parts are built by *ab initio* modeling. In the final step, the fragment assembly is repeated with the newly generated structure, to remove steric clashes and refine global topology of the cluster centroids. The results are ranked based on the confidence score of the I-TASSER structural models, structural similarity between models and templates, and the sequence identity in the structurally aligned regions.^[262–264]

I-TASSER was used for the first structural prediction of FrsA and FrsA_{TE}.

6.9.3. ExpasyProtParam

To generate the first information about a protein of interest, the Expasy ProtParam tool was used, which is a free online tool. It computes several physicochemical properties which can be deduced from the protein sequence.^[265] Therefore, the amino acid sequence can be submitted, or the UniProtKB accession number or ID can be used. The computed parameters include molecular weight, amino acid composition, extinction coefficient, and the theoretical pI. Thereby, it is noteworthy that the ProtParam tool is not able to specify post-translational modifications. The ExpasyProtParam tool was used for the calculation of the molecular weight and the extinction coefficient for each protein used in this work.

6.9.4. SIM (Expasy Alignment tool)

SIM is a free available alignment tool for protein sequences. It finds a user-defined number of best non-intersecting alignments between two protein sequences, which can be submitted in single letter code, or by designation of the UniProtKB accession number or ID. For each alignment, the default method was used: 20 computed alignments, the Gap open penalty was set to 12, the gap extension penalty was set to 4, and the comparison matrix was BLOSUM62. SIM was used for the alignment of FrsA_{TE} and FrsG8_{TE}.

6.9.5. AlphaFold

ColabFold is an open-source software, offering protein structure prediction by combining MMSeqs2 with AlphaFold2.^[266,267] ColabFold is a fast and easy method to use a Jupyter Notebook inside Google Collaboratory, enabling access to AlphaFold2 from any computer. Through the input of the protein sequence, AlphaFold2 builds a multiple sequence alignment (MSA), generated by

several protein sequence databases. Divers and deep MSAs lead to the identification of similarities and co-evolutions, resulting in profound structural predictions. Using AlphaFold2 in the ColabFold interface, the structures of FrsA, FrsA_{T+TE}, and FrsG8_{TE} were predicted.

6.9.6. SwissDock

Molecular docking is a computational method used to predict the preferred orientation of one molecule (ligand) to another (receptor) when bound to form a stable complex. For this study, SwissDock^[268–271], a web-based platform, was used to investigate potential interactions between FrsA_{T+TE} and FR-Core. SwissDock offers two docking algorithms: Attracting Cavities 2.0 and AutoDock Vina (Vina)^[272,273]. Given the computational efficiency and accuracy, the Vina algorithm was selected for this research. The Vina scoring function incorporates a potential energy term representing van der Waals interactions, an empirical hydrogen bond contribution, a hydrophobic term, and a penalty for the loss of conformational entropy. Vina calculates intermolecular interactions using trilinear interpolation of precomputed grid maps centered on the target molecule. Post-docking energy minimization is performed using the target structure as a rigid reference. For the analysis, the elucidated structure of FrsA_{T+TE} was applied, and the search box comprised the active center and the lid structure (search box center: -35/56/-4; search box size: 22/28/18).

6.9.7. Dali server

The elucidated protein structure of FrsA_{T+TE} was prepared for Dali analysis by removing water molecules and heteroatoms. The processed structure was then submitted to the Dali server for heuristic PDB search. This search algorithm compares the query structure against all structures in the Protein Data Bank (PDB).

The Dali server generates a list of structurally similar proteins ranked by Z-score, a measure of the statistical significance of the alignment. For this study, the top 83 Dali hits with a Z-score greater than 22 were considered for further analysis. Structural alignments and superposition of the query protein with the identified homologs were visualized using PyMOL.

6.9.8. Visualization with PyMOL

PyMOL is a commercially available three-dimensional graphics software, designed for visualization of molecular structures. To display molecules in 3D, desired data files in .pdb and .cif

format are required, containing information about the types and positions of atoms and bonds. PyMOL supports most relevant data formats and allows a rapid interactive visualization and analysis of molecular models, including measuring distances, identifying binding partners and neighboring residues, visualizing surfaces and cavities, and aligning multiple structures. Furthermore, PyMOL enables the modification of the 3D structure by various design options of individual atoms, amino acids, secondary structures, and molecular surfaces in shapes and colors. By zooming, rotating, and moving, the 3D structure can be examined and displayed from any direction. Finally, the produced image is rendered, producing a 3D scene, and can be saved in .png format. PyMOL has also the capability to create a 3D animation of the displayed molecules. Due to the fact, that PyMOL is an open-source software, it offers the addition of numerous plugins, supporting a detailed analysis of molecules.

All protein illustrations in this work were generated using PyMOL version 2.5.2. For each protein, all ligands as well as the crystal water were removed. The colors were adjusted, and the alignments were obtained with the “Alignment” plugin. Secondary structures were depicted in “cartoon” style, whereby important amino acids were highlighted in altered colors and shown in the “stick” style.

7. References

- [1] O. Mosunova, J. C. Navarro-Muñoz, J. Collemare, *Encyclopedia of Mycology: Volume 1,2* **2021**, 2, 458–476.
- [2] J. O'Brien, G. D. Wright, *Curr Opin Biotechnol* **2011**, 22, 552–558.
- [3] V. L. Challinor, H. B. Bode, *Ann N Y Acad Sci* **2015**, 1354, 82–97.
- [4] S. Sanchez, A. L. Demain, *Comprehensive Biotechnology, Second Edition* **2011**, 1, 155–167.
- [5] J. V. Ekman, A. Kruglov, M. A. Andersson, R. Mikkola, M. Raulio, M. Salkinoja-Salonen, *Microbiology (N Y)* **2012**, 158, 1106–1116.
- [6] A. Oladipo, V. Enwemiwe, O. Ejeromedoghene, A. Adebayo, O. Ogunyemi, F. Fu, *Metabolites* **2022**, 12.
- [7] A. Fleming, *Br J Exp Pathol* **1929**, 10, 226–236.
- [8] A. Schatz, E. Bugie, S. Waksman, *Proc Soc Exp Biol Med* **1944**, 55, 66–69.
- [9] E. K. Perry, L. A. Meirelles, D. K. Newman, *Nat Rev Microbiol* **2022**, 20, 129–142.
- [10] M. Fujioka, S. Koda, Y. Morimoto, K. Biemann, *J. Org. Chem.* **1988**, 53, 2820–25.
- [11] L. Ma, W. Li, H. Wang, X. Kuang, Q. Li, Y. Wang, P. Xie, K. Koike, *J Pharm Biomed Anal* **2015**, 102, 400–408.
- [12] E. G. De Mejía, M. V. Ramírez-Mares, *Toxicol Mech Methods* **2011**, 21, 667–674.
- [13] H. Kobayashi, E. De Mejía, *J Ethnopharmacol* **2005**, 96, 347–354.
- [14] Z. Jia, K. Koike, T. Ohmoto, M. Ni, *Phytochemistry* **1994**, 37, 1389–1396.
- [15] M. Crüsemann, R. Reher, I. Schamari, A. O. Brachmann, T. Ohbayashi, M. Kuschak, D. Malfacini, A. Seidinger, M. Pinto-Carbó, R. Richarz, T. Reuter, S. Kehraus, A. Hallab, M. Attwood, H. B. Schiöth, P. Mergaert, Y. Kikuchi, T. F. Schäberle, E. Kostenis, D. Wenzel, C. E. Müller, J. Piel, A. Carlier, L. Eberl, G. M. König, *Angewandte Chemie - International Edition* **2018**, 57, 836–840.
- [16] A. Miyamae, M. Fujioka, S. Koda, Y. Morimoto, *J Chem Soc Perkin 1* **1989**, 873–878.
- [17] I. M. Miller, *Adv Bot Res* **1990**, 17, 163–234.
- [18] B. Lemaire, E. Smets, S. Dessein, *Res Microbiol* **2011**, 162, 528–534.
- [19] A. S. Hauser, M. M. Attwood, M. Rask-Andersen, H. B. Schiöth, D. E. Gloriam, *Nat Rev Drug Discov* **2017**, 16, 829–842.
- [20] A. Carlier, L. Fehr, M. Pinto-carbó, T. Schäberle, R. Reher, S. Dessein, G. König, L. Eberl, **2016**, 18, 2507–2522.
- [21] 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, *Nat Commun* **2021**, 12, 1–33.
- [22] S. Soby, S. Gadagkar, C. Contreras, F. Caruso, *Int J Syst Evol Microbiol* **2013**, 63, 1840–46.

-
- [23] K. Vöing, A. Harrison, S. D. Soby, *Genome Announc* **2015**, *3*, 1–2.
- [24] D. Pistorius, K. Buntin, E. Weber, E. Richard, C. Bouquet, S. Wollbrett, H. Regenass, V. Peón, M. Böhm, R. Kessler, T. Gempeler, A. Haberkorn, L. Wimmer, C. Lanshoeft, J. Davis, D. Hainzl, J. A. D'Alessio, E. Manchado, F. Petersen, *Chemistry - A European Journal* **2022**, *28*.
- [25] D. A. Wirtz, N. Schneberger, S. Klöppel, R. Richarz, M. Geyer, G. M. König, G. Hagelueken, M. Crüsemann, *ACS Chem Biol* **2023**, *18*, 1748–1759.
- [26] S. Klöppel, R. Richarz, D. A. Wirtz, N. Vasenda, G. M. König, M. Crüsemann, *ChemBioChem* **2022**, *23*.
- [27] M. Taniguchi, K. Nagai, N. Arao, T. Kawasaki, T. Saito, Y. Moritani, J. Takasaki, K. Hayashi, S. Fujita, K. I. Suzuki, S. I. Tsukamoto, *Journal of Antibiotics* **2003**, *56*, 358–363.
- [28] C. Hermes, G. M. König, M. Crüsemann, *Nat Prod Rep* **2021**, *38*, 2276–2292.
- [29] T. Hashimoto, H. Suenaga, K. Amagai, J. Hashimoto, I. Kozono, S. Takahashi, K. Shin-ya, *Angewandte Chemie - International Edition* **2024**, 202317805.
- [30] R. Reher, M. Kuschak, N. Heycke, S. Annala, S. Kehraus, H. F. Dai, C. E. Müller, E. Kostenis, G. M. König, M. Crüsemann, *J Nat Prod* **2018**, *81*, 1628–1635.
- [31] R. Reher, T. Kühn, S. Annala, T. Benkel, D. Kaufmann, B. Nubbemeyer, J. P. Odhiambo, P. Heimer, C. A. Bäuml, S. Kehraus, M. Crüsemann, E. Kostenis, D. Tietze, G. M. König, D. Imhof, *ChemMedChem* **2018**, *13*, 1634–1643.
- [32] D. Pistorius, E. Richard, K. Buntin, K. Dresen, S. Wollbrett, E. Weber, A. Haberkorn, E. Manchado, F. Petersen, *Chembiochem* **2024**.
- [33] R. Fredriksson, M. C. Lagerström, L.-G. Lundin, H. B. Schiöth, *Mol Pharmacol* **2003**, *63*, 1256–1272.
- [34] D. Calebiro, Z. Koszegi, *Mol Cell Endocrinol* **2019**, *483*, 24–30.
- [35] D. Hilger, M. Masureel, B. K. Kobilka, *Nat Struct Mol Biol* **2018**, *25*, 4–12.
- [36] J. P. Mahoney, R. K. Sunahara, *Curr Opin Struct Biol* **2016**, *41*, 247–254.
- [37] S. Offermanns, *Biol. Chem* **2005**, *381*, 389–396.
- [38] R. K. Sunahara, P. A. Insel, *Mol Pharmacol* **2016**, *89*, 585–592.
- [39] C. Southan, J. L. Sharman, H. E. Benson, E. Faccenda, A. J. Pawson, S. P. H. Alexander, O. P. Buneman, A. P. Davenport, J. C. McGrath, J. A. Peters, M. Spedding, W. A. Catterall, D. Fabbro, J. A. Davies, *Nucleic Acids Res* **2016**, *44*, D1054–D1068.
- [40] W. M. Oldham, H. E. Hamm, *Nat Rev Mol Cell Biol* **2008**, *9*, 60–71.
- [41] S. R. Sprang, *Biopolymers* **2016**, *105*, 449–462.
- [42] G. Milligan, E. Kostenis, *Br J Pharmacol* **2006**, *147*.
- [43] D. Tietze, D. Kaufmann, A. A. Tietze, A. Voll, R. Reher, G. Ko, F. Hausch, *Journal of Chemical Information and Modeling* **2019**, *59*.
- [44] H. Zhang, A. L. Nielsen, K. Strømgaard, *Med Res Rev* **2020**, *40*, 135–157.
-

-
- [45] T. Flock, A. S. Hauser, N. Lund, D. E. Gloriam, S. Balaji, M. M. Babu, *Nature* **2017**, *545*, 317–322.
- [46] M. I. Simon, M. P. Strathmann, Nara. Gautam, *Nucl Instrum Methods Phys Res Sect B* **1971**, *32*, 1785.
- [47] S. J. Taylor, H. Z. Chae, S. G. Rhee, J. H. Exton, *Nature* **1991**, *350*, 516–518.
- [48] N. Suzuki, N. Hajicek, T. Kozasa, *Neurosignals* **2009**, *17*, 55–70.
- [49] G. Sánchez-Fernández, S. Cabezudo, C. García-Hoz, C. Benincá, A. M. Aragay, F. Mayor, C. Ribas, *Cell Signal* **2014**, *26*, 833–848.
- [50] E. Kostenis, E. M. Pfeil, S. Annala, *Journal of Biological Chemistry* **2020**, *295*, 5206–5215.
- [51] R. Schrage, A. Schmitz, E. Gaffal, S. Annala, S. Kehraus, D. Wenzel, K. Büllersbach, T. Bald, A. Inoue, Y. Shinjo, S. Galandrin, N. Shridhar, M. Hesse, M. Grundmann, N. Merten, T. Charpentier, M. Martz, A. Butcher, T. Slodczyk, S. Armando, M. Effern, Y. Namkung, L. Jenkins, V. Horn, A. Stößel, H. Dargatz, D. Tietze, D. Imhof, C. Galés, C. Drewke, C. Müller, M. Hölzel, G. Milligan, A. Tobin, J. Gomeza, H. Dohlman, J. Sondek, T. Harden, M. Bouvier, S. Laporte, J. Aoki, B. Fleischmann, K. Mohr, G. König, T. Tüting, E. Kostenis, *Nat Commun* **2015**, *14*, 1–7.
- [52] J. Takasaki, T. Saito, M. Taniguchi, T. Kawasaki, Y. Moritani, K. Hayashi, M. Kobori, *Journal of Biological Chemistry* **2004**, *279*, 47438–47445.
- [53] A. Nishimura, K. Kitano, J. Takasaki, M. Taniguchi, N. Mizuno, K. Tago, T. Hakoshima, H. Itoh, *Proc Natl Acad Sci U S A* **2010**, *107*, 13666–13671.
- [54] C. Hermes, G. M. König, M. Crüsemann, *Nat Prod Rep* **2021**, *38*, 2276–2292.
- [55] J. Mühle, M. J. Rodrigues, J. Alenfelder, L. Jürgenliemke, R. Guixà-González, A. Bacchin, F. Andres, M. Hennig, H. Schihada, M. Crüsemann, G. M. König, E. Kostenis, G. Schertler, X. Deupi, *Preprint* **2024**, DOI 10.1101/2023.10.24.563737.
- [56] J. H. Voss, M. Crüsemann, C. R. O. Bartling, S. Kehraus, A. Inoue, G. M. König, K. Strømgaard, C. E. Müller, *iScience* **2023**, *26*.
- [57] M. Kuschak, J. G. Schlegel, M. Schneider, S. Kehraus, J. H. Voss, A. Seidinger, M. Matthey, D. Wenzel, B. K. Fleischmann, G. M. König, C. E. Müller, *Front Chem* **2020**, *8*.
- [58] J. H. Voss, J. Nagel, M. Rafehi, R. Guixà-González, D. Malfacini, J. Patt, S. Kehraus, A. Inoue, G. M. König, E. Kostenis, X. Deupi, V. Namasivayam, C. E. Müller, *Pharmacol Res* **2021**, *173*.
- [59] K. Zaima, J. Deguchi, Y. Matsuno, T. Kaneda, Y. Hirasawa, H. Morita, *J Nat Med* **2013**, *67*, 196–201.
- [60] V. Inamdar, A. Patel, B. K. Manne, C. Dangelmaier, S. P. Kunapuli, *Platelets* **2015**, *26*, 771–778.
- [61] M. M. Meleka, A. J. Edwards, J. Xia, S. A. Dahlen, I. Mohanty, M. Medcalf, S. Aggarwal, K. D. Moeller, O. V. Mortensen, P. Osei-Owusu, *Pharmacol Res* **2019**, *141*, 264–275.
- [62] R. Carr, C. Koziol-White, J. Zhang, H. Lam, S. S. An, G. G. Tall, R. A. Panettieri, J. L. Benovic, *Mol Pharmacol* **2016**, *89*, 94–104.
- [63] M. Matthey, R. Roberts, A. Seidinger, A. Simon, R. Schröder, M. Kuschak, S. Annala, G. M. König, C. E. Müller, I. P. Hall, E. Kostenis, B. K. Fleischmann, D. Wenzel, *Sci Transl Med* **2017**, *9*, 1–12.
-

-
- [64] E. S. Rantala, M. M. Hernberg, S. Piperno-Neumann, H. E. Grossniklaus, T. T. Kivelä, *Prog Retin Eye Res* **2022**, *90*.
- [65] M. Wespiser, E. Neidhardt, S. Negrier, *Cancer Treat Rev* **2023**, *119*, 102599.
- [66] C. Chattopadhyay, D. Won Kim, D. Gombos, J. M. Oba, Y. Qin, M. Williams, B. Esmaeli, E. Grimm, J. Wargo, S. Woodman, S. Patel, C. Author, *Cancer* **2016**, *122*, 2299–2312.
- [67] C. D. Van Raamsdonk, V. Bezrookove, G. Green, J. Bauer, L. Gaugler, J. M. O'Brien, E. M. Simpson, G. S. Barsh, B. C. Bastian, *Nature* **2009**, *457*, 599–602.
- [68] C. D. Van Raamsdonk, K. G. Griewank, M. B. Crosby, M. C. Garrido, S. Vemula, T. Wiesner, A. C. Obenaus, W. Wackernagel, G. Green, N. Bouvier, M. M. Sozen, G. Baimukanova, R. Roy, A. Heguy, I. Dolgalev, R. Khanin, K. Busam, M. R. Speicher, J. O'Brien, B. C. Bastian, *New England Journal of Medicine* **2010**, *363*, 2191–2199.
- [69] M. D. Onken, S. E. Noda, K. M. Kaltenbronn, C. Frankfater, C. M. Makepeace, N. Fetting, K. D. Piggott, P. L. Custer, J. E. Ippolito, K. J. Blumer, *Journal of Biological Chemistry* **2022**, *298*, 101495.
- [70] T. D. Hitchman, G. Bayshtok, E. Ceraudo, A. R. Moore, C. Lee, R. Jia, N. Wang, M. R. Pachai, A. N. Shoushtari, J. H. Francis, Y. Guan, J. Chen, M. T. Chang, B. S. Taylor, T. P. Sakmar, T. Huber, P. Chi, Y. Chen, *Clinical Cancer Research* **2021**, *27*, 1476–1490.
- [71] M. D. Onken, C. M. Makepeace, K. M. Kaltenbronn, J. Choi, L. Hernandez-Aya, K. N. Weilbaecher, K. D. Piggott, P. Kumar Rao, C. M. Yuede, A. J. Dixon, P. Osei-Owusu, J. A. Cooper, K. J. Blumer, *Journal of Biological Chemistry* **2021**, *296*, 100403.
- [72] S. Annala, X. Feng, N. Shridhar, F. Eryilmaz, J. Patt, J. H. Yang, E. M. Pfeil, R. D. Cervantes-Villagrana, A. Inoue, F. Häberlein, T. Slodczyk, R. Reher, S. Kehraus, S. Monteleone, R. Schrage, N. Heycke, U. Rick, S. Engel, A. Pfeifer, P. Kolb, G. König, M. Bünemann, T. Tüting, J. Vázquez-Prado, J. S. Gutkind, E. Gaffal, E. Kostenis, *Sci Signal* **2019**, *12*.
- [73] K. Klepac, A. Kilić, T. Gnad, L. M. Brown, B. Herrmann, A. Wilderman, A. Balkow, A. Glöde, K. Simon, M. E. Lidell, M. J. Betz, S. Enerbäck, J. Wess, M. Freichel, M. Blüher, G. König, E. Kostenis, P. A. Insel, A. Pfeifer, *Nat Commun* **2016**, *7*.
- [74] G. Seravalle, G. Grassi, *Pharmacol Res* **2017**, *122*, 1–7.
- [75] *Cold-Activated Brown Adipose Tissue in Healthy Men A Bs Tr Ac t*, **2009**.
- [76] J. H. Voss, *ACS Pharmacol Transl Sci* **2023**, *6*, 1790–1800.
- [77] C. H. Chau, P. S. Steeg, W. D. Figg, *Lancet* **2019**, *394*, 793–804.
- [78] S. Coats, M. Williams, B. Kebble, R. Dixit, L. Tseng, N. S. Yao, D. A. Tice, J. C. Soria, *Clinical Cancer Research* **2019**, *25*, 5441–5448.
- [79] S. Baah, M. Laws, K. M. Rahman, *Molecules* **2021**, *Vol. 26*, Page 2943 **2021**, *26*, 2943.
- [80] R. D. Süßmuth, A. Mainz, *Angewandte Chemie - International Edition* **2017**, *56*, 3770–3821.
- [81] D. A. Alonzo, T. M. Schmeing, *Protein Science* **2020**, *29*, 2316–2347.
- [82] B. Mach, E. Reich, E. L. Tatum, *J. Biol. Chem* **1963**, *50*, 175–181.
-

-
- [83] T. L. Berg, L. O. Froholm, S. G. Laland, *Biochem J* **1965**, *96*, 43–52.
- [84] R. Roskoski, W. Gevers, H. Kleinkauf, F. Lipmann, *Biochemistry* **1970**, *9*, 4839–4845.
- [85] R. Roskoski, H. Kleinkauf, W. Gevers, F. Lipmann, *Biochemistry* **1970**, *9*, 4846–4851.
- [86] W. Gevert, H. Kleinkauf, T. And, F. Lipmann, *Proc Natl Acad Sci U S A* **1969**, *63*, 1335–1342.
- [87] F. Ishikawa, S. Nakamura, I. Nakanishi, G. Tanabe, *J Pept Sci* **2024**, *30*.
- [88] T. Stein, J. Vater, V. Kruft, A. Otto, B. Wittmann-Liebold, P. Franke, M. Panico, R. McDowell, H. R. Morris, *J Biol Chem* **1996**, *271*, 15428–15435.
- [89] A. M. Gulick, *Curr Opin Chem Biol* **2016**, *Dec*, 89–96.
- [90] J. M. Reimer, A. S. Haque, M. J. Tarry, T. M. Schmeing, *Curr Opin Struct Biol* **2018**, *49*, 104–113.
- [91] M. A. Marahiel, *Journal of Peptide Science* **2009**, *15*, 799–807.
- [92] D. Schwarzer, R. Finking, M. A. Marahiel, *Nat Prod Rep* **2003**, *20*, 275–287.
- [93] M. E. Horsman, T. P. A. Hari, C. N. Boddy, *Nat Prod Rep* **2016**, *33*, 183–202.
- [94] A. M. Gulick, *ACS Chem Biol* **2009**, *4*, 811–827.
- [95] G. L. Challis, J. Ravel, C. A. Townsend, *Chem Biol* **2000**, *7*, 211–224.
- [96] E. Conti, T. Stachelhaus, M. A. Marahiel, P. Brick, *EMBO J* **1997**, *16*, 4174–4183.
- [97] T. Stachelhaus, H. D. Mootz, M. A. Marahiel, *Chem Biol* **1999**, *6*, 493–505.
- [98] S. Caboche, V. Leclère, M. Pupin, G. Kuchero, P. Jacques, *J Bacteriol* **2010**, *192*, 5143–5150.
- [99] G. H. Hur, C. R. Vickery, M. D. Burkart, **2012**, 1074–1098.
- [100] K. Eppelmann, T. Stachelhaus, M. A. Marahiel, *Biochemistry* **2002**, *41*, 9718–9726.
- [101] R. H. Baltz, *J Ind Microbiol Biotechnol* **2017**, *44*, 573–588.
- [102] D. A. Herbst, B. Boll, G. Zocher, T. Stehle, L. Heide, *Journal of Biological Chemistry* **2013**, *288*, 1991–2003.
- [103] L. E. Quadri, J. Sello, T. A. Keating, P. H. Weinreb, C. T. Walsh, *Chem Biol* **1998**, *5*, 631–45.
- [104] R. D. Zwahlen, C. Pohl, R. A. L. Bovenberg, A. J. M. Driessen, *ACS Synth Biol* **2019**, *8*, 1776–1787.
- [105] B. Boll, T. Taubitz, L. Heide, *Journal of Biological Chemistry* **2011**, *286*, 36281–36290.
- [106] M. Wolpert, B. Gust, B. Kammerer, L. Heide, *Microbiology (N Y)* **2007**, *153*, 1413–1423.
- [107] S. Lautru, D. Oves-Costales, J. L. Pernodet, G. L. Challis, *Microbiology (N Y)* **2007**, *153*, 1405–1412.
- [108] E. A. Felnagle, J. J. Barkei, H. Park, A. M. Podevels, M. D. McMahon, D. W. Drott, M. G. Thomas, *Biochemistry* **2010**, *49*, 8815–8817.
- [109] W. Zhang, J. R. Heemstra, C. T. Walsh, H. J. Imker, *Biochemistry* **2010**, *49*, 9946–9947.
- [110] K. Bloudoff, D. A. Alonzo, T. M. Schmeing, *Cell Chem Biol* **2016**, *23*, 331–339.
-

-
- [111] K. Bloudoff, T. M. Schmeing, *Biochim Biophys Acta Proteins Proteom* **2017**, 1865, 1587–1604.
- [112] S. L. Clugston, S. A. Sieber, M. A. Marahiel, C. T. Walsh, *Biochemistry* **2003**, 42, 12095–12104.
- [113] C. Rausch, I. Hoof, T. Weber, W. Wohlleben, D. H. Huson, *BMC Evol Biol* **2007**, 7.
- [114] De Crécy-Lagard V, Marlière P, Saurin W., *C R Acad Sci III*. **1995**, 318, 927–936.
- [115] D. Konzl, A. Klensl, K. Schörgendorfer, M. A. Marahiell, *Chem Biol* **1997**, 4, 927–937.
- [116] K. Haslinger, M. Peschke, C. Brieke, E. Maximowitsch, M. J. Cryle, *Nature* **2015**, 521, 105–109.
- [117] V. Ulrich, M. Peschke, C. Brieke, M. J. Cryle, *Mol Biosyst* **2016**, 12, 2992–3004.
- [118] N. Ziemert, S. Podell, K. Penn, J. H. Badger, E. Allen, P. R. Jensen, *PLoS One* **2012**, 7.
- [119] L. Du, C. Sanchez, M. Chen, D. J. Edwards, B. Shen, *Chem Biol* **2000**, 7, 623–642.
- [120] T. A. Keating, C. G. Marshall, C. T. Walsh, A. E. Keating, *Nat Struct Biol* **2002**, 9, 522–526.
- [121] M. A. Marahiel, T. Stachelhaus, H. D. Mootz, *Chem Rev* **1997**, 97, 2651–2674.
- [122] S. A. Samel, G. Schoenafinger, T. A. Knappe, M. A. Marahiel, L. O. Essen, *Structure* **2007**, 15, 781–792.
- [123] T. Weber, R. Baumgartner, C. Renner, M. A. Marahiel, T. A. Holak, *Structure* **2000**, 8, 407–4018.
- [124] A. Koglin, M. R. Mofid, F. Löhr, B. Schäfer, V. V Rogov, M.-M. Blum, T. Mittag, M. A. Marahiel, F. Bernhard, V. Dötsch, *Science (1979)* **2006**, 312, 273–276.
- [125] J. Beld, E. C. Sonnenschein, C. R. Vickery, J. P. Noel, M. D. Burkart, *Nat Prod Rep* **2014**, 31, 61–108.
- [126] T. Kittilä, M. Schoppet, M. Cryle, *ChemBioChem* **2016**, 17, 576–584.
- [127] P. Tufar, S. Rahighi, F. I. Kraas, D. K. Kirchner, F. Löhr, E. Henrich, J. Köpke, I. Dikic, P. Güntert, M. A. Marahiel, V. Dötsch, *Chem Biol* **2014**, 21, 552–562.
- [128] E. J. Drake, B. R. Miller, C. Shi, J. T. Tarrasch, J. A. Sundlov, C. Leigh Allen, G. Skinotis, C. C. Aldrich, A. M. Gulick, *Nature* **2016**, 529, 235–238.
- [129] W. H. Chen, K. Li, N. S. Guntaka, S. D. Bruner, *ACS Chem Biol* **2016**, 11, 2293–2303.
- [130] Y. Liu, T. Zheng, S. D. Bruner, *Chem Biol* **2011**, 18, 1482–1488.
- [131] D. P. Frueh, H. Arthanari, A. Koglin, D. A. Vosburg, A. E. Bennett, C. T. Walsh, G. Wagner, *Nature* **2008**, 454, 903–906.
- [132] L. Hedstrom, *Chem Rev* **2002**, 102, 4501–4523.
- [133] M. Kotowska, K. Pawlik, *Appl Microbiol Biotechnol* **2014**, 98, 7735–7746.
- [134] D. Schwarzer, H. D. Mootz, M. A. Marahiel, *Chem Biol* **2001**, 8, 997–1010.
- [135] L. Du, L. Lou, *Nat Prod Rep* **2010**, 27, 255–278.
- [136] R. F. Little, C. Hertweck, *Nat Prod Rep* **2022**, 39, 163–205.
- [137] A. T. Keatinge-Clay, *Nat Prod Rep* **2012**, 29, 1050–1073.
-

-
- [138] R. M. Kohli, C. T. Walsh, *Chemical Communications* **2003**, 3, 297–307.
- [139] F. Kopp, M. A. Marahiel, *Nat Prod Rep* **2007**, 24, 735–749.
- [140] S. Bruner, T. Weber, R. Kohli, D. Schwarzer, M. Marahiel, C. Walsh, M. Stubbs, *Structure* **2002**, 10, 301–310.
- [141] S. A. Samel, P. Czodrowski, L. O. Essen, *Acta Crystallogr D Biol Crystallogr* **2014**, 70, 1442–1452.
- [142] K. J. Labby, S. G. Watsula, S. Garneau-Tsodikova, *Nat Prod Rep* **2015**, 32, 641–653.
- [143] M. Z. Ansari, J. Sharma, R. S. Gokhale, D. Mohanty, *BMC Bioinformatics* **2008**, 9.
- [144] S. Wang, Q. Fang, Z. Lu, Y. Gao, L. Trembleau, R. Ebel, J. H. Andersen, C. Philips, S. Law, H. Deng, *Angewandte Chemie International Edition* **2021**, 60, 3229–3237.
- [145] N. M. Gaudelli, D. H. Long, C. A. Townsend, *Nature* **2015**, 520, 383.
- [146] V. V. Phelan, Y. Du, J. A. McLean, B. O. Bachmann, *Chem Biol* **2009**, 16, 473–478.
- [147] C. Hermes, Investigations into Side Chain Assembly and Attachment during Biosynthesis of the G Protein Inhibitor FR900359, **2020**.
- [148] D. Malfacini, J. Patt, S. Annala, K. Harpsøe, F. Eryilmaz, R. Reher, M. Crüsemann, W. Hanke, H. Zhang, D. Tietze, D. E. Gloriam, H. Bräuner-Osborne, K. Strømgaard, G. M. König, A. Inoue, J. Gomeza, E. Kostenis, *Journal of Biological Chemistry* **2019**, 294, 5747–5758.
- [149] L. Ray, K. Yamanaka, B. S. Moore, *Angewandte Chemie - International Edition* **2016**, 55, 364–367.
- [150] J. J. May, T. M. Wendrich, M. A. Marahiel, *Journal of Biological Chemistry* **2001**, 276, 7209–7217.
- [151] 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, *Angewandte Chemie - International Edition* **2021**, 60, 13579–13586.
- [152] H. Chen, C. T. Walsh, *Chem Biol* **2001**, 8, 301–312.
- [153] S. Uhlmann, R. D. Süßmuth, M. J. Cryle, *ACS Chem Biol* **2013**, 8, 2586–2596.
- [154] T. M. Makris, C. J. Knoot, C. M. Wilmot, J. D. Lipscomb, *Biochemistry* **2013**, 52, 6662–6671.
- [155] T. Izoré, M. J. Cryle, *Nat Prod Rep* **2018**, 35, 1120–1139.
- [156] X. F. Xiong, H. Zhang, C. R. Underwood, K. Harpsøe, T. Gardella, M. Wöldike, M. Mannstadt, D. Gloriam, H. Bräuner-Osborne, K. Strømgaard, *Nat Chem* **2016**, 8, 1035–1041.
- [157] J. Zhang, X. Li, *Biotechnol Lett* **2018**, 40, 165–171.
- [158] P. Minárik, N. Tomášková, M. Kollárová, M. Antalík, *Gen Physiol Biophys* **2002**, 21, 257–65.
- [159] S. Wright, M. Kish, R. Viola, *J Biol Chem* **2000**, 275, 31689–31694.
- [160] P. Lavermicocca, F. Valerio, A. Evidente, S. Lazzaroni, A. Corsetti, M. Gobbetti, *Appl Environ Microbiol* **2000**, 66, 4084–4090.
-

-
- [161] X. Li, B. Jiang, B. Pan, W. Mu, T. Zhang, *J Agric Food Chem* **2008**, *56*, 2392–2399.
- [162] L. Sellés Vidal, C. L. Kelly, P. M. Mordaka, J. T. Heap, *Biochim Biophys Acta Proteins Proteom* **2018**, *1866*, 327–347.
- [163] D. A. Wirtz, Biochemical and Structural Characterization of β -Hydroxylases in Non-Ribosomal Peptide Biosynthesis, **2021**.
- [164] C. M. Heckmann, F. Paradisi, *ChemCatChem* **2020**, *12*, 6082–6102.
- [165] A. McPherson, J. A. Gavira, *Acta Crystallogr F Struct Biol Commun* **2014**, *70*, 2–20.
- [166] M. Till, A. Robson, M. J. Byrne, A. V. Nair, S. A. Kolek, P. D. Shaw Stewart, P. R. Race, *J Vis Exp* **2013**.
- [167] A. D'Arcy, T. Bergfors, S. W. Cowan-Jacob, M. Marsh, *Acta Crystallogr F Struct Biol Commun* **2014**, *70*, 1117–1126.
- [168] A. D'Arcy, F. Villard, M. Marsh, *Acta Crystallogr D Biol Crystallogr* **2007**, *63*, 550–554.
- [169] G. E. Dale, C. Oefner, A. D'Arcy, *J Struct Biol* **2003**, *142*, 88–97.
- [170] T. Izoré, Y. T. Candace Ho, J. A. Kaczmarek, A. Gavrilidou, K. H. Chow, D. L. Steer, R. J. A. Goode, R. B. Schittenhelm, J. Tailhades, M. Tosin, G. L. Challis, E. H. Krenske, N. Ziemert, C. J. Jackson, M. J. Cryle, *Nat Commun* **2021**, *12*, 1–14.
- [171] S. Sogabe, M. Masubuchi, K. Sakata, T. A. Fukami, K. Morikami, Y. Shiratori, H. Ebiiike, K. Kawasaki, Y. Aoki, N. Shimma, A. D'Arcy, F. K. Winkler, D. W. Banner, T. Ohtsuka, *Chem Biol* **2002**, *9*, 1119–1128.
- [172] V. M. Leppänen, C. V. Parast, K. K. Wong, J. W. Kozarich, A. Goldman, *Acta Crystallogr D Biol Crystallogr* **1999**, *55*, 531–533.
- [173] H. M. Holden, G. Wesenberg, D. A. Raynes, D. J. Hartshorne, V. Guerriero, I. Rayment, *Acta Crystallogr D Biol Crystallogr* **1996**, *52*, 1153–1160.
- [174] B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane, C. Khosla, *Science* **2001**, *291*, 1790–1792.
- [175] A. Shirai, A. Matsuyama, Y. Yashiroda, A. Hashimoto, Y. Kawamura, R. Arai, Y. Komatsu, S. Horinouchi, M. Yoshida, *J Biol Chem* **2008**, *283*, 10745–10752.
- [176] D. P. Fewer, M. Metsä-Ketelä, *FEBS J* **2020**, *287*, 1429–1449.
- [177] M. G. Chevette, K. Gutiérrez-García, N. Selem-Mojica, C. Aguilar-Martínez, A. Yañez-Olvera, H. E. Ramos-Aboites, P. A. Hoskisson, F. Barona-Gómez, *Nat Prod Rep* **2020**, *37*, 566–599.
- [178] S. Costa, A. Almeida, A. Castro, L. Domingues, *Front Microbiol* **2014**, *5*, 1–20.
- [179] R. B. Kapust, D. S. Waugh, *Protein Sci* **1999**, *8*, 1668–1674.
- [180] J. D. Fox, R. B. Kapust, D. S. Waugh, *Protein Sci* **2001**, *10*, 622–630.
- [181] H. Nikaido, *FEBS Lett* **1994**, *346*, 55–58.
- [182] K. D. Pryor, B. Leiting, *Protein Expr Purif* **1997**, *10*, 309–319.
-

-
- [183] D. C. Prasher, V. K. Eckenrode, W. W. Ward, F. G. Prendergast, M. J. Cormier, *Gene* **1992**, *111*, 229–233.
- [184] C. M, T. Y, E. G, W. WW, P. DC, *Science* **1994**, *263*, 1766–1767.
- [185] R. Y. Tsien, *Annu Rev Biochem* **1998**, *67*, 509–544.
- [186] T. Nagai, K. Ibata, E. S. Park, M. Kubota, K. Mikoshiba, A. Miyawaki, *Nat Biotechnol* **2002**, *20*, 87–90.
- [187] M. A. Rizzo, G. H. Springer, B. Granada, D. W. Piston, *Nat Biotechnol* **2004**, *22*, 445–449.
- [188] G. J. Kremers, J. Goedhart, E. B. Van Munster, T. W. J. Gadella, *Biochemistry* **2006**, *45*, 6570–6580.
- [189] A. Rath, M. Glibowicka, V. G. Nadeau, G. Chen, C. M. Deber, *Proc Natl Acad Sci U S A* **2009**, *106*, 1760–1765.
- [190] E. R. Geertsma, N. A. B. Nik Mahmood, G. K. Schuurman-Wolters, B. Poolman, *Nat Protoc* **2008**, *3*, 256–266.
- [191] J. M. Reimer, M. Eivaskhani, I. Harb, A. Guarné, M. Weigt, T. Martin Schmeing, *Science (1979)* **2019**, *366*.
- [192] M. J. Tarry, A. S. Haque, K. H. Bui, T. M. Schmeing, *Structure* **2017**, *25*, 783–793.e4.
- [193] N. J. Marianayagam, M. Sunde, J. M. Matthews, *Trends Biochem Sci* **2004**, *29*, 618–625.
- [194] M. A. Dessau, Y. Modis, *J Vis Exp* **2011**.
- [195] B. R. Miller, E. J. Drake, C. Shi, C. C. Aldrich, A. M. Gulick, *J Biol Chem* **2016**, *291*, 22559.
- [196] J. T. Mindrebo, C. M. Narthey, Y. Seto, M. D. Burkart, J. P. Noel, *Curr Opin Struct Biol* **2016**, *41*, 233–246.
- [197] M. L. Adrover-Castellano, J. J. Schmidt, D. H. Sherman, *ChemCatChem* **2021**, *13*, 2095–2116.
- [198] K. D. Patel, F. B. d’Andrea, N. M. Gaudelli, A. R. Buller, C. A. Townsend, A. M. Gulick, *Nat Commun* **2019**, *10*.
- [199] N. Huguenin-Dezot, D. A. Alonzo, G. W. Heberlig, M. Mahesh, D. P. Nguyen, M. H. Dornan, C. N. Boddy, T. M. Schmeing, J. W. Chin, *Nature* **2019**, *565*, 112.
- [200] S. A. Samel, B. Wagner, M. A. Marahiel, L. O. Essen, *J Mol Biol* **2006**, *359*, 876–889.
- [201] J. Yu, J. Song, C. Chi, T. Liu, T. Geng, Z. Cai, W. Dong, C. Shi, X. Ma, Z. Zhang, X. Ma, B. Xing, H. Jin, L. Zhang, S. Dong, D. Yang, M. Ma, *ACS Catal* **2021**, *11*, 11733–11741.
- [202] D. F. Kreitler, E. M. Gemmell, J. E. Schaffer, T. A. Wencewicz, A. M. Gulick, *Nature Communications* **2019 10:1** **2019**, *10*, 1–13.
- [203] K. D. Patel, R. A. Oliver, M. S. Lichstrahl, R. Li, C. A. Townsend, A. M. Gulick, *bioRxiv* **2024**.
- [204] A. E. Fraley, C. L. Dieterich, M. F. J. Mabesoone, H. A. Minas, R. A. Meoded, F. Hemmerling, J. Piel, *Angewandte Chemie - International Edition* **2022**, *61*.
-

-
- [205] A. Koglin, F. Löhr, F. Bernhard, V. V. Rogov, D. P. Frueh, E. R. Strieter, M. R. Mofid, P. Güntert, G. Wagner, C. T. Walsh, M. A. Marahiel, V. Dötsch, *Nature* **2008**, *454*, 907–911.
- [206] P. Argyropoulos, F. Bergeret, C. Pardin, J. M. Reimer, A. Pinto, C. N. Boddy, T. M. Schmeing, *Biochim Biophys Acta Gen Subj* **2016**, *1860*, 486–497.
- [207] D. L. Akey, J. D. Kittendorf, J. W. Giraldez, R. A. Fecik, D. H. Sherman, J. L. Smith, *Nat Chem Biol* **2006**, *2*, 537–542.
- [208] J. C. Corpuz, J. O. Sanlley, M. D. Burkart, *Synth Syst Biotechnol* **2022**, *7*, 677–688.
- [209] A. M. Gulick, C. C. Aldrich, *Nat Prod Rep* **2018**, *35*, 1156–1184.
- [210] B. Wienen-Schmidt, M. Oebbeke, K. Ngo, A. Heine, G. Klebe, *ChemMedChem* **2021**, *16*, 292.
- [211] K. Murata, M. Wolf, *Biochim Biophys Acta Gen Subj* **2018**, *1862*, 324–334.
- [212] D. J, A. M, C. JJ, H. JC, L. J, M. AW, S. P, *Q Rev Biophys* **1988**, *21*, 129–228.
- [213] H. R. Saibil, *Mol Cell* **2022**, *82*, 274–284.
- [214] X. Mu, C. Gillman, C. Nguyen, T. Gonen, *Annu Rev Biochem* **2021**, *90*, 431–450.
- [215] R. Castells-Graells, K. Meador, M. A. Arbing, M. R. Sawaya, M. Gee, D. Cascio, E. Gleave, J. Debreczeni, J. Breed, K. Leopold, A. Patel, D. Jahagirdar, B. Lyons, S. Subramaniam, C. Phillips, T. O. Yeates, *Proc Natl Acad Sci U S A* **2023**, *120*.
- [216] I. R. Kleckner, M. P. Foster, *Biochim Biophys Acta Proteins Proteom* **2011**, *1814*, 942–968.
- [217] Y. Hu, K. Cheng, L. He, X. Zhang, B. Jiang, L. Jiang, C. Li, G. Wang, Y. Yang, M. Liu, *Anal Chem* **2021**, *93*, 1866–1879.
- [218] K. R. Andersen, N. C. Leska, T. U. Schwartz, *Proteins* **2013**, *81*, 1857–1861.
- [219] G. B. Irvine, *Curr Protoc Cell Biol* **2001**, Chapter 5.
- [220] R. K. Harris, E. D. Becker, S. M. Cabral De Menezes, P. Granger, R. E. Hoffman, K. W. Zilm, *Pure and Applied Chemistry* **2008**, *80*, 59–84.
- [221] Y. Q. Cheng, M. Yang, A. M. Matter, *Appl Environ Microbiol* **2007**, *73*, 3460.
- [222] T. T. Hoang, R. R. Karkhoff-Schweizer, A. J. Kutchma, H. P. Schweizer, *Gene* **1998**, *212*, 77–86.
- [223] J. Hou, L. Robbel, M. A. Marahiel, *Chem Biol* **2011**, *18*, 655–664.
- [224] M. Kaniusaite, J. Tailhades, T. Kittilä, C. D. Fage, R. J. A. Goode, R. B. Schittenhelm, M. J. Cryle, *FEBS Journal* **2021**, *288*, 507–529.
- [225] B. Shen, L. Du, C. Sanchez, D. J. Edwards, M. Chen, J. M. Murrell, *J Nat Prod* **2002**, *65*, 422–431.
- [226] A. A. Yunis, *Ann. Rev. Pharmacol. Toxicol* **1988**, *28*, 83–100.
- [227] G. M. Singh, P. D. Fortin, A. Koglin, C. T. Walsh, *Biochemistry* **2008**, *47*, 11310–11320.
- [228] H. Chen, L. Zhong, H. Zhou, T. Sun, G. Zhong, Q. Tu, Y. Zhuang, X. Bai, X. Wang, J. Xu, L. Xia, Y. Shen, Y. Zhang, X. Bian, *Angewandte Chemie - International Edition* **2022**, *61*.
- [229] T. M. Makris, C. J. Koot, C. M. Wilmot, J. D. Lipscomb, *Biochemistry* **2013**, *52*, 6662–6671.
-

-
- [230] X. Luo, Y. Zhang, L. Yin, W. Zheng, Y. Fu, *3 Biotech* **2020**, *10*, 1–9.
- [231] H. Zhang, A. L. Nielsen, M. W. Boesgaard, K. Harpsøe, N. L. Daly, X. F. Xiong, C. R. Underwood, L. M. Haugaard-Kedström, H. Bräuner-Osborne, D. E. Gloriam, K. Strømgaard, *Eur J Med Chem* **2018**, *156*, 847–860.
- [232] J. Müller, S. C. Feifel, T. Schmiederer, R. Zocher, R. D. Süssmuth, *ChemBioChem* **2009**, *10*, 323–328.
- [233] T. Sasaki, M. Takagi, T. Yaguchi, S. Miyadoh, T. Okada, M. Koyama, *J Antibiot (Tokyo)* **1992**, *45*, 692–697.
- [234] A. Harder, G. Samson-Himmelstjerna, *Parasitol Res* **2001**, *87*, 924–928.
- [235] J. Scherkenbeck, P. Jeschke, A. Harder, *Curr Top Med Chem* **2005**, *2*, 759–777.
- [236] T. Stadelmann, G. Subramanian, S. Menon, C. E. Townsend, R. S. Lokey, M. O. Ebert, S. Riniker, *Org Biomol Chem* **2020**, *18*, 7110–7126.
- [237] G. Von Samson-Himmelstjerna, A. Harder, T. Schnieder, J. Kalbe, N. Mencke, *Parasitol Res* **2000**, *86*, 194–199.
- [238] J. C. Kwan, R. Ratnayake, K. A. Abboud, V. J. Paul, H. Luesch, *Journal of Organic Chemistry* **2010**, *75*, 8012–8023.
- [239] J. C. Kwan, Y. Liu, R. Ratnayake, R. Hatano, A. Kuribara, C. Morimoto, K. Ohnuma, V. J. Paul, H. Luesch, *ChemBioChem* **2014**, *15*, 799–804.
- [240] Z. Zheng, C. Ma, C. Gao, F. Li, J. Qin, H. Zhang, K. Wang, *PLoS One* **2011**, *6*, 23–26.
- [241] F. Valerio, P. Lavermicocca, M. Pascale, A. Visconti, *FEMS Microbiol Lett* **2004**, *233*, 289–295.
- [242] W. Mu, S. Yu, B. Jiang, X. Li, *Biotechnol Lett* **2012**, *34*, 907–911.
- [243] M. G. Chevette, K. Gutiérrez-García, N. Selem-Mojica, C. Aguilar-Martínez, A. Yañez-Olvera, H. E. Ramos-Aboites, P. A. Hoskisson, F. Barona-Gómez, *Nat Prod Rep* **2020**, *37*, 566–599.
- [244] D. Madern, *J Mol Evol* **2002**, *54*, 825–40.
- [245] A. Chapman, A. Cortés, T. Dafforn, A. Clarke, R. Brady, *J Mol Biol.* **1999**, *285*, 703–12.
- [246] D. T. Dang, *Front Chem* **2022**, *10*.
- [247] F. Flett, V. Mersinias, C. P. Smith, *FEMS Microbiol Lett* **1997**, *155*, 223–229.
- [248] S. C. Gill, P. H. von Hippel, *Anal Biochem* **1989**, *182*, 319–326.
- [249] M. Ciani, G. Bourenkov, G. Pompidor, I. Karpics, J. Kallio, I. Bento, M. Roessle, F. Cipriani, S. Fiedler, T. R. Schneider, *J Synchrotron Radiat* **2017**, *24*, 323–332.
- [250] M. F. Incardona, G. P. Bourenkov, K. Levik, R. A. Pieritz, A. N. Popov, O. Svensson, *J Synchrotron Radiat* **2009**, *16*, 872–879.
- [251] W. Kabsch, *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 125–132.
- [252] P. Evans, *Acta Crystallogr D Biol Crystallogr* **2006**, *62*, 72–82.
- [253] P. R. Evans, G. N. Murshudov, *Acta Crystallogr D Biol Crystallogr* **2013**, *69*, 1204–1214.
-

- [254] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin, K. S. Wilson, *Acta Crystallogr D Biol Crystallogr* **2011**, *67*, 235–242.
- [255] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J Appl Crystallogr* **2007**, *40*, 658–674.
- [256] D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkoczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, *Acta Crystallogr D Struct Biol* **2019**, *75*, 861–877.
- [257] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*:7873 **2021**, *596*, 583–589.
- [258] P. V. Afonine, B. K. Poon, R. J. Read, O. V. Sobolev, T. C. Terwilliger, A. Urzhumtsev, P. D. Adams, *Acta Crystallogr D Struct Biol* **2018**, *74*, 531–544.
- [259] P. Emsley, K. Cowtan, *Acta Crystallogr D Biol Crystallogr* **2004**, *60*, 2126–2132.
- [260] V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, D. C. Richardson, *Acta Crystallogr D Biol Crystallogr* **2010**, *66*, 12–21.
- [261] K. Blin, S. Shaw, H. E. Augustijn, Z. L. Reitz, F. Biermann, M. Alanjary, A. Fetter, B. R. Terlouw, W. W. Metcalf, E. J. N. Helfrich, G. P. Van Wezel, M. H. Medema, T. Weber, *Nucleic Acids Res* **2023**, *51*, W46–W50.
- [262] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, Y. Zhang, *Nat Methods* **2014**, *12*, 7–8.
- [263] A. Roy, A. Kucukural, Y. Zhang, *Nat Protoc* **2010**, *5*, 725–738.
- [264] Y. Zhang, *BMC Bioinformatics* **2008**, *9*.
- [265] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, A. Bairoch, *The Proteomics Protocols Handbook* **2005**, 571–607.
- [266] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, M. Steinegger, *Nat Methods* **2022**, *19*, 679–682.
- [267] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583–589.
- [268] V. Zoete, T. Schuepbach, C. Bovigny, P. Chaskar, A. Daina, U. F. Röhrig, O. Michielin, *J Comput Chem* **2016**, *37*, 437–447.

- [269] M. Bugnon, U. F. Röhrig, M. Goullieux, M. A. S. Perez, A. Daina, O. Michielin, V. Zoete, *Nucleic Acids Res* **2024**, *52*, W324–W332.
- [270] U. F. Röhrig, M. Goullieux, M. Bugnon, V. Zoete, *J Chem Inf Model* **2023**, *63*, 3925–3940.
- [271] A. Grosdidier, V. Zoete, O. Michielin, *Nucleic Acids Res* **2011**, *39*.
- [272] O. Trott, A. J. Olson, *J Comput Chem* **2010**, *31*, 455–461.
- [273] J. Eberhardt, D. Santos-Martins, A. F. Tillack, S. Forli, *J Chem Inf Model* **2021**, *61*, 3891–3898.

8. List of Abbreviations

A domain.....	Adenylation domain
<i>A. crenata</i>	<i>Ardisia crenata</i>
a.u.	Arbitrary unit
AA	Amino acid
AC.....	Adenylate cyclases
A _{core}	A coredomain
ACP	Acyl carrier protein
Ala.....	L-Alanine
AMP-AA	Amino acid-adenylate
Asub	A subdomain
ATP	Adenosine triphosphate
AUC.....	Area under the curve
BGC.....	Biosynthetic gene cluster
C domain	Condensation domain
<i>C. vaccinii</i>	<i>Chromobacterium vaccinii</i> MWU205
<i>Ca. B. crenata</i>	<i>Candidatus</i> Burkholderia crenata
cAMP	Cyclic AMP
CD.....	Cyclic depsipeptide
CoA	Coenzyme A
cryo-EM	Cryogenic electron microscope
C _{start}	Starter C domain
CYP450	Cytochrom P450-dependent monooxygenases
DAG.....	Diacylglycerole
DARPin.....	Designed ankyrin repeat proteins
DLS	dynamic light scattering
DNA	Desoxyribonucleic acid
D-PLA.....	D-Phenyllactate
DSS	Trimethylsilylpropanesulfonate
DTT.....	Dithiothreitol, Dithiothreitol
E domain	Epimerization domain
<i>E. coli</i>	<i>Escherichia coli</i>
EIC	Extracted ion chromatograms
ELSD.....	Evaporative light scattering
ESI-LC/MS.....	Electrospray ionization-liquid chromatography-mass spectrometry
FP	Fusion protein
FPLC.....	Fast protein liquid chromatography
FR.....	FR900359
FRT	Flippase recognition target
G proteins	guanine-nucleotide binding proteins
GC-MS.....	Gas chromatography/mass spectrometry
GDI.....	Guanine nucleotide dissociation inhibitors
gDNA.....	Genomic DNA
GDP.....	Guanosine diphosphate
GFP	Green fluorescent protein
GOI.....	Gene of interest
GPCR.....	G protein-coupled receptor
GTP	Guanosine triphosphate
Hle	Hydroxy-L-leucine
HPLC	High-performance liquid chromatography

IP3	Inositol triphosphate
IPTG	Isopropyl- β -D-thiogalactopyranosid
I-TASSER	Iterative threading assembly refinement
ITC	Isothermal titration calorimetry
KO	Knock-out
LAB	Lactic acid bacteria
LB	Lysogenic-Broth
LDH	Lactate dehydrogenase
LFR	Linearized FR
LFRC	Linearized FR-Core
MALS	Multiangle light scattering
MBP	Maltose binding protein
MDH	Malate dehydrogenase
mEGFP	Monomeric enhanced green fluorescent protein
MLP	MbtH-like protein
modAA C domains	Modified AA C domains
MS	Mass spectrometry
MSA	Multiple sequence alignment
MT domain	Methyltransferase domain
mVenus	Monomeric Venus
N,O-Me ₂ -Thr	<i>N,O</i> -Dimethyl-L-threonine
N-Ac-Hle	<i>N</i> -Acetylhydroxy-L-leucine
NAD(P)	Nicotinamide adenine dinucleotid (phosphate)
NAD(P)H	Nicotinamide adenine dinucleotid (phosphate) hydrogen
nanoDSF	Nano differential scanning fluorimetry
NHDM	Nonheme diiron monooxygenases
Ni-NTA	Nickel-nitrilotriacetic acid
N-Me-Ala	<i>N</i> -Methyl-L-alanine
N-Me-Dha	<i>N</i> -Methyl-L-dehydroalanine
NMR	Nuclear magnetic resonance
NP	Natural product
N-Pp-Hle	<i>N</i> -Propionyl-(2 <i>S</i> ,3 <i>R</i>)-3-hydroxyleucine
NRP	Non-ribosomal peptide
NRPS	Non-ribosomal peptide synthetases
OD ₆₀₀	Optical density at 600 nm
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylen glycol
PIP ₃	Phosphatidylinositol (3,4,5)-triphosphate
PKS	Polyketide synthetases
PLA	Phenyllactate
PLC- β	Phospholipase C- β isozymes
POI	Protein of interest
PPA	Phenylpyruvic acid
Ppant	4'-phosphopantetheinyl
PP _i	Pyrophosphate
PPTase	Phosphopantetheinyl transferase
RE	Restriction enzyme
RhoGEFs	Rho guanine nucleotide exchange factors
rmds	Root mean square displacement
rMMS	Random microseed matrix screening
SAM	S-adenosylmethionine

SCA	Side chain assembly
SDS-PAGE.....	Sodium-dodecyl sulfate-polyacrylamide gel electrophoresis
SEC.....	Size exclusion chromatography
SEM	Standard error of the mean
SM	Specialized metabolites
SOC	Super optimal broth with catabolite repression
TB.....	Terrific-Broth
TE domain.....	Thioesterase domain
TEV	Tobacco etch virus
TROSY	2D-transverse relaxation-optimized spectroscopy
UM	Uveal melanoma
UV	Ultraviolet
vioA.....	Violacein A
WT	Wild type
YM.....	YM-254890

9. Appendix

9.1. Protein sequences

Table 22: List of FrsA_{TE} constructs including amino acid sequences and size.

Construct name	Amino acid sequence	Size [kDa]
His-TEV-FrsA _{TE}	MKHHHHHHHHGGENLYFQGGSAEGGDSKSPYQVLMPIRATGGRHPLFCIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLDGMSELAPSIPDMAADYIEQIRSIQPNGPYHLLGWSLGGVIAQEVAVQLERVGEKTALLAILDTFPIEILHEAMFGKQACAYDLFARVVQEMYLMPIEEARLKSMYLIGLNHMKITAAFSSSHYGGDLLLFRSLIPYAEDALMPEADTWQPYLSGQLEVHDIECTHMDMMQRDVLKIIGPVLESKLSVTAVKQ*	30.4
FrsA _{TE}	GGSAEGGDSKSPYQVLMPIRATGGRHPLFCIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLDGMSELAPSIPDMAADYIEQIRSIQPNGPYHLLGWSLGGVIAQEVAVQLERVGEKTALLAILDTFPIEILHEAMFGKQACAYDLFARVVQEMYLMPIEEARLKSMYLIGLNHMKITAAFSSSHYGGDLLLFRSLIPYAEDALMPEADTWQPYLSGQLEVHDIECTHMDMMQRDVLKIIGPVLESKLSVTAVKQ*	28.2
His-TEV-FrsA _{TE} Ed	MKHHHHHHHHGGENLYFQGGSVLMPIRATGGRHPLFCIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLDGMSELAPSIPDMAADYIEQIRSIQPNGPYHLLGWSLGGVIAQEVAVQLERVGEKTALLAILDTFPIEILHEAMFGKQACAYDLFARVVQEMYLMPIEEARLKSMYLIGLNHMKITAAFSSSHYGGDLLLFRSLIPYAEDALMPEADTWQPYLSGQLEVHDIECTHMDMMQRDVLKIIGPVLESKLSVTAVKQ*	29.3
FrsA _{TE} Ed	GGSVLMPIRATGGRHPLFCIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLDGMSELAPSIPDMAADYIEQIRSIQPNGPYHLLGWSLGGVIAQEVAVQLERVGEKTALLAILDTFPIEILHEAMFGKQACAYDLFARVVQEMYLMPIEEARLKSMYLIGLNHMKITAAFSSSHYGGDLLLFRSLIPYAEALMPEADTWQPYLSGQLEVHDIECTHMDMMQRDVLKIIGPVLESKLSVTAVKQ*	27.0
His-TEV-FrsA _{TE} short	MKHHHHHHHHGGENLYFQGGSPFCIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLDGMSELAPSIPDMAADYIEQIRSIQPNGPYHLLGWSLGGVIAQEVAVQLERVGEKTALLAILDTFPIEILHEAMFGKQACAYDLFARVVQEMYLMPIEEARLKSMYLIGLNHMKITAAFSSSHYGGDLLLFRSLIPYAEDALMPEADTWQPYLSGQLEVHDIECTHMDMMQRDVLKIIGPVLESKLS*	27.4
His-MBP-TEV-FrsA _{TE} short	MGSSHHHHHHSSGLVPRGSHMKIEEGKLVWINGDKGYNGLAIEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDREFGGYASGGLLAEITPDKAFQDKLYPFTWDVAVRYNGKLIAYPIAVEALSIIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIADGGYAFKYENGKYDIKDVGVNDAGAKAGLTFLVDLIKNKHMNADTDYSIAEAFNKGETAMTINGPWAWSNIDTSKVNNGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYYEELAKDPRIAATMENAQKGEIMPNIQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSHHHHHHANSVPLVPRGSENLYFQSPLFCIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLDGMSELAPSIPDMAADYIEQIRSIQPNGPYHLLGWSLGGVIAQEVAVQLERVGEKTALLAILDTFPIEILHEAMFGKQACAYDLFARVVQEMYLMPIEEARLKSMYLIGLNHMKITAAFSSSHYGGDLLLFRSLIPYAEDALMPEADTWQPYLSGQLEVHDIECTHMDMMQRDVLKIIGPVLESKLS*	70.6

FrsA _{TEshort}	GGSPFCIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLDGMSELA PSIPDMAADYIEQIRSIQPNNGPYHLLGWSLGGVIAQEVAVQLERVGE KTALLAILDTFPIEILHEAMFGKQACAYDLFARVVQEMYLMPIEEARL KSMYLI GLNHMKITAAFSSSHYGGDLLLFRSLIPYAEDALMPEADTW QPYL SGQLEVHDIECTHMDMMQRDVLKIIGPVLESKLS*	25.1
His-MBP-TEV- Linker- FrsA _{TEshort}	MGSSHHHHHHSSGLVPRGSHMKIEEGKLVWINGDKGYNGLAEVG KKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHD RFGGYA QSGLLAEITPDKAFQDKLYPFTWDVAVRYNGKLIAYPIAVEALS LIYNK DLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGG YAFKYENGKYDIKDVGVNDAGAKAGLTFLVDLIKNKHMNADTDYSIA EAAFNKGETAMTINGPWAWSNIDTSKVN YGVTVLPTFKGQPSKPFV GVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKS YEEELAKDPRIAATMENAQKGEIMPNI PQMSAFWYAVRTAVINAASG RQTVDEALKDAQTNSSSHHHHHHANSVPLVPRGSENLYFQSGGGS PLFCIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLDGMSELAPSIP DMAADYIEQIRSIQPNNGPYHLLGWSLGGVIAQEVAVQLERVGEKTAL LAILDTFPIEILHEAMFGKQACAYDLFARVVQEMYLMPIEEARL KSMY LI GLNHMKITAAFSSSHYGGDLLLFRSLIPYAEDALMPEADTWQPYL SGQLEVHDIECTHMDMMQRDVLKIIGPVLESKLS*	70.8
Linker- FrsA _{TEshort}	GGGSPFCIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLDGMSEL APSIPDMAADYIEQIRSIQPNNGPYHLLGWSLGGVIAQEVAVQLERVGE KTALLAILDTFPIEILHEAMFGKQACAYDLFARVVQEMYLMPIEEAR L KSMYLI GLNHMKITAAFSSSHYGGDLLLFRSLIPYAEDALMPEADT WQPYL SGQLEVHDIECTHMDMMQRDVLKIIGPVLESKLS*	25.2
His-MBP-TEV- Linker ² - FrsA _{TEshort}	MGSSHHHHHHSSGLVPRGSHMKIEEGKLVWINGDKGYNGLAEVG KKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHD RFGGYA QSGLLAEITPDKAFQDKLYPFTWDVAVRYNGKLIAYPIAVEALS LIYNK DLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGG YAFKYENGKYDIKDVGVNDAGAKAGLTFLVDLIKNKHMNADTDYSIA EAAFNKGETAMTINGPWAWSNIDTSKVN YGVTVLPTFKGQPSKPFV GVLSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKS YEEELAKDPRIAATMENAQKGEIMPNI PQMSAFWYAVRTAVINAASG RQTVDEALKDAQTNSSSHHHHHHANSVPLVPRGSENLYFQSGGGS GGGSPFCIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLDGMSEL APSIPDMAADYIEQIRSIQPNNGPYHLLGWSLGGVIAQEVAVQLERVGE KTALLAILDTFPIEILHEAMFGKQACAYDLFARVVQEMYLMPIEEAR L KSMYLI GLNHMKITAAFSSSHYGGDLLLFRSLIPYAEDALMPEADT WQPYL SGQLEVHDIECTHMDMMQRDVLKIIGPVLESKLS*	71.1
His-mVenus- TEV-Linker ² - FrsA _{TEshort}	MGSSHHHHHHSSGLVPRGSHMVSKGEELFTGVVPILVELDGDVNG HKFSVSGEGEGDATYGKLTCLKICTTGKLPVPWPTLVTTLG YGLQCF ARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRA EVKFEG DTLVNRIELKGIDFKEDGNILGHKLEYNYN SHNVYITADKQKNGIKAN FKIRHNIEDGGVQLADHYQQNTPIGDGPVLLPDNHYLSYQSALSKD PNEKRDHMLLEFVTAAGITLGMDELYKENLYFQSGGGSGGGSP LF CIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLDGMSELAPSIPDMA ADYIEQIRSIQPNNGPYHLLGWSLGGVIAQEVAVQLERVGEKTALLAIL DTFPIEILHEAMFGKQACAYDLFARVVQEMYLMPIEEARL KSMYLI GL NHMKITAAFSSSHYGGDLLLFRSLIPYAEDALMPEADTWQPYL SGQ LEVHDIECTHMDMMQRDVLKIIGPVLESKLS*	55.3
Linker ² - FrsA _{TEshort}	SGGGSGGGSPFCIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLD GMSELAPSIPDMAADYIEQIRSIQPNNGPYHLLGWSLGGVIAQEVAVQ LERVGEKTALLAILDTFPIEILHEAMFGKQACAYDLFARVVQEMYLMPI IEEARL KSMYLI GLNHMKITAAFSSSHYGGDLLLFRSLIPYAEDALMP EADTWQPYL SGQLEVHDIECTHMDMMQRDVLKIIGPVLESKLS*	25.5

His-TEV-FrsA _{TElong}	MKHHHHHHHHGGENLYFQGGSPFCIHPEGGLGWSYIGLALHLDH EQPIYTLQARGLDGMSELAPSIPDMAADYIEQIRSIQPNPGPYHLLGW SLGGVIAQEVAVQLERVGEKTALLAILDTFPIELHEAMFGKQACAYD LFARVVQEMYLMPIEEARLKSMYLIGLNHMKITAAFSSSHYGGDLLL FRSLIPYAEDALMPEADTWQPYLSGQLEVHDIECTHMDMMQRDVL KIIGPVLESKLSVTAVKQ*	28.0
His-MBP-TEV-FrsA _{TElong}	MGSSHHHHHHSSGLVPRGSHMKIEEGKLVWINGDKGYNGLAEVG KKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYA QSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSIIYNK DLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGG YAFKYENGKYDIKDVGVNDAGAKAGLTFLVDLIKNKHMNADTDYSIA EAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFV GVL SAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKS YEEELAKDPRIAATMENAQAQGEIMPNIQMSAFWYAVRTAVINAASG RQTVDEALKDAQTNSSSHHHHHHANSVPLVPRGSENLYFQSGPLFCI HPEGGLGWSYIGLALHLDHEQPIYTLQARGLDGMSELAPSIPDMAA DYIEQIRSIQPNPGPYHLLGWSLGGVIAQEVAVQLERVGEKTALLAILD TFPIELHEAMFGKQACAYDLFARVVQEMYLMPIEEARLKSMYLIGLN HMKITAAFSSSHYGGDLLLFRSLIPYAEDALMPEADTWQPYLSGQLE VHDIECTHMDMMQRDVLKIIGPVLESKLSVTAVKQ*	71.2
FrsA _{TElong}	GGSPFCIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLDGMSELA PSIPDMAADYIEQIRSIQPNPGPYHLLGWSLGGVIAQEVAVQLERVGE KTALLAILDTFPIELHEAMFGKQACAYDLFARVVQEMYLMPIEEARL KSMYLIGLNHMKITAAFSSSHYGGDLLLFRSLIPYAEDALMPEADTW QPYLSGQLEVHDIECTHMDMMQRDVLKIIGPVLESKLSVTAVKQ*	25.7
His-MBP-TEV-Linker-FrsA _{TElong}	MGSSHHHHHHSSGLVPRGSHMKIEEGKLVWINGDKGYNGLAEVG KKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYA QSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSIIYNK DLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGG YAFKYENGKYDIKDVGVNDAGAKAGLTFLVDLIKNKHMNADTDYSIA EAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFV GVL SAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKS YEEELAKDPRIAATMENAQAQGEIMPNIQMSAFWYAVRTAVINAASG RQTVDEALKDAQTNSSSHHHHHHANSVPLVPRGSENLYFQSGGGS PLFCIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLDGMSELAPSIP DMAADYIEQIRSIQPNPGPYHLLGWSLGGVIAQEVAVQLERVGEKTAL LAILDTFPIELHEAMFGKQACAYDLFARVVQEMYLMPIEEARLKSMY LIGLNHMKITAAFSSSHYGGDLLLFRSLIPYAEDALMPEADTWQPYL SGQLEVHDIECTHMDMMQRDVLKIIGPVLESKLSVTAVKQ*	71.4
Linker-FrsA _{TElong}	SGGGSPFCIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLDGMSE LAPSIPDMAADYIEQIRSIQPNPGPYHLLGWSLGGVIAQEVAVQLERV GEKTALLAILDTFPIELHEAMFGKQACAYDLFARVVQEMYLMPIEEA RLKSMYLIGLNHMKITAAFSSSHYGGDLLLFRSLIPYAEDALMPEADT WQPYLSGQLEVHDIECTHMDMMQRDVLKIIGPVLESKLSVTAVKQ*	25.9
His-MBP-TEV-Linker ² -FrsA _{TElong}	MGSSHHHHHHSSGLVPRGSHMKIEEGKLVWINGDKGYNGLAEVG KKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYA QSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSIIYNK DLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGG YAFKYENGKYDIKDVGVNDAGAKAGLTFLVDLIKNKHMNADTDYSIA EAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFV GVL SAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKS YEEELAKDPRIAATMENAQAQGEIMPNIQMSAFWYAVRTAVINAASG RQTVDEALKDAQTNSSSHHHHHHANSVPLVPRGSENLYFQSGGGS	71.7

	GGGSPLFCIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLDGMSEL APSIPDMAADYIEQIRSIQPNPGPYHLLGWSLGGVIAQEVAVQLERV EKTALLAILDTFPIELHEAMFGKQACAYDLFARVVQEMYLMPIEEAR LKSMYLIGLNHNMKITAAFSSSHYGGDLLLFRSLIPYAEDALMPEADT WQPYLSGQLEVHDIECTHMDMMQRDVLKIIGPVLESKLSVTAVKQ*	
His-mVenus- TEV-Linker ² - FrsA _{TElong}	MGSSHHHHHHSSGLVPRGSHMVSKGEELFTGVVPILVELDGDVNG HKFSVSGEGEGDATYGKLTCLKICTTGKLPVPWPTLVTTLG YGLQCF ARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEG DTLVNRIELKGIDFKEDGNILGHKLEYNNSHNVIYITADKQKNGIKAN FKIRHNIEDGGVQLADHYQNTPIGDGPVLLPDNHLYSYQSALSKD PNEKRDHMLLEFVTAAGITLGMDELYKENLYFQSGGGSGGGSPFL CIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLDGMSELAPSIPDMA ADYIEQIRSIQPNPGPYHLLGWSLGGVIAQEVAVQLERVGEKTALLAIL DTFPIELHEAMFGKQACAYDLFARVVQEMYLMPIEEARLKSMYLI GLNHNMKITAAFSSSHYGGDLLLFRSLIPYAEDALMPEADTWQPYLSGQ LEVHDIECTHMDMMQRDVLKIIGPVLESKLSVTAVKQ*	55.9
Linker ² - FrsA _{TElong}	SGGGSGGGSPFLCIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLD GMSELAPSIPDMAADYIEQIRSIQPNPGPYHLLGWSLGGVIAQEVAVQ LERVGEKTALLAILDTFPIELHEAMFGKQACAYDLFARVVQEMYLM PIEEARLKSMYLIGLNHNMKITAAFSSSHYGGDLLLFRSLIPYAEDALM PEADTWQPYLSGQLEVHDIECTHMDMMQRDVLKIIGPVLESKLSVT AVKQ*	26.1
His-TEV-FrsA _{T+TE}	MKHHHHHHHHGGENLYFQGGQRPRNAQEEMLCGLFAEVLDMK VGRGDSFFDLGGHSLLATRLIRRIRETLDVELSIRDLEAPCVTELSRH IAEGGDSKSPYQVLMPIRATGGRHPLFCIHPEGGLGWSYIGLALHLD HEQPIYTLQARGLDGMSELAPSIPDMAADYIEQIRSIQPNPGPYHLLG WSLGGVIAQEVAVQLERVGEKTALLAILDTFPIELHEAMFGKQACAY DLFARVVQEMYLMPIEEARLKSMYLIGLNHNMKITAAFSSSHYGGDLL LFRSLIPYAEDALMPEADTWQPYLSGQLEVHDIECTHMDMMQRDVL KIIGPVLESKLSVTAVKQ*	38.7
FrsA _{T+TE}	GGQRPRNAQEEMLCGLFAEVLDMKVGGRGDSFFDLGGHSLLATRL IRRIRETLDVELSIRDLEAPCVTELSRHIAEGGDSKSPYQVLMPIRA TGGRHPLFCIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLDGMSE LAPSIPDMAADYIEQIRSIQPNPGPYHLLGWSLGGVIAQEVAVQLERV GEKTALLAILDTFPIELHEAMFGKQACAYDLFARVVQEMYLMPIEEA RLKSMYLIGLNHNMKITAAFSSSHYGGDLLLFRSLIPYAEDALMPEADT WQPYLSGQLEVHDIECTHMDMMQRDVLKIIGPVLESKLSVTAVKQ*	36.4

9.2. Plasmids

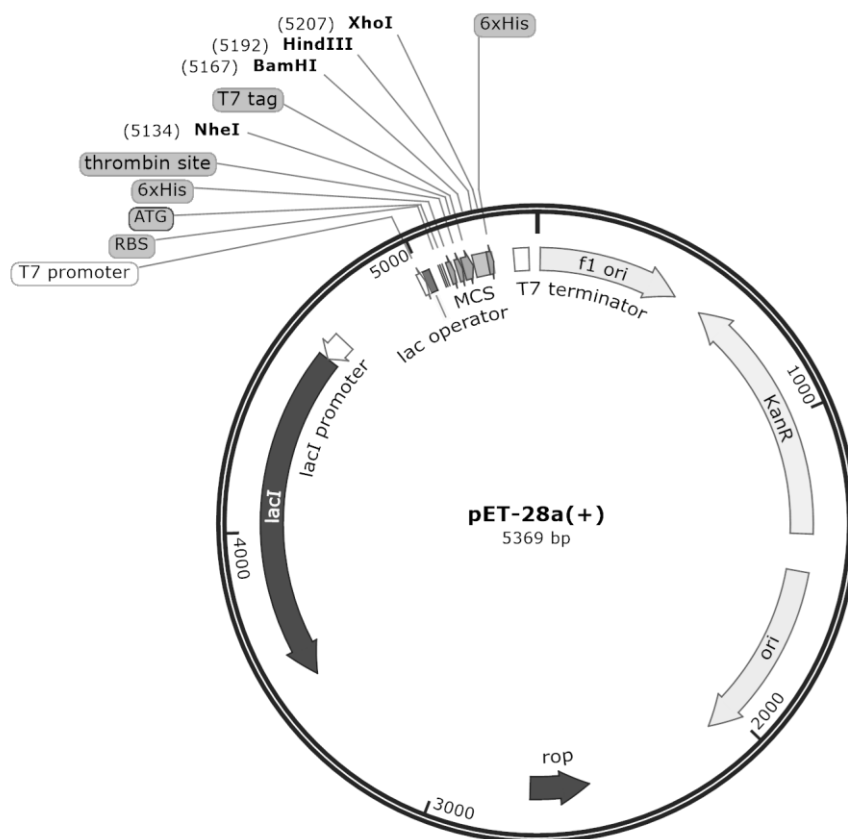


Figure 64: Plasmid map of pET28a(+).

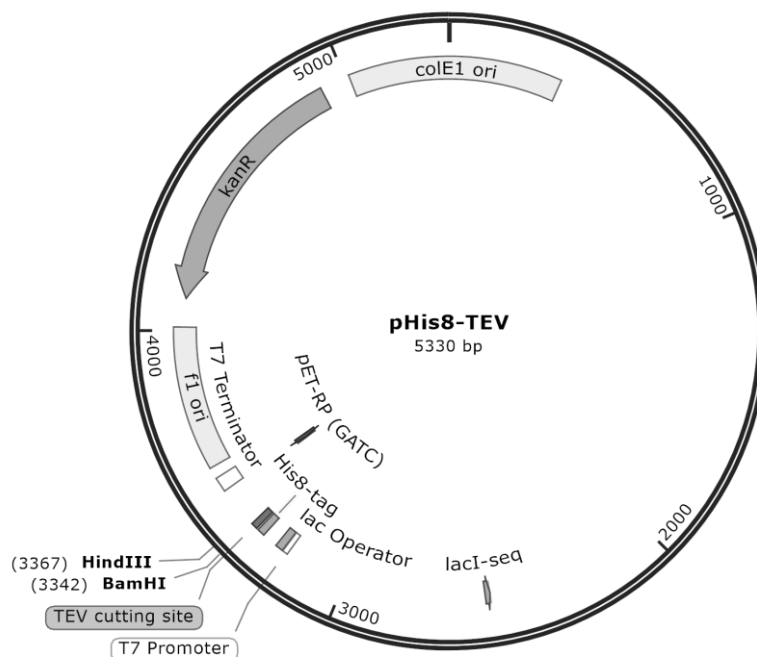


Figure 65: Plasmid map of pHis8-TEV.

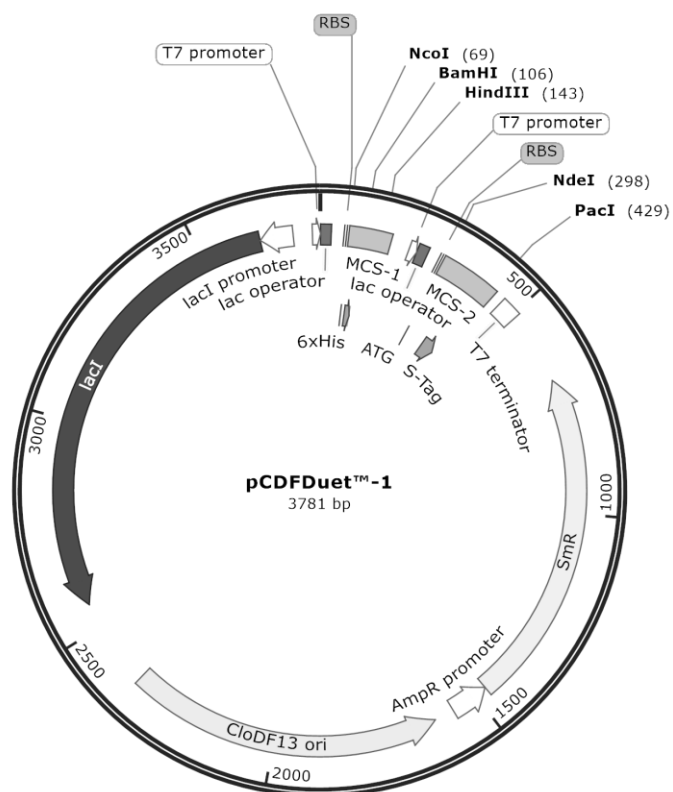


Figure 66: Plasmid map of pCDFDuet-1.

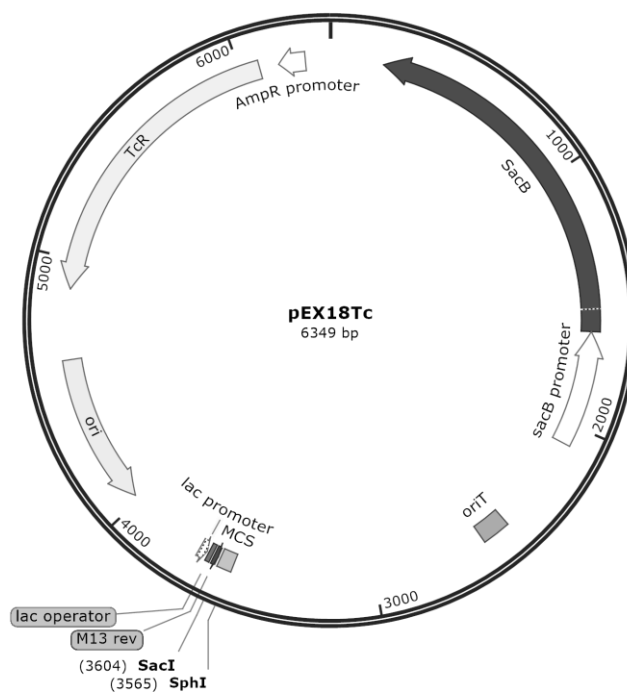


Figure 67: Plasmid map of pEX18Tc.

9.3. SDS-PAGE

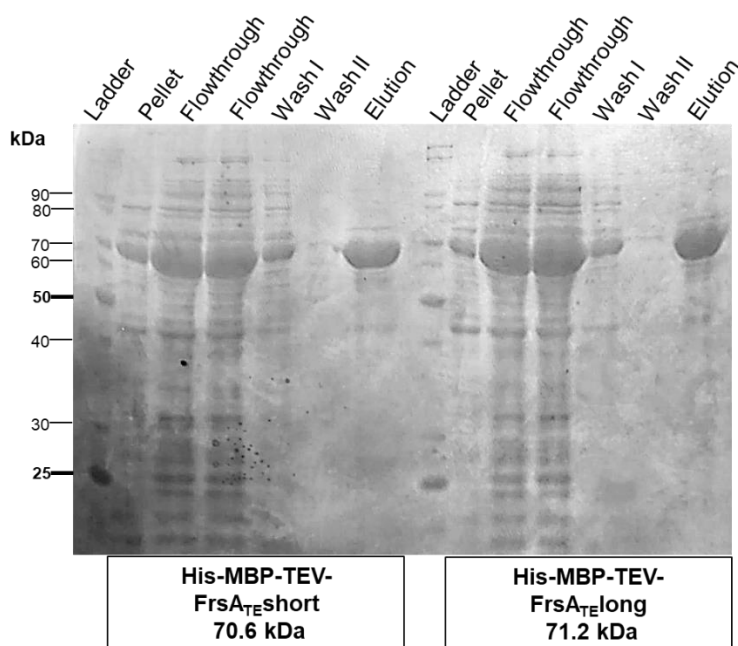


Figure 68: SDS-PAGE of the expression in *E. coli* BL21 and purification of the constructs His-MBP-TEV-FrsA_{TE}short (left, expected size = 70.6 kDa) and His-MBP-TEV-FrsA_{TE}long (right, expected size = 71.2 kDa). Applied volumes: purification samples = 15 μ L, ladder = 3 μ L.

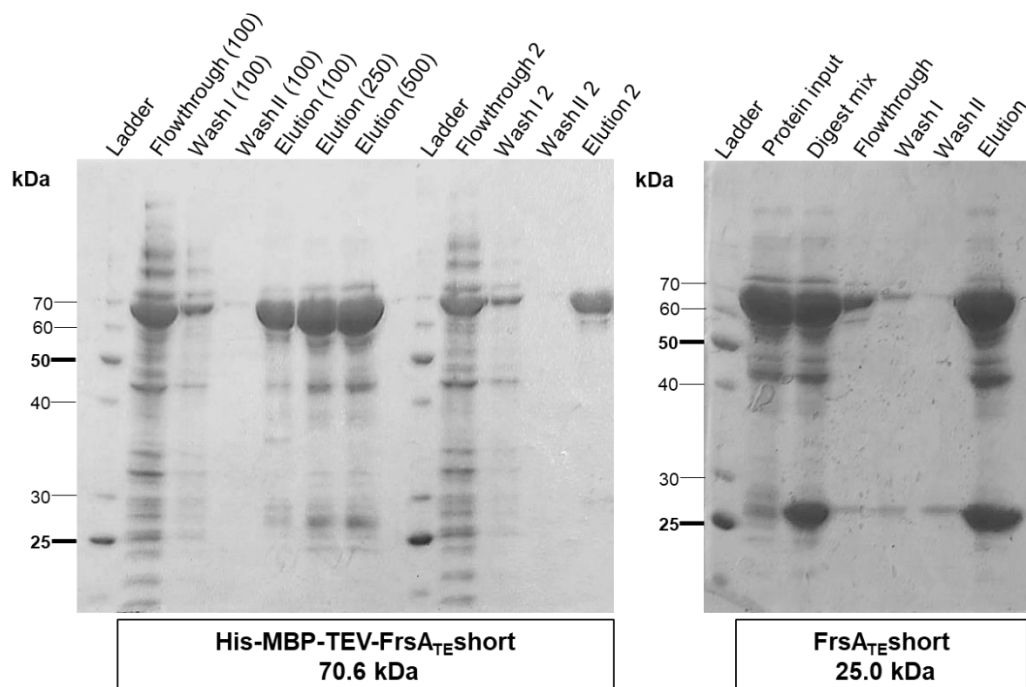


Figure 69: SDS-PAGE of the upscaling experiments (left, in parentheses the culture size in mL), purification optimization (middle, 2), and TEV removal of the His-tag (right) of the construct His-MBP-TEV-FrsA_{TE}short (expected size = 70.6 kDa, after His-Tag removal = 25 kDa) expressed in *E. coli* BL21. Applied volumes: purification samples = 15 μ L, ladder = 3 μ L.

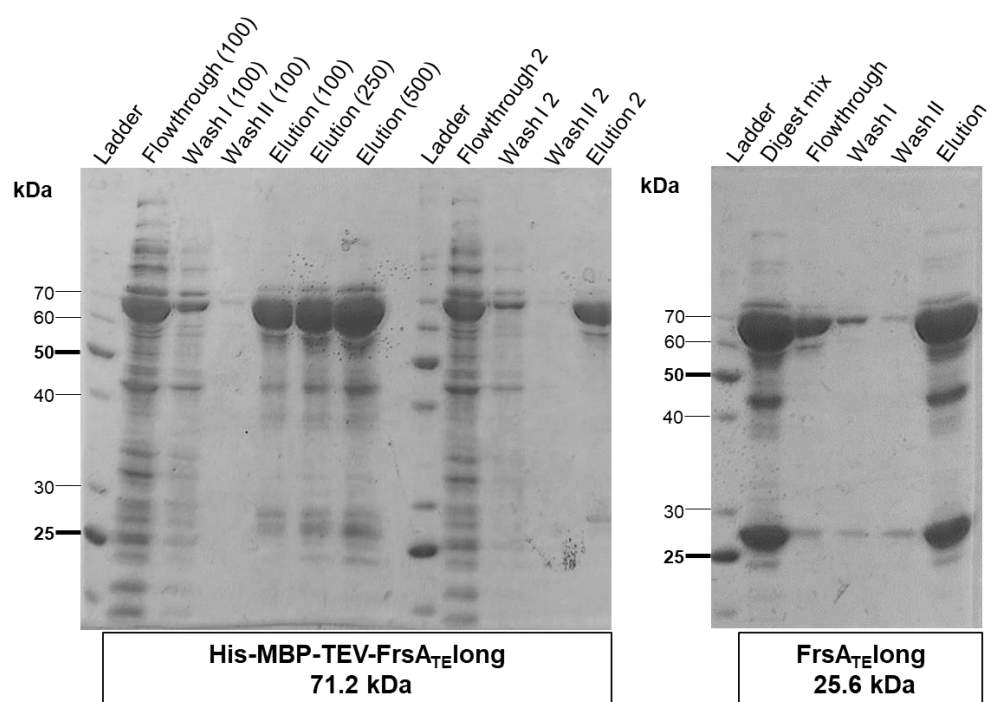


Figure 70: SDS-PAGE of the upscaling experiments (left, in parentheses the culture size in mL), purification optimization (middle, 2), and TEV removal of the His-tag (right) of the construct His-MBP-TEV-FrsA_{TE}long (expected size = 71.2 kDa, after His-Tag removal = 25.6 kDa) expressed in *E. coli* BL21. Applied volumes: purification samples = 15 µL, ladder = 3 µL.

9.4. HPLC/MS-Chromatograms

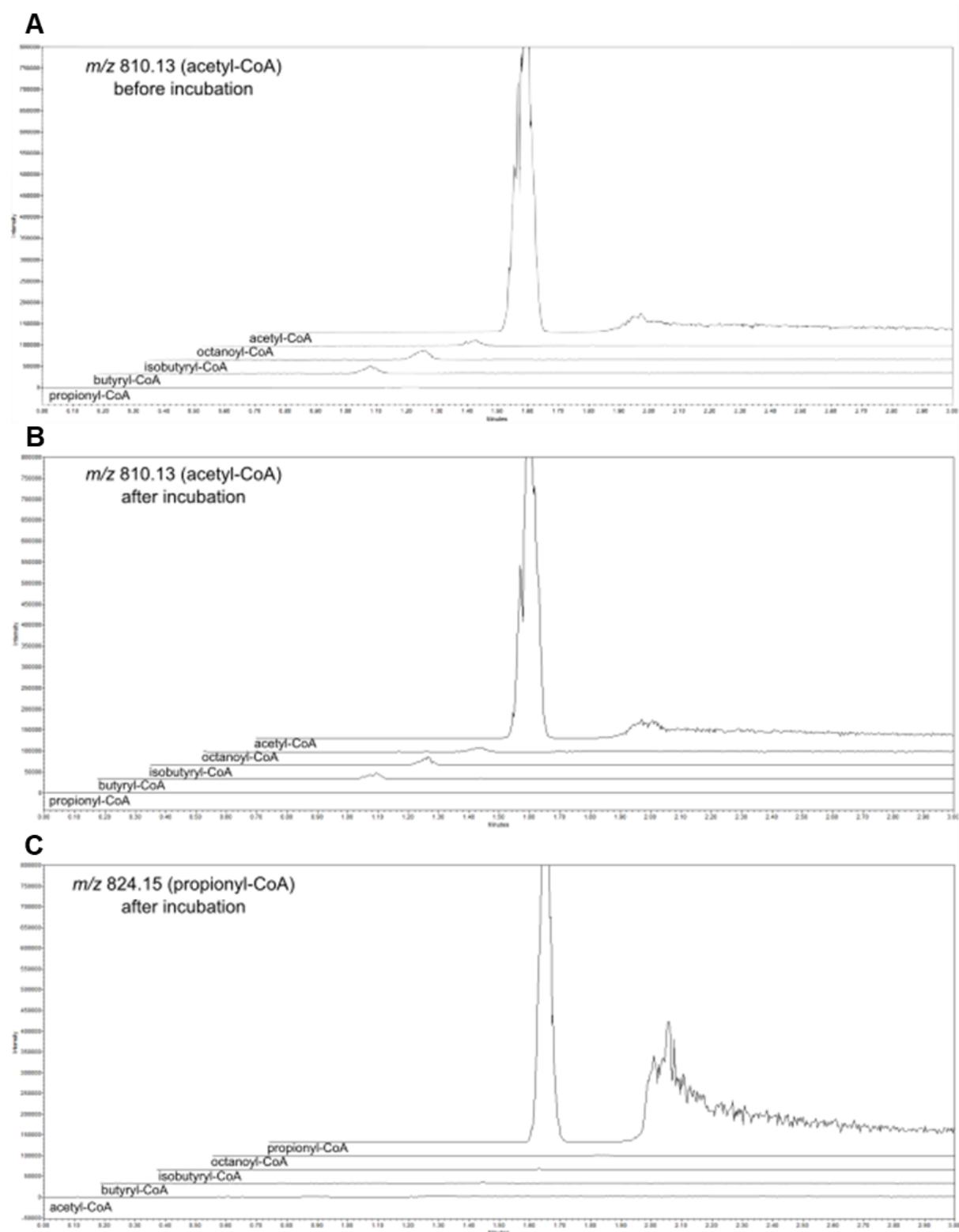


Figure 71: Extracted ion chromatograms of the HPLC-MS measurements of acyl-CoA stocks at the specific m/z characteristic for (A/B) acetyl-CoA (m/z 810.13) and (C) propionyl-CoA (m/z 824.15). Samples were measured (A) after storage at -20 °C and (B/C) after incubation under side chain assembly and transfer assay conditions.

10. Publications

Research publications

S. Klöppel, R. Richarz, D. A. Wirtz, N. Vasenda, G. M. König, M. Crüsemann „A Specialized Dehydrogenase Provides L-Phenyllactate for FR900359 Biosynthesis” *ChemBioChem* **2022**, *23*, 10. DOI: 10.1002/cbic.202100569 - This publication is reproduced in **Chapter 4.6** of this dissertation.

D. A. Wirtz, N. Schneberger, **S. Klöppel**, R. Richarz, M. Geyer, G. M. König, G. Hagelüken, M. Crüsemann „Adenylation Domain-Guided Recruitment of Trans-Acting Nonheme Monooxygenases in Nonribosomal Peptide Biosynthesis” *ACS Chem. Biol.* **2023**, *18*, 8. DOI: 10.1021/acschembio.3c00106 - This publication is reproduced in **Chapter 4.5** of this dissertation.

Oral presentation

S. Klöppel, C. Bonifer, J. Mühle, C. Hermes, G. Schertler, C. Glaubitz, G. M. König, M. Crüsemann „Structural and biochemical insights into the transesterifying thioesterase domain from FR900359 biosynthesis” Workshop of the International Association for General and Applied Microbiology (VAAM) 2022, Dortmund, Germany.

Poster presentations

S. Klöppel, C. Hermes, J. Mühle, C. Bonifer, C. Glaubitz, G. Schertler, G. M. König, M. Crüsemann „Structural insights into the noncanonical TE domain from FR900359 biosynthesis” Workshop of the International Association for General and Applied Microbiology (VAAM) 2021 (online)

S. Klöppel, G. Kowalski, C. Bonifer, J. Mühle, C. Hermes, G. Schertler, C. Glaubitz, G. M. König, M. Crüsemann „Structural and biochemical insights into the transesterifying thioesterase domain from FR900359 biosynthesis” Workshop of the International Association for General and Applied Microbiology (VAAM) 2023, Saarbrücken, Germany.

S. Klöppel, C. Bonifer, J. Mühle, C. Hermes, G. Schertler, C. Glaubitz, G. M. König, M. Crüsemann „Structural and biochemical insights into the transesterifying thioesterase domain from FR900359 biosynthesis” International Conference “Cellular Mechanisms of Antibiotic Action and Production” organised by CellMAP-TRR261, 2023, Bad Boll, Germany.