Investigations into side chain assembly and attachment during biosynthesis of the G protein inhibitor FR900359

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Cornelia Hermes

aus

Hameln

Bonn 2021

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

1. Gutachter: Prof. Dr. G. M. König

2. Gutachter: Prof. Dr. E. Kostenis

Tag der Promotion: 03.05.2021

Erscheinungsjahr: 2021

Table of contents

1 Abstract

The natural product FR900359 (FR) selectively inhibits Gαq proteins and thus intracellular signalling of many G protein-coupled receptors. This unique mechanism of action makes FR an indispensable pharmacological tool to study Gαq-related processes, as well as a promising drug candidate. FR is a complex cyclic depsipeptide with seven nonproteinogenic building blocks; it was isolated from the plant *Ardisia crenata* but is not produced by the latter. Instead, the endosymbiotic bacterium "*Candidatus* Burkholderia crenata" contains the biosynthetic gene cluster (BGC) *frs*, encoding two nonribosomal peptide synthetase (NRPS) systems. Recently, the soil bacterium *Chromobacterium vaccinii* was found to harbour a very similar *frs* BGC and to produce FR under laboratory conditions, facilitating *in vitro* and *in vivo* biosynthesis studies on FR.

In this work, the successive assembly of the FR side chain *N*-propionylhydroxyleucine was achieved *in vitro* by utilising the purified monomodular NRPS FrsA, the MbtH-like chaperon protein FrsB and the non-heme diiron monooxygenase FrsH. The final step of FR biosynthesis is an unusual intermolecular transesterification reaction, connecting the side chain with a macrocyclic intermediate (FR-Core), that is assembled by the heptamodular NRPS FrsD-G. FR-Core was isolated from the deletion mutant *C. vaccinii* Δ*frsA*. It has been shown that the FrsA thioesterase domain catalyses this transesterification and the substrate promiscuity of the FrsA domains was utilised for the chemoenzymatic production of FR derivatives with altered side chains. A new and unnatural derivative, FR-5, was produced in an *in vitro* assay and then isolated from *C. vaccinii* after precursor-directed biosynthesis, induced by feeding of butyric acid. This new analogue contains *N*-butyrylhydroxyleucine instead of *N*-propionylhydroxyleucine as side chain. The Gαq protein inhibiting activity of this new compound was 7-fold decreased in comparison to FR, demonstrating that this position is unsuitable for further modifications.

Comparative *in vivo* and *in vitro* studies on FR-Core and FR supported by *in silico* docking to Gαq demonstrate that the side chain of FR is crucial for its remarkable Gαq inhibition properties. An evolutionary scenario is presented, leading to this important biosynthetic modification. Phylogenetic analysis of the starter condensation and thioesterase domains of the two *frs* clusters revealed their closest relatives to be inside the BGC, and overall no closely related BGCs could be found in a global BGC search. So, our hypothesis involves duplication of the highly similar NRPS module FrsD and the incorporation and possible modification of the unusual thioesterase domain.

Lastly, we started preparations for the structural investigation of the three-dimensional structure of the NRPS FrsA and its thioesterase domain using highly pure protein for first crystallisation and cryo-EM trials.

2 Introduction

The manuscript of the review article “The Chromodepsins – Chemistry, Biosynthesis and Ecology of a selective Gαq inhibitor natural product family” written by the author of this thesis supervised by Dr. Max Crüsemann and Prof. Gabriele König should serve as an introduction to this work. It contains detailed information on the natural product FR900359 (FR), its derivatives and the state of research concerning its structure, bioactivity, and biosynthesis up to Dec. 2020.

2.1 Importance of Gαq inhibitors with focus on FR900359

In the last two decades, two natural products with the ability to selectively inhibit Gαq proteins have drawn the attention of researchers focusing on signalling due to their emergence as valuable pharmacological tools. The structurally highly similar cyclic depsipeptides FR900359 (FR) and YM-254890 (YM) were shown to selectively inhibit the guanosine triphosphate/diphosphate (GTP/GDP) exchange in Gαq proteins at nanomolar potency.[1,2] As there are nearly 1000 G protein-coupled receptors (GPCRs) encoded in the human genome,[3] GPCRs play indispensable roles in human and other mammalian physiological processes.[4,5] FR and YM „trap“ the Gαq protein, one of the major G protein families, in its GDP-bound inactive form and thus prohibit the dissociation of the heterotrimeric G proteins and any subsequent downstream signalling. This novel and very effective mechanism of action for natural products is shown in Figure 2.1 and was evidenced by co-crystallisation of Gαqβγ with YM.[2]

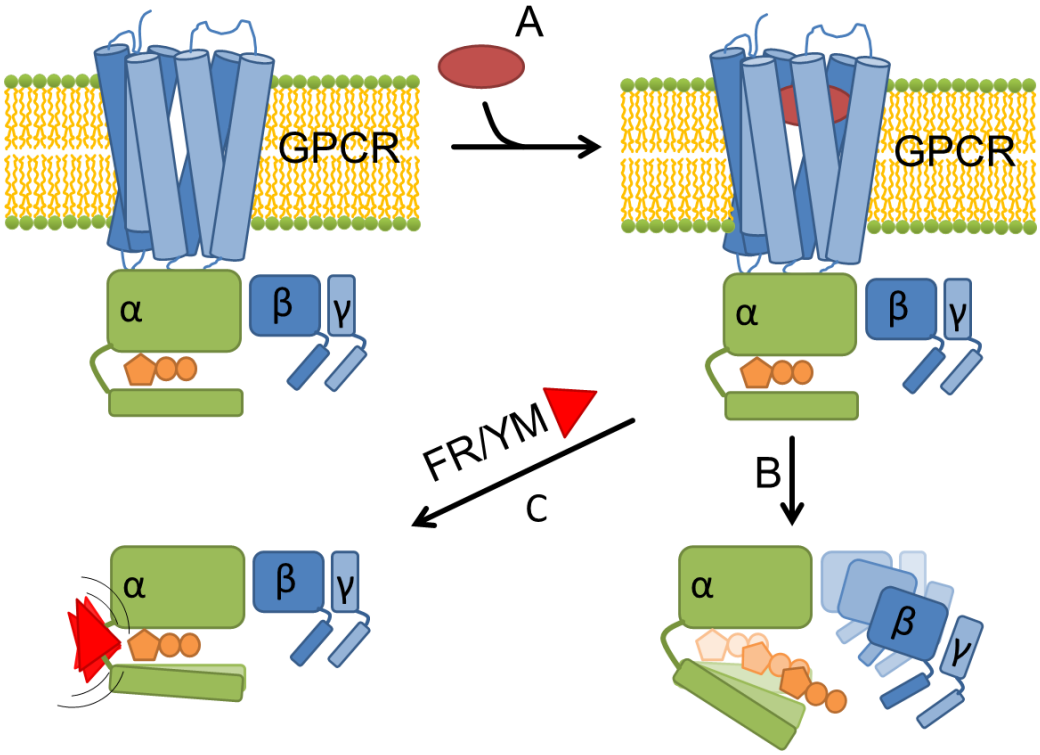


Figure 2.1: Molecular mechanism of G protein inhibition by YM and FR describing how YM/FR impair the opening motion of the Gαq protein, adapted from Tietze *et al.* A. After activation of the GPCR by a ligand (red circle), the GPCR activates the G proteins by acting as a Guanine-nucleotide exchange factor via conformational changes **B.** GDP is released, and GTP can be bound. If there is no inhibitor present, the activated heterotrimer will dissociate into the α subunit and the βγ complex. Both the α subunit and the βγ complex can activate downstream effectors **C.** In the presence of the inhibitor, the

domain opening is blocked and the GDP cannot be exchanged for GTP. This inhibits the dissociation of the subunits and downstream signalling.[6]

As Gαq proteins are involved in diverse pathological conditions, e.g. induced by mutations,[7] compounds directly targeting these proteins are not only valuable for pharmacological studies to investigate G protein-related signalling but might also be utilised for the treatment of diseases. Approximately 34% of all current drugs target GPCRs,[8] highlighting the importance of this heterogenic group of receptors and their associated G proteins for pathologic signal transduction. Targeting not a particular GPCR subtype, but instead, a G protein like Gαq that is coupled to many different GPCRs might result in therapeutic advantages for managing complex diseases with many GPCRs involved.[9]

For the Gαi/o protein family, pertussis toxin (PTX), has proven to be a valuable tool to suppress downstream signalling.[10] Analogously, FR and YM show very strong inhibition of the Gαq family and have thus become important tools for the investigation and characterisation of Gαq protein-related processes. This is reflected in an explosion of citations in the last 10 years (Figure 2.2). So far, despite various efforts, no synthetic molecule with higher Gαq inhibition potency or selectivity than the natural products FR or YM was detected or generated,[11,12] showing that these compounds possibly represent optimal scaffolds for the specific inhibition of Gαq proteins.

This review mainly focuses on the chemistry, biosynthesis, ecology and evolution of this depsipeptide natural product family, for which we propose the name chromodepsins, and follows the history from their initial discovery to their current importance as pharmacological tools. Additionally, we will discuss opportunities for future developments of these extraordinary natural products.

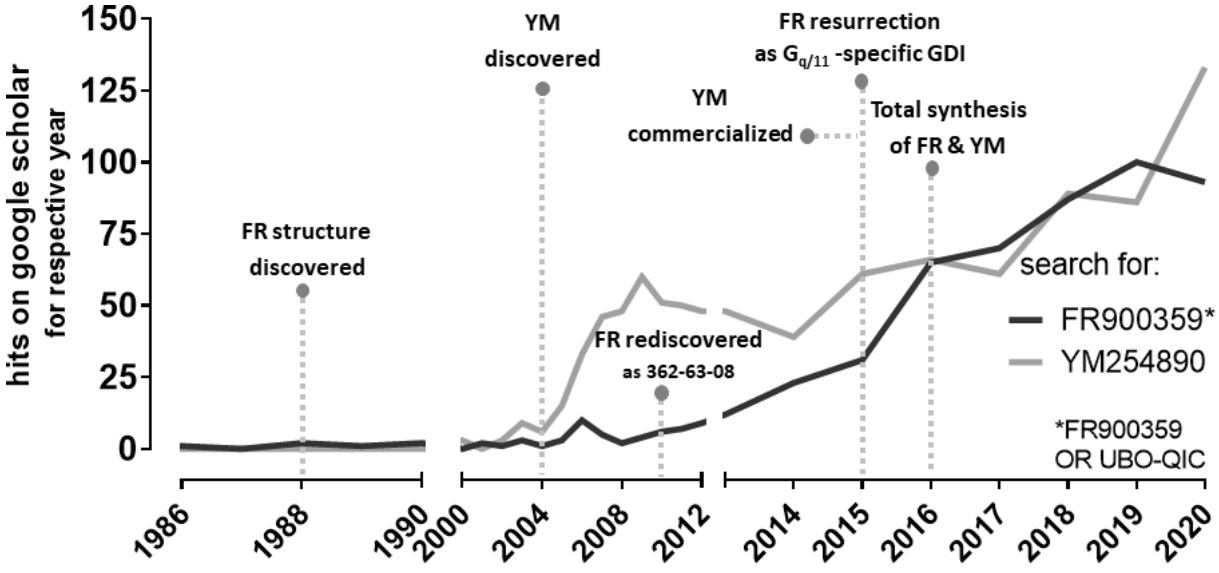


Figure 2.2: Google scholar hits for Gαq inhibitors FR and YM, adapted from Kostenis et al.[9] and updated (Dec. 2020). 1988: Isolation and structure elucidation of FR; biology and mechanism of action unknown.[13] **2004:** Discovery of the structurally close analogue YM by Yamanouchi Pharmaceutical Co.,[14] later combined in a merger with Fujisawa to form Astellas Pharma, which chose to provide YM to the scientific community in a rather restrictive manner. Until commercialization (see below), YM was available for a small number of researchers only. **2010:** Rediscovery of FR, code-named "362-63-08," from a plant extract library as inhibitor of the Gαq-coupled cholecystokinin CCK1 receptor.[15] **2015:** Resurrection of FR by in-depth characterisation of its *in vitro* specificity and mechanism of action by a concerted effort of members of the signal transduction community.[1] **2016:** Commercialization of YM by Fujifilm Wako Chemicals, as well as total synthesis of YM and

FR.[16] Coincidentally, worldwide awareness of and interest in FR and YM has risen steeply. During a short period of time, FR was commercialized under the code name "UBO-QIC" (University of Bonn–Gq-inhibiting component), which indicated market potential and, in turn, encouraged commercialization of the competing molecule YM.[9]

2.2 Discovery and Structure Elucidation

2.2.1 Discovery and structure elucidation of FR900359 (FR)

In 1986, the Japanese Analytical Research Laboratory, Fujisawa Pharmaceutical Co., Ltd isolated a bioactive cyclic depsipeptide which they named FR900359 (Figure 2.3 A, **1**). This natural product was isolated from a methanolic extract of the plant *Ardisia crenata* Sims. The authors stated, that the fractionation during isolation was driven by following the blood pressure decreasing activity in anesthetised normotensive rats.[13] In a subsequent study, the isolated compound was shown to inhibit platelet aggregation *in vitro* and *ex vivo* in rabbits. This investigation also confirmed the decrease in blood pressure by showing dose-related hypotensive action in anaesthetised normotensive rats.[17] Additionally, FR was stated to be cytotoxic in cultured rat fibroblasts and myelocytic leukaemia cells. However, only its structure elucidation was described in detail. A study from 2013 confirmed the stated cardiovascular bioactivity of FR. Its cytotoxicity could however not be proven so far.[18]

First attempts to resolve the structure of FR included mass spectrometric (MS) studies, ^{1}H and ^{13}C nuclear magnetic resonance (NMR) spectroscopy and chemical analytical methods like hydrolytic cleavage of the building blocks. The molecular weight was determined as 1001.5402 g/mol and the molecular formula found to be $C_{46}H_{75}N_7O_{15}$. The cyclic depsipeptide contains several non-proteinogenic amino and hydroxy acids. The core structure consists of the following units (the numbers indicate the number of the biosynthetic module responsible for this building block, compare Figure 2.3 and Figure 2.5): L-alanine (4), L-*N*-methylalanine (3), two β-hydroxy-L-leucine residues (2,7), one of them *N*-acetylated (2), D-3-phenyllactic acid (6), and the rare amino acids L-*N*-methyldehydroalanine (5) and L-*N*,*O*-dimethylthreonine (8). Additionally, a side chain consisting of an *N*-propionylated β-hydroxy- L-leucine residue (1) is attached to the hydroxy group of residue (7) in the core molecule. The FR structure is depicted in Figure 2.3 A.[13] One year after the discovery of FR, Miyamae *et al.* published the threedimensional structure of FR giving the absolute configuration of all eleven stereocenters. The group used gas chromatography coupled to MS detection (GC-MS) and X-ray crystallographic analysis to determine five chiral centres as *R* and the other six as *S*. Furthermore, five intramolecular hydrogen bonds were found to stabilise the FR structure resulting in two *cis* configured peptide bonds, rarely found in peptides and depsipeptides. The crystal structure also suggested a hydrophobic surface of the FR molecule with no intermolecular hydrogen bonds.[17]

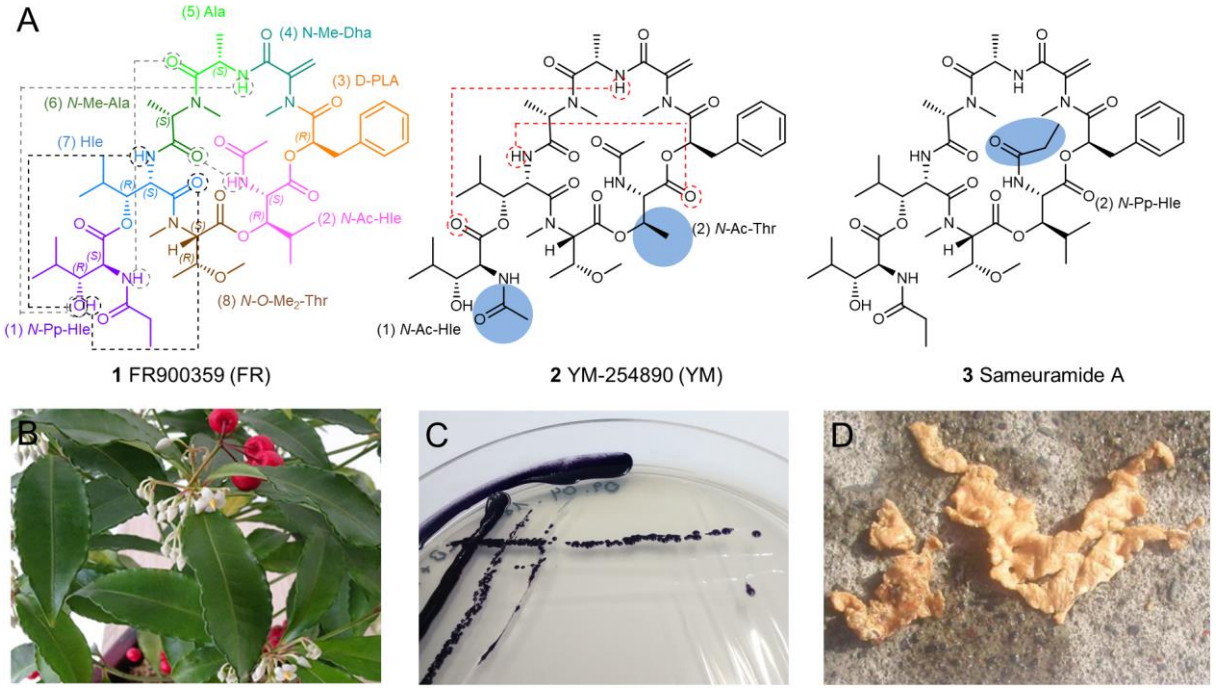


Figure 2.3: Structures of FR, YM and Sameuramide A. The intramolecular hydrogen bonds are highlighted in dash lines and the atoms involved in the formation of these bonds are marked with dashed circles. Hle = 3-hydroxyleucine, PLA = phenyllactic acid, Dha = dehydroalanine. **B.** Plant *A. crenata* (picture taken by Dr. Raphael Reher, AG König). **C.** Bacterial colonies of *C. vaccinii* (picture taken by Dr. René Richarz, AG König). **D.** The didemnid ascidian, source of sameuramide A (picture from graphical abstract of Machida *et al.*).[19]

At the time when FR was discovered, the *N*,*O*-dimethylthreonine building block had not been found in any other natural product, and to date, only one further compound containing *N*,*O*-dimethylthreonine, a cyclic dipeptide from *Streptomyces* species, was reported.[20] *N*-methyldehydroalanine is also a rare feature, the reactive functionality of which led to the assumption, that nucleophilic attacks by the exomethylene group could be relevant for the bioactivity.[13] This assumption was disproven 2004 by Taniguchi *et al.* for the structurally very similar compound YM, showing that hydrogenation of the double bond does not result in significantly abolished bioactivity, leading to the conclusion that Gαq/11 is not covalently modified.[21] Equivalent results were obtained for FR in 2015 with the hydrogenated FR-red.[1]

After these two reports on the discovery and initial bioactivity of FR from the 1980s, this compound was not further investigated for almost fifteen years, until in 2003 a natural product with high structural similarity to FR, named YM-254890 was isolated (see 2.2.2). In 2010, FR was rediscovered in a plant extract screening for inhibitors of the gut hormone cholecystokinin type 1 receptor under the name "compound 362-63-08".[15]

Interestingly, *A. crenata*, the plant found to contain FR (Figure 2.3 B), has been used in traditional Chinese medicine for a long time. It is a low-growing evergreen shrub naturally occurring in south-eastern subtropical and tropical regions of Asia and invasive in Florida (USA).[22] The roots of *A. crenata*

have been used to treat respiratory tract infections, menstrual disorders, tonsillitis, toothaches, trauma and arthralgia. For that reason, *A. crenata* is listed in the Chinese Pharmacopeia.[23] Though FR was believed to be mainly produced in the leaves of *A. crenata*, a group recently reported the isolation of FR from dried roots of *A. crenata*.[24] Future research will have to show whether FR is the only bioactive compound resulting in this traditional medicinal use or if rather a variety of triterpenoid saponins found in high concentrations in *Ardisia* roots has led to the application of *A. crenata* for medicinal purposes.[25]

2.2.2 Discovery and structure elucidation of YM-254890 (YM)

In 2003, the cyclic depsipeptide YM-254890 (YM) was described by Taniguchi *et al.*, after isolation from the culture broth of *Chromobacterium* sp. QS3666, a strain isolated from soil collected at Okutama, Tokyo, Japan.[26] The structure of YM was elucidated using MS, 1D and 2D NMR studies; for the 2D structure with additional Marfey's analysis and chiral HPLC analysis for the absolute stereochemistry, as is shown in Figure 2.3 A, **2**.[27] It differs from the structure of FR only at two sites, i.e. one amino acid and one acyl group. Instead of β-hydroxyleucine (7), threonine is incorporated in the backbone, and in the side chain (1), an *N*-acetyl group replaces the *N*-propionyl residue as compared to FR.

This novel depsipeptide was discovered during a screening for new platelet aggregation inhibitors, and the first bioactivity tests revealed astonishing properties. Experiments with YM on G**α**q and Gαi signalling pathways suggested inhibition of Gαq, but not Gαi, by YM.[27] Following investigations confirmed the selective Gαq/11 inhibitory effect of YM by targeting the GTP/GDP exchange in Gαq/11 activation states.[14] These findings indicated that YM could be a valuable tool to study Gαq/11-related processes, similar to the established tool for Gαi-coupled pathways, Pertussis Toxin (PTX).[10]

2.2.3 Actual producer of FR discovered

FR was isolated from the plant *A. crenata* but its structure is non-typical for a plant metabolite. The bacterial origin of the highly similar YM also indicated that *A. crenata* may only store FR, but not primarily produce it.[27] It is long known that *Ardisia* species harbour endosymbiotic bacteria in their leaf nodules. Therefore it seemed obvious to search for a bacterial FR producer.[28,29] In 2015, Carlier *et al.* isolated bacterial DNA from the leaf nodules of *A. crenata* which allowed to identify the bacterial symbionts belonging to the genus *Burkholderia*. The endosymbiotic bacterium was named "*Candidatus* Burkholderia crenata" ("*Ca*. B. crenata") and genome sequencing revealed an eroded genome of only 2.85 Mb, common for obligate symbionts.[30] To date, it is the second smallest *Burkholderia* genome. Endosymbiontic *Burkholderia* range from 2.4 to 6.1 Mb while free-living varieties have an average of ca. 8 Mb.[31] So far the symbiotic bacteria could not be cultivated in the laboratory. On an extrachromosomal plasmid, a 34 kb, 8 open reading frame nonribosomal peptide synthetase (NRPS) biosynthetic gene cluster (BGC) termed *frs* was identified (Figure 2.5). It was hypothesised that the *frs* gene cluster may be responsible for the biosynthesis of FR, supported by coincidence of location of the endosymbionts and the highest FR concentration in the leave nodules of *A. crenata*.[30] FR biosynthesis

by *frsA-H* was confirmed in 2018 by achieving the heterologous production of FR in *E. coli*, although in minute amounts.[32]

2.2.4 New producer of FR: *Chromobacterium vaccinii*

Our group aimed at discovering cultivable bacterial producers of FR or similar compounds. The recent study of Hermes *et al.* describes the detection of several short DNA sequences with high identity to parts of the *frs* BGC via BLAST searches.[33] These sequences belonged to the draft genome of *Chromobacterium vaccinii*, a bacterial strain isolated from soil and roots of cranberry plants in Massachusetts, USA (Figure 2.3 C).[34,35] Using further Illumina genome sequencing and targeted PCR amplification, the complete *cv_frs* BGC from *C. vaccinii* could be assembled. Apart from slight differences in GC content and the length of the intergenic regions, the BGC is identical to *bc_frs* from "*Ca*. B. crenata" in terms of gene number, organisation and domain architecture. This was confirmed by the isolation of FR from the culture broth of *C. vaccinii* in yields of 2.5 mg/L,[33] adding another source for isolation of FR and paving the way for *in vivo* interrogation of FR biosynthesis.

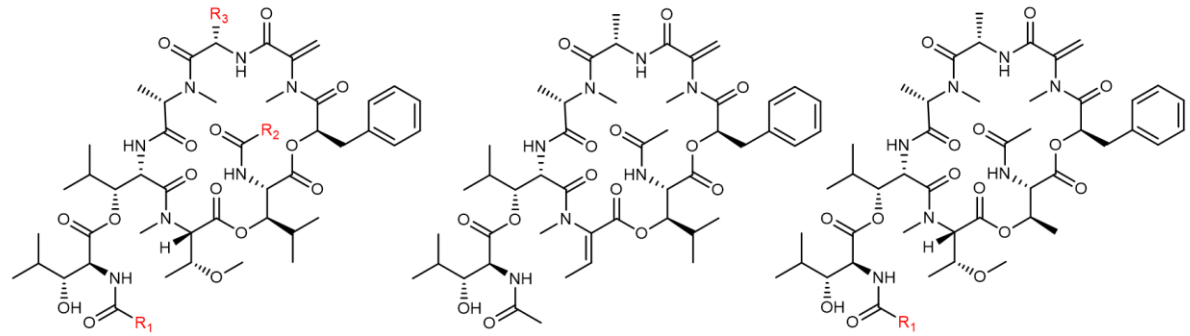
2.2.5 Naturally occurring FR and YM derivatives

Up to now, five natural FR analogues were isolated from *A. crenata* and *C. vaccinii*, indicating flexible substrate specificity of some biosynthetic enzymes (Figure 2.4). The first derivative, FR-1 (formerly termed AC-1, **4**), was isolated from the leaves of *A. crenata*. It has an altered *N*-acylation pattern at the hydroxyleucine (2) residue. The acetate unit is replaced by a 3-hydroxypropionate, an uncommon acyl residue for bacterial natural products. Its structure elucidation was performed by 2D NMR analysis and high-resolution MS/MS.[32] The same methods were used for FR-2 (formerly called AC-0, **5**), where the propionyl residue of the side chain (1) is replaced by an acetyl moiety.[11] Two more analogues, FR-3 (**3**) and FR-4 (**6**), again isolated from plant leaves, are isomers with the same molecular formula. FR-3, analogous to FR-1, has a propionyl replacing the acetyl residue at the hydroxyleucine (2) and represents the identical planar structure as sameuramide A, a natural product isolated from a didemnid ascidian described in 2.2.6.[19] FR-4 is the first derivative of FR with an altered amino acid. Here, L-alanine (4) is replaced by L-homoalanine. All these natural derivatives showed Gαq inhibition capacities in a range similar to FR.[36] Recently another natural FR derivative was isolated from *C. vaccinii*. FR-6 (**7**) has an acetyl residue in the side chain (1), and residue (8) is O-demethylated and dehydrogenated, resulting in an additional double bond. The G**α**q inhibiting activity of this compound is nine times decreased compared to FR (Results of Wiebke Hanke, manuscript in preparation).

With the help of MS/MS-based GNPS molecular networking[37] several further FR derivatives lacking the side chain (1) were annotated in methanolic extracts of *A. crenata.*[36] Unfortunately, the yields of these putative biosynthetic intermediates were not high enough for isolation and NMR structure elucidation. Only after the generation of the *frsA* deletion mutant of *C. vaccinii* was it possible to isolate the macrocyclic FR-Core (formerly called FR-SC, **8**) in preparative amounts and to verify its structure

by 2D NMR spectroscopy. FR-Core is 16-fold less active against Gαq proteins than FR,[33] underlining the importance of the *N*-acylhydroxyleucine side chain for effective Gαq inhibition.

Three natural YM analogues were isolated from *Chromobacterium* sp. QS3666 by Taniguchi *et al.* in 2004 (Figure 2.4) and their structures determined by MS and 1D- and 2D-NMR studies. YM-254891 (**9**) and YM-254892 (**10**) differ in the acyl residue at the side chain (1). **9** carries a propionyl and **10** a methylthioacetyl moiety, instead of the acetyl group in YM. Both analogues showed similarly potent Gαq inhibition as YM. YM-280193 (**11**), similar to FR-Core, is a YM derivative without the *N*-acetylhydroxyleucine side chain (1) and 40-fold less potent than YM.[21] These natural YM derivatives, varying in the acyl residue at the side chain, analogous to the natural FR derivatives, indicate a likewise biosynthetic route for this group of compounds. However, up to date, no BGC for YM has been published.



FR (**1** $R_1 = CH_2CH_3$, $R_2 = CH_3$, $R_3 = CH_3$)

FR-1 (**4** $R_1 = CH_2CH_3$, $R_2 = CH_2CH_2OH$, $R_3 = CH_3$)

FR-2 (**5** $R_1 = CH_3$, $R_2 = CH_3$, $R_3 = CH_3$)

FR-3/Sameuramide A (**3** $R_1 = CH_2CH_3$, $R_2 = CH_2CH_3$, $R_3 = CH_3$)

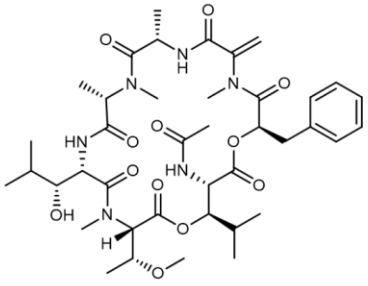
FR-4 (**6** $R_1 = CH_2CH_3$, $R_2 = CH_3$, $R_3 = CH_2CH_3$)

FR-6 (**7**)

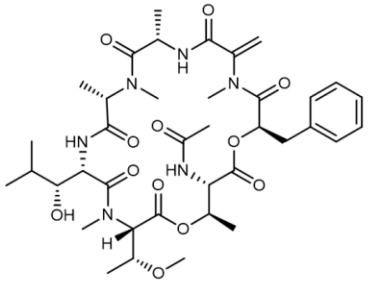
YM (**2** $R_1 = CH_3$)

YM-254891 (**9** $R_1 = CH_2CH_3$)

YM-254892 (**10** $R_1 = CH_{2S}CH_3$)



FR-Core (**8**)



YM-254893 (**11**)

Figure 2.4: Structures of naturally occurring derivatives of FR and YM.

2.2.6 Sameuramide A

In 2018 Machida *et al.* reported the isolation and structure elucidation of the cyclic depsipeptide sameuramide A, from a didemnid ascidian, collected at Sameura Bay at the north-west coast of Japan (Figure 2.3 D).[19] The taxonomy of the marine source organism was not further investigated in this study. The structure of sameuramide A is nearly identical to the structure of FR but contains a propionyl residue

replacing the acetyl at the hydroxyleucine (2) in the core peptide (see Figure 2.3, **3**). The planar structure is thus identical to FR-3, isolated from *A. crenata* leaves (section 2.2.5).[36] Sameuramide A showed some bioactivity in maintaining colony formation of murine embryonic stem cells without leukaemia inhibitory factor during a high throughput screening. There are no insights about a BGC or the actual producer of sameuramide A, but a bacterial origin is likely in the light of the different producers of FR, YM and their derivatives.[19]

2.3 Biosynthesis of FR

The NRPS BGC *frs*, responsible for the production of FR, was first described by Carlier *et al.* in 2015. It is located on the extrachromosomal plasmid pBCRE02 of the endosymbiont "*Ca*. B. crenata".[30] A second *frs* BGC with identical architecture was recently sequenced from the chromosome of the soil bacterium *C. vaccinii*.[33] Six of eight *frs* genes encode non-ribosomal peptide synthetase components, while the other two are tailoring enzymes (see Figure 2.5). NRPS are multimodular megaenzymes, that assemble peptides in a thiotemplated manner. A minimal NRPS module consists of adenylation (A), thiolation (T), and condensation (C) domains, to recruit and elongate the peptide chain with a specific building block.[38] The A domain activates this specific building block, which can be proteinogenic amino acids but also nonproteinogenic amino acids, fatty acids or α-hydroxy acids.[39] The activated substrate is attached to the adjacent T domain, which, in its *holo* form, carries a 4'-phosphopantetheine (Ppant) arm that binds the substrate and shuttles it to the next domain. The C domain catalyses the coupling to the upstream peptide chain. While the substrate is bound to the T domain it can be also shuttled to optional modifying domains, either being part of the NRPS or as *trans*-acting tailoring enzymes that add further modifications. Typical modifying NRPS domains are methylation (MT), epimerization (E), formylation (F), heterocyclization (Cy), reduction (R), and oxidation (Ox) domains.[38] There can be any number of modules following the first module, forming a growing peptide chain that is usually terminated by the final thioesterase (TE) domain. The TE catalyses the hydrolytic release of the assembled peptide from the mega enzyme, either as a linear peptide or, after intramolecular esterification, as a macrolactone or -lactam.[40] Bioinformatic analysis of the *frs* NRPS revealed eight adenylation (A) domains, corresponding to the number of amino acid building blocks of the FR backbone.[30] The presence of two thioesterase (TE) domains indicated two distinct NRPS machineries, one with seven A domains (FrsD-FrsG) hypothesised for the synthesis of the cyclic core molecule and one with only one A domain (FrsA) for the synthesis of the side chain.

For FrsC, bioinformatic analysis revealed similarities to malate and L-lactate dehydrogenases. It was proposed that FrsC catalyzes the reductive formation of L-phenyllactate from phenylpyruvate, an intermediate from phenylalanine metabolism. As D-phenyllactate is the building block present in FR, FrsC would provide the substrate for the A domain that is epimerized on the E domain of FrsE.[32] This biosynthetic route has been recently confirmed through analysis of the *in vitro* reaction product of FrsC and experimental characterisation of the FrsE3 A domain (unpublished).

The other modifying enzyme, FrsH, does not exhibit high sequence homology to any characterized protein. However, a more detailed bioinformatic search showed homologies to the active centre and overall tertiary structure of the non-heme diiron monooxygenase CmlA.[32] CmlA catalyses the β-hydroxylation of the thiolation domain (T)-bound L-*para*-aminophenylalanine in the biosynthesis of the antibiotic chloramphenicol.[41] In analogy to chloramphenicol biosynthesis, it was suggested that the hydroxylation of leucines in FR could take place on the leucinyl-T domains of FrsA, FrsD and the first module of FrsG. Experiments showing preferred activation of L-leucine by the A domains of FrsA and FrsD support this hypothesis.[32,33] Recent investigations with heterologously expressed FrsH confirm the proposal: In an enzymatic *in vitro* assay, FrsH combined with FrsA produced *N*-propionylhydroxyleucine, proving its proposed function.[33]

The smallest gene within the BGC, i.e. *frsB* encodes an MbtH-like protein (MLP).[30] MLPs are small, highly conserved proteins that are frequently associated with bacterial NRPSs and are in many, but not all cases crucial for A domain activity or solubility.[42] The heterologously expressed modules FrsA and FrsD from *cv_frs* were soluble and showed adenylating activity only when coexpressed with FrsB.[33] Unpublished results of our group indicate that also the A domains FrsE3 and E4 likewise are only active when coexpressed with FrsB, highlighting the importance of FrsB for FR biosynthesis.

FrsA is a monomodular NRPS consisting of a starter condensation (C) domain, an A, T and TE domain. It was proposed that FrsA synthesises the *N*-propionylhydroxyleucine side chain and that the TE domain catalyses an unusual intermolecular transesterification attaching it to the free β-hydroxy moiety of the cycloheptapeptide core molecule.[32] So far, a similar mechanism was only described for the biosynthesis of salinamides from a marine *Streptomyces*, performed by a hybrid NRPS/PKS system.[43] As outlined above, after leucine activation and loading onto $FrsA_T$, FrsH conducts the β-hydroxylation on leucinyl-$FrsA_T$. The C domain of FrsA is phylogenetically related to so-called Starter C domains ($C_{starter}$), known to conduct a transfer of acyl units onto the first amino group of a peptide chain.[32,44] The *in vitro* production of *N*-propionylhydroxyleucine by purified FrsA/B and FrsH confirmed this prediction. Additionally, bioassays without FrsH gave only traces of *N*-propionylleucine, implying that ß-hydroxylation of leucinyl-$FrsA_T$ had to take place before acylation. To investigate the unusual TE domain, different approaches were chosen including the deletion of the whole *frsA* in *C. vaccinii*. This deletion mutant stopped to produce FR and instead overproduced FR-Core (**8**), an analogue lacking the side chain, that was also found in traces in extracts of *A. crenata* and *C. vaccinii* (see above). This not only confirmed the function of FrsA, but also provided enough material of FR-Core for structure elucidation and further assays. FR-Core was used as substrate to prove the side chain transesterification catalysed by the FrsA TE domain with *in vitro* synthesized *N*-propionylhydroxyleucine. The conducted assays were used for the analysis of substrate specificity of the domains: The A domain of FrsA also accepted D-leucine and L-isoleucine and the C domain could incorporate acetyl-CoA as well as butyryl-CoA into the side chain. The TE domain was able to transfer these altered side chains upon FR-Core

too, leading to one already known (FR-2) and one completely new FR derivative (FR-5) (see Figure 2.6, **19**).[33] This opens possibilities for the production of novel bioactive molecules by precursor-directed biosynthesis and engineering of FR biosynthesis.

FrsD is the first module responsible for the assembly of the cyclic heptapeptide core of FR. The DNA sequence encoding these three domains is 94.3% identical to the sequence encoding the C, A and T domains of FrsA. We thus suggested a duplication event during evolution of the *frs* BGC leading to the production of an additional acylated hydroxyleucine moiety and its incorporation into the natural product. Unlike FrsA, the FrsD $C_{starter}$ domain preferably transfers an acetyl moiety onto hydroxyleucine during FR synthesis. *In vitro* assays with expressed FrsD showed its ability to accept acetyl- as well as propionyl-CoA as a substrate and, in combination with the FrsA TE, to transfer its assembled product onto FR-Core, even though the acetyl residue is clearly preferred.[33]

FrsE contains the third and fourth NRPS module. The Stachelhaus code of the first A domain indicated a carboxylic acid as a substrate.[30,45] Therefore this domain was hypothesised to activate L-phenyllactate produced by FrsC and the epimerisation (E) domain, which appears by bioinformatics analyses to be functional, would then catalyze the epimerization to D-phenyllactate,[32] which was recently shown in our laboratory (unpublished). The fourth module assembles *N*-methyldehydroalanine and so far little is known about the biosynthesis of this rare building block. The nearest Stachelhaus code for this A domain, however, refers to serine.[30] We thus hypothesised that the C5 domain performs dehydration of serine with a mechanism comparable to the nocardicin NRPS.[46] A methyl transferase (MT) domain is then predicted to perform N-methylation of dehydroalanine.[30]

FrsF with the fifth and sixth module is supposed to incorporate one common L-alanine and one *N*-methylated L-alanine. An MT domain is located after the second A domain, fitting the predicted methylation of L-alanine.[32] The A domain of module seven has the same Stachelhaus code as the A domains of FrsA and FrsD and is supposed to incorporate β-hydroxyleucine.[30] Until now, FrsF and FrsG were not investigated further.

The eighth and last module of the FR biosynthesis is responsible for the incorporation of *N*-*O*-dimethylthreonine and the cyclisation of the core molecule. Here, two MT domains are interrupting the A domain, presumably responsible for the *N*- and *O*-methylation of this residue.[30] There are only a few known A domains with two interrupting MT domains and just the A domain of FrsG is thereby divided into three parts, which is, to our knowledge an unprecedented architecture.[47] The TE domain of FrsG is supposed to catalyse the intramolecular cyclization and offloading of FR-Core by the formation of an ester bond between threonine and the hydroxy group of the first β-hydroxyleucine (see Figure 2.5). Evidence for this theory was provided by the *frsA* deletion mutant of *C. vaccinii*, producing high amounts of this intermediate.[33]

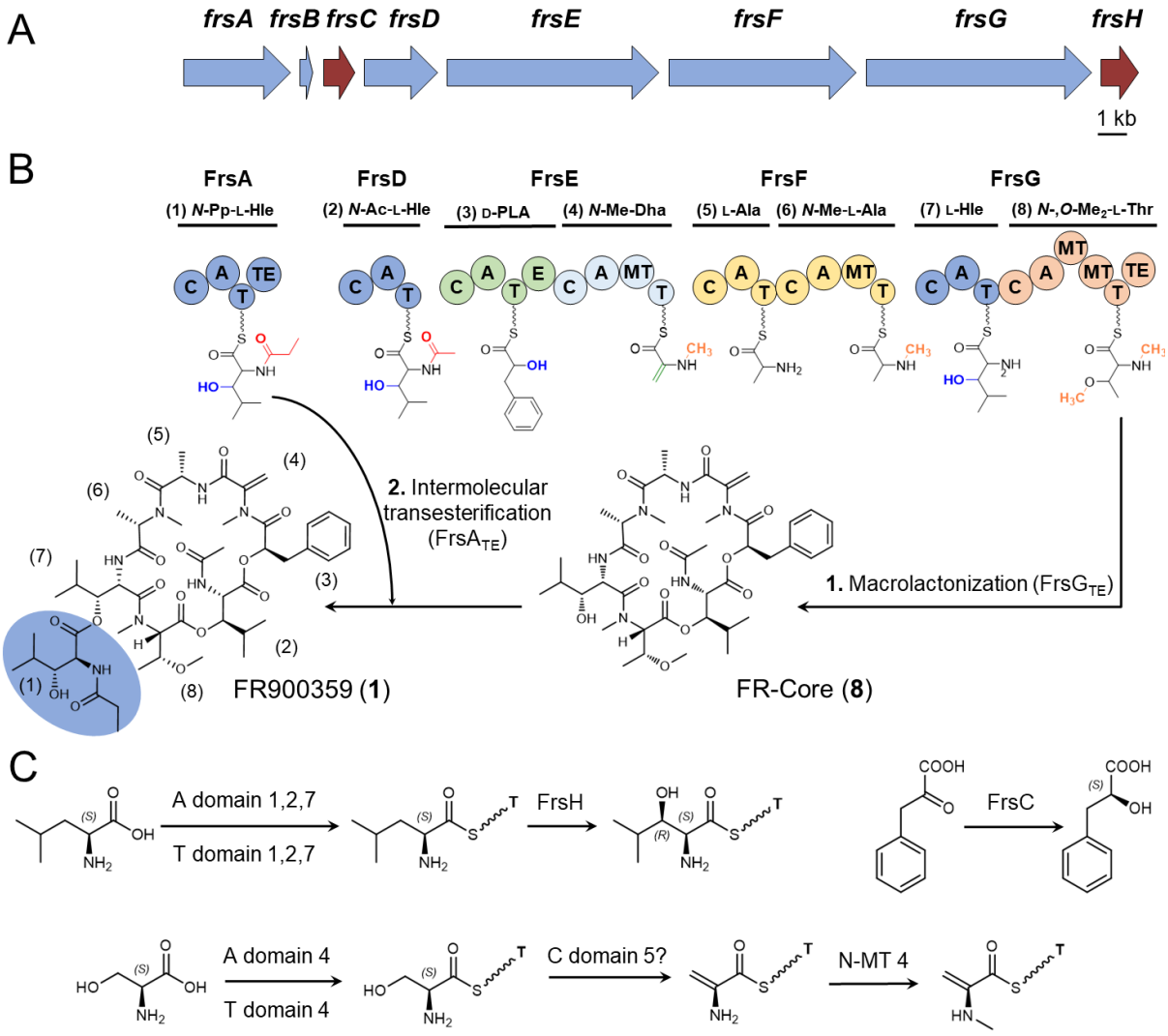


Figure 2.5: Biosynthesis of FR. A. Organisation and size of the *frs* BGCs from "*Ca.* B. crenata" and *C. vaccinii* MWU205 (blue = NRPS, red = modifying enzyme) **B.** Proposed biosynthetic pathway of FR900359. In a first step (1.) the NRPS assembly line FrsDEFG forms a seven-membered linear peptide chain which is then cyclized by the FrsG TE domain to FR-Core. Following this (2.) the FrsA TE domain catalyzes the intermolecular transfer of the *N*-propionyl hydroxyleucine side chain onto FR-Core to yield FR. **C.** Proposed biosynthesis of non-proteinogenic building blocks. Hle = Hydroxyleucine, PLA = Phenyllactic acid, Dha = Dehydroalanine. C = condensation domain, A = adenylation domain T = thiolation domain, TE = thioesterase domain, E = epimerase domain, MT = methyltransferase domain.

2.4 Total synthesis of FR, YM, derivatives and structure-activity relationship studies

Due to their structural complexity, both FR and YM are highly challenging synthetic target molecules. In 2012, a worldwide challenge promised $100,000 for the total synthesis of 1 mg YM, but it was not met in time.[12] In 2015, the total synthesis of WU-07047, a simplified YM analogue with lower Gαq inhibition potency, was reported.[48] Also, the total synthesis of YM-280193, the YM derivative without side chain was achieved in the same year. Here, the monomer and dipeptide fragments were prepared using conventional chemistry and subsequently assembled by Fmoc-solid-phase peptide synthesis.[49] In 2016 the total synthesis of both FR and YM was finally reported by the Strømgaard group, using a

combination of solution-phase synthesis for depsipeptide building blocks and solid-phase approaches. This synthesis also confirmed the original structural assignment of the corresponding natural products.[16]

Next, the Strømgaard group synthesized several YM derivatives for structure-activity relationship (SAR) studies to provide information on the structural requirements for inhibition of Gαq signalling. YM-1 to YM-35 are comprehensively listed and compared in the review of Zhang *et al.*[12] Taken together, these SAR studies showed that any kind of structural variation at most positions of the molecule led to a drastic loss of potency, however with the following exceptions: Hydrogenation of dehydroalanine (4) leading to D-Ala in YM-385781 (**12**); the variation of *N*-Me- L-Ala (6) to *N*-Me- D-Ala in YM-13 (**13**) or to *N*-Me- L-Phe in YM-14 (**14**) and the exchange of L-Ala (5) to L-Phe in YM-18 (**15**) (see Figure 2.6 A). This indicated some structural flexibility in these positions and the respective derivatives showed IC_{50} values comparable to those of YM, but up to date, no synthetic derivative with higher potency than YM has been generated so far.[12]

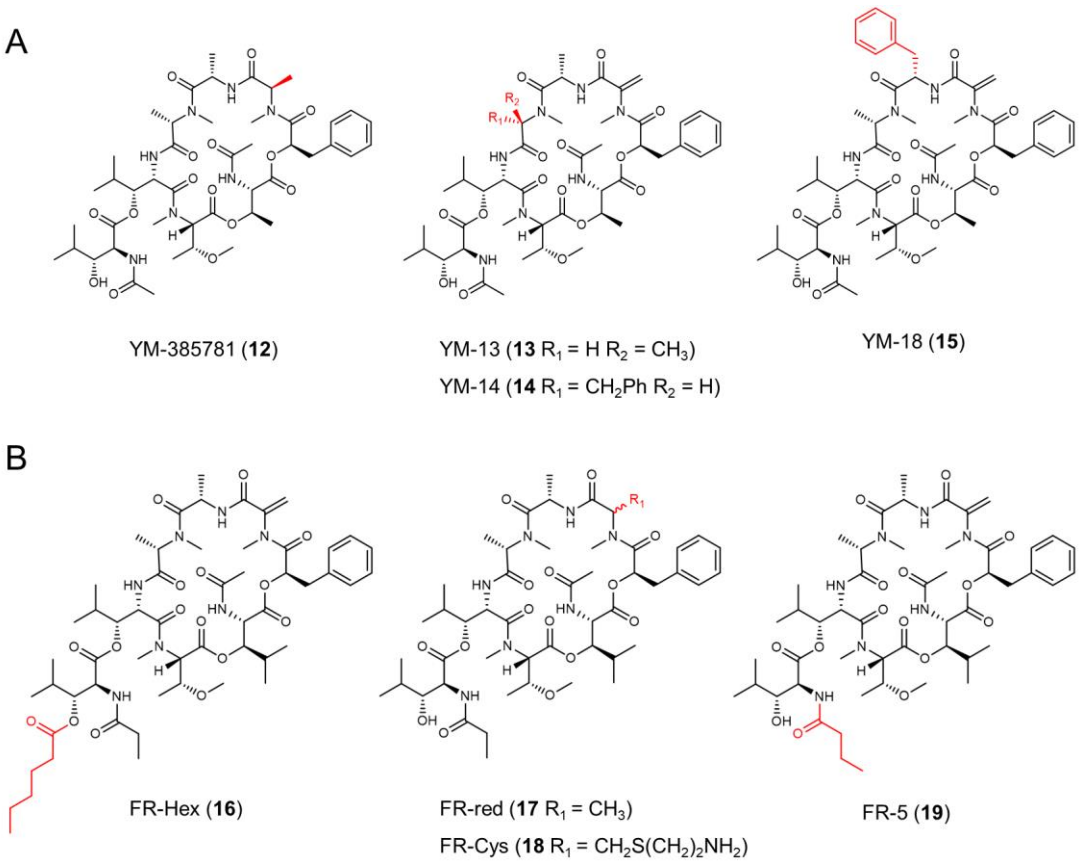


Figure 2.6: Synthetic derivatives of YM and FR. A. YM derivatives with comparable potency to YM. **B.** Structures of semi-synthetic derivatives of FR.

The (semi)synthetic analogue of FR, FR-red (**17**), was generated by hydrogenation of the exocyclic double bond, analogous to the semi-synthesis of YM-385780 and YM-385781 (**12**) from YM. FR-red is a mixture of the two possible stereoisomers and was only slightly less potent in Gαq inhibition than FR.[1] Another two semi-synthetic derivatives were described in 2018, i.e. FR-Hex (**16**) was obtained by hexanoylation of the hydroxyl group of FR side chain Hle (1), whereas FR-Cys (**18**) was synthesized by

Michael addition of 2-aminoethanethiol hydrochloride to FR to form the corresponding thioether at the double bond of *N*-Me-Dha (4). FR-Hex showed an IC_{50} value of approximately 90 µM and FR-Cys was completely inactive toward Gαq, proving these positions difficult for structural variation as well.[11] For FR-Cys the inactivity might also be due to the presence of a primary amine, which is charged at physiological pH values. In 2020 another derivative was produced by precursor-directed biosynthesis: Feeding of butyric acid to *C. vaccinii* resulted in the formation of FR-5 (**19**), carrying a butyryl moiety instead of propionyl at the side chain (1). This chain elongation by one methylene group resulted in a 7-fold loss of activity against Gαq compared to FR. The structures of the semi-synthetic derivatives of FR are shown in Figure 2.6 B. Taken together SAR studies for FR and YM may invoke the conclusion, that the structure of this molecule was optimised by nature for Gαq inhibition. The predominant part of the molecule seems to represent the "pharmacophore", with some exceptions for the building blocks incorporated into the depsipeptides by modules (4), (5) and (6.) Even small changes in the structure can lead to a dramatic loss of bioactivity and so far there is no synthetic analogue developed with higher potency than the naturally occurring structures.

2.5 Ecological and Evolutionary Aspects

Natural products with the chromodepsin scaffold have been found in different ecosystems like plant leaf nodules,[13] soil,[27,33]and the marine habitat,[19] leading to the question of the ecological role of these depsipeptides in their natural environment. Up to date, only for FR and derivatives thereof some aspects of their function in nature have been investigated, especially in the context of the *A. crenata* bacterial symbiosis. Schrage *et al.* tested FR on the plant G protein Gpa1 to see if it has any regulatory effects on plant G protein signalling. As expected, FR did not affect the kinetics of Gpa1 nucleotide exchange or thermal stability of the G proteins, in contrast to its inhibitory effect on mammalian Gαq signalling.[1] This is not surprising, taking into account that G protein-dependent signalling in plants has taken a very different evolutionary path.[50] Carlier *et al.*, who clarified FR to be a bacterial secondary metabolite, suggested a protective role for the host in its symbiotic relationship, e.g. as herbivore deterrent.[30] This hypothesis was fortified in 2018, revealing that the oral uptake of FR in mice resulted in a significant reduction of blood pressure (Figure 2.7 B). Additionally, insect toxicity was tested, which showed the killing and prevention of moulting of bean bug nymphs, as well as the high affinity to Gαq proteins of the pest insects *Bemisia tabaci* and *Bombyx mori* (Figure 2.7 C). It was thus proposed that the host plant *A. crenata* is protected from a large range of enemies by FR, and the metabolically very limited bacterium "*Ca.* B. crenata" profits from primary metabolites of the plant. Additionally, this study showed the distribution of FR in *A. crenata* leaves using MALDI imaging, which correlates perfectly with the location of the endosymbiotic bacteria in the leaf nodules (see Figure 2.7 A).[32] The symbiont is transmitted vertically during the live cycle of the plant. Additional to the dead-end leaf nodules, the symbionts are present in the buds, fruits and seeds to inoculate the next plant generation.[31] Interestingly, Reher *et al.* proved the production of FR in an *Ardisia* species lacking specific leaf nodules *A. lucida*. This species might represent a very early stage of a plant-bacteria symbiosis, yet without specific

morphological features. FR was also found in four other nodulated *Ardisia* species, *A. hanceana*, *A. villosa*, *A. mamillata*, and *A. crispa*, proving the wide-spread occurrence of chromodepsins in this genus.[36] Together with the discovery of sameuramide A in a marine tunicate,[19] and YM and FR also in soil bacteria,[27,33] this implies a broad distribution of the FR molecular family in nature. This could be related to an important ecological role of chromodepsins in several contexts.[36]

The suggestion, that the chromodepsin structure was highly optimized for specific Gαq protein inhibition,[11] leads to questions concerning the evolution of these compounds. Hermes *et al.* compared FR and the biosynthetic intermediate FR-Core (**8**, Figure 2.4) in different ecologically relevant bioassays. FR-Core, in direct comparison to FR, is 16-fold less potent in inhibition of human Gαq. For a significant conclusion on the ecology, the assay needed to be repeated with Gαq proteins of potential predators like insects. However, competition binding assays against a radiolabeled FR-derivative revealed a 207-fold decrease in binding affinity for FR-Core in human platelet membrane. FR-Core was not tested for the affinity against insect cells, but for FR the affinity was similar for human (pIC_{50} = 7.88)[33] and insect (pIC_{50} = 8.13–9.27)[32] Gαq proteins, compare Figure 2.7 C and E. To investigate ecological relevance of the side chain, FR and FR-Core were fed to nymphs of a stinkbug: While 0.2 µg/µl of both metabolites killed all insects after nine days, in lower concentrations only feeding of FR led to the death of insects, while FR-Core did not affect the animals (Figure 2.7 F). This demonstrated the enhanced *in vivo* toxicity of FR compared to FR-Core.

On a genetic level, bioinformatic analyses and the sequence comparisons of the NRPS modules FrsA and FrsD suggested a gene duplication event during evolution of the FR biosynthesis. Phylogenetic analysis revealed no close relative to the $FrsA/D_{Cstarter}$ or $FrsA_{TE}$ that would hint to horizontal acquisition of these genes. Furthermore, a global bioinformatics analysis with BiG-FAM[51] and BiG-SCAPE[52] reveals no closely related BGC to the two *frs* in the 1.2 million BGC present in the databases.[33] We consequently hypothesised that FR-Core may have been the less active ancestor molecule of FR, optimized via the addition of a side chain enabled through duplication events during evolution of the BGC. This scenario would be conform with recently published evolutionary frameworks on natural product evolution,[53,54] considering improved Gαq inhibition as trait for positive selection of the BGC. Remarkably, the natural derivative YM-280193 (**11**), which is the YM equivalent without side chain, is also a less potent Gαq inhibitor than YM, reassuring this hypothesis.[21] Ultimately, sequencing of the YM and sameuramide BGCs may lead to further insights into the evolution of this molecular family.

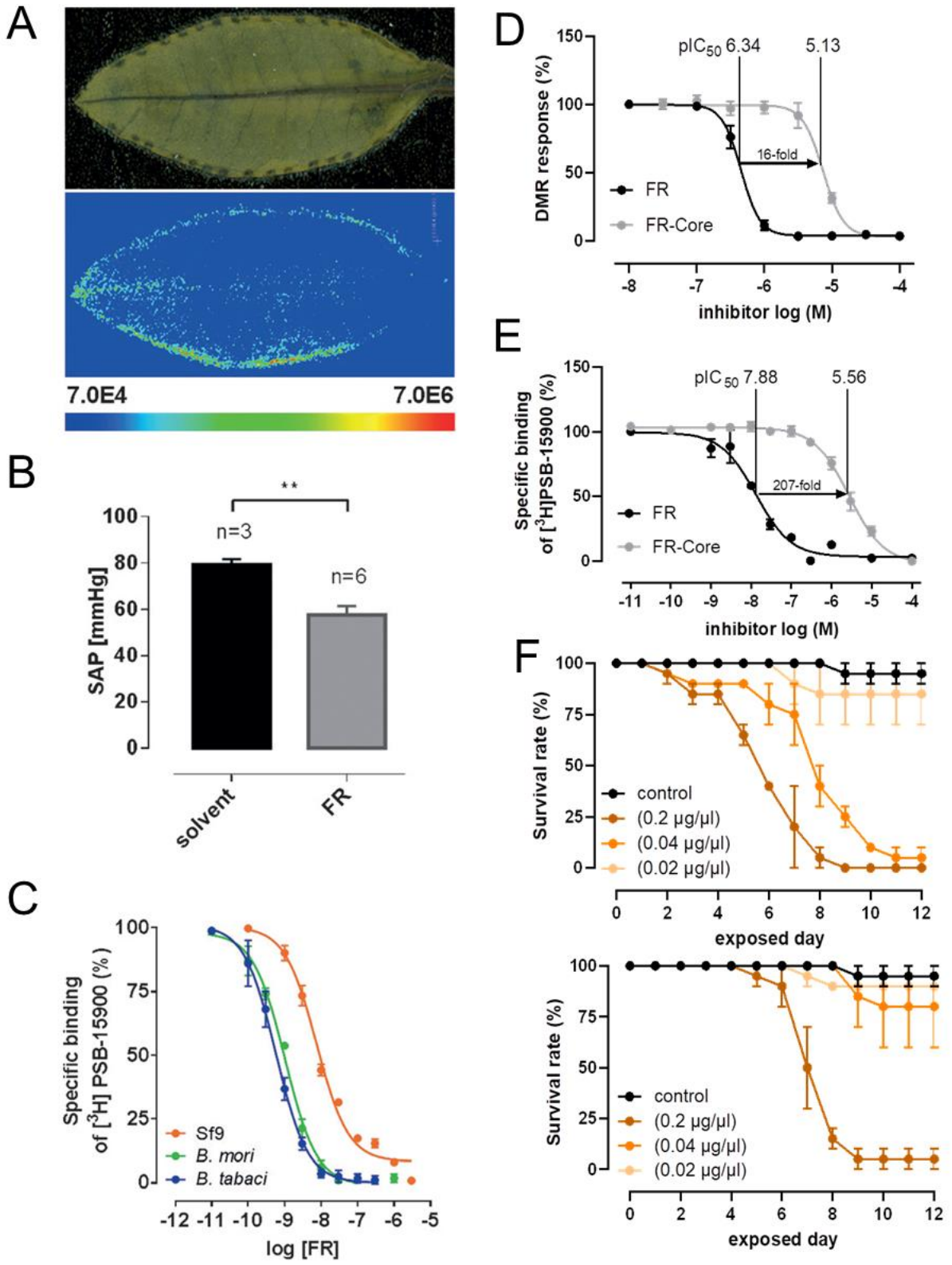


Figure 2.7: Evaluating the ecological function of FR. A. MALDI imaging mass spectrometry of an *Ardisia crenata* leaf. The potassium adduct of FR (*m/z* 1040.49) is highlighted according to intensity. **B.** Statistical analysis of *in vivo* blood pressure recordings in mice revealed a strong reduction of systolic arterial pressure (SAP) in the aorta 1 h after oral FR application. **C.** Competition binding study of FR versus [3H]PSB-15900 (5 nm), the tritiated derivative of FR, at Sf9 insect cell membranes, and at Gαq proteins of *Bombyx mori* and *Bemisia tabaci* expressed in Gαq -knockout HEK cell membrane preparations. Values represent means ± SEM of three independent experiments. pIC50 values of 8.13–9.27 were determined. **A-C** was taken from Crüsemann *et al.*[32] **D.** Concentration-dependent inhibition of activated $G\alpha_q$ proteins by FR and FR-Core as determined by label-free whole cell DMR biosensing. DMR recordings are representative (mean + s.e.m.) of at least four independent biological replicates conducted in triplicate. **E.** Competition binding experiments of FR and FR-Core versus the FR-derived radiotracer

[^{3}H]PSB-15900 at human platelet membrane preparation (50 µg protein per vial), incubated at 37 °C for 1 h. **F.** Exposure of nymphs of a stink bug (*Riptortus pedestris*) to different concentrations of FR (top) and FR-Core (bottom), survival rate was measured. D-F was taken from Hermes *et al.*[33]

2.6 Pharmacology of FR and YM

2.6.1 FR and YM are Gαq protein inhibitors

Initial pharmacological studies[14,26,27,55] indicated that YM is not only a strong and selective P2Y1 antagonist but also actively inhibits other Gαq/11-coupled receptors, meanwhile having little effect on Gαi-mediated Ca^{2+} mobilization. More detailed *in vitro* and *in vivo* experiments suggested that YM targets the exchange step of GDP for GTP in Gαq/11 activation states, meaning that YM functions as a guanine nucleotide dissociation inhibitor (GDI).[14] At this point, it became clear that YM and derivatives thereof might be used as promising tools for studying Gαq/11 protein activation, Gαq/11-coupled receptor signalling, and Gαq/11-mediated biological events.

In 2010 the molecular mechanism of action of YM was unravelled by utilisation of a [^{35}S]GTPγS binding assay with Gαq protein, and the subsequent determination of an X-ray crystal structure of the Gαi/qβγ–YM complex with a 2.9 Å resolution.[2] In this publication, Nishimura *et al.* unambiguously demonstrated that YM blocks GDP dissociation from Gαq in a concentration-dependent manner, while no effect occurs with Gαs, Gαi1, Gαo and Gα13. The crystal structure revealed that YM is bound to Gαq in near distance to the binding site of GDP in the inactive state with no direct contact to Gβ or Gγ. YM is localised between linker 1 and linker 2, where it docks into a hydrophobic cleft formed by the GTPase and the helical domain of Gα (analogously to the structure with docked FR in Figure 2.8). Bound there, YM is believed to stabilise the inactive conformation and to prohibit the hinge motion of linker 1 and linker 2, thus inhibiting the rearrangement to the GTP-bound active state.[2]

For FR, no crystal structure complexed with a G protein is reported so far. But, five years after the publication of the Gαi/qβγ–YM structure, Schrage *et al.* performed the first detailed investigations into FR as a Gαq protein inhibitor and concluded after different experiments, that FR has the same principal mode of action as YM.[1] Since then, there have been numerous studies to further analyse and determine the potential of FR and YM as selective Gαq inhibitors. as well as different structure-activity relationship studies, recently reviewed by Zhang *et al.* [12] Thus, we will not discuss this topic here in detail.

A recent publication, however, gave striking new insights in terms of the different kinetics between FR and YM. Kuschak *et al.* synthesised tritium labelled probes by hydrogenation of the exocyclic double bond of both FR and YM, of which the L-configured isomer showed high-affinity binding to Gαq.[56] These new radiotracers [^{3}H]PSB-16254, derived from YM, and [^{3}H]PSB-15900, derived from FR, were used for kinetic and molecular docking studies. [^{3}H]PSB-15900 showed an extraordinarily slow dissociation rate and was therefore characterized as a pseudoirreversible Gαq binder, while the dissociation of [^{3}H]PSB-16254 was quite rapid. Based on the molecular docking studies, the authors suggested, that this is due to FRs additional lipophilic “handles”. FR seems to be anchored in the binding

pocket like a dowel forming a latch, while YM lacks those anchor points and is, therefore, more readily released, leading to the conclusion that YM and FR behave unexpectedly very different and therefore the pharmacological effects of these structurally highly similar compounds differ. While FR could be advantageous for applications with long residence time at Gαq, YM could be preferred for experiments where fast reversibility is necessary.[56]

While previous NMR studies had shown that YM forms two major conformers in water,[57] a new study from Tietze *et al.* classified these structures as the organic solvent-derived *trans*-YM and the water-derived *cis*-YM conformations (referring to the predominant *cis* or *trans* conformation of the peptide bonds). In this study, the new structural information from the NMR data were used to reanalyse the crystal structure of Gαq bound YM. After re-evaluation of fitting parameters and X-ray data analysis, it was concluded that both conformers have the same or nearly the same fit, which would imply that the *cis* conformer of YM is also a valid representation of the crystal structure. Molecular dynamic binding studies revealed however much shorter residence times in the Gαq binding pocket of the *trans* compared to the *cis* isomer. Thus, the conformational stability of the inhibitor seems to be of high importance for effective G protein inhibition. FR instead has only the *cis*-conformation in water, which could be another explanation for the approximately three times higher Gαq inhibitory activity of FR compared to YM.[16] Taken together, these data combined with previous studies give further insights into the extremely tight SARs observed for FR and YM.[6]

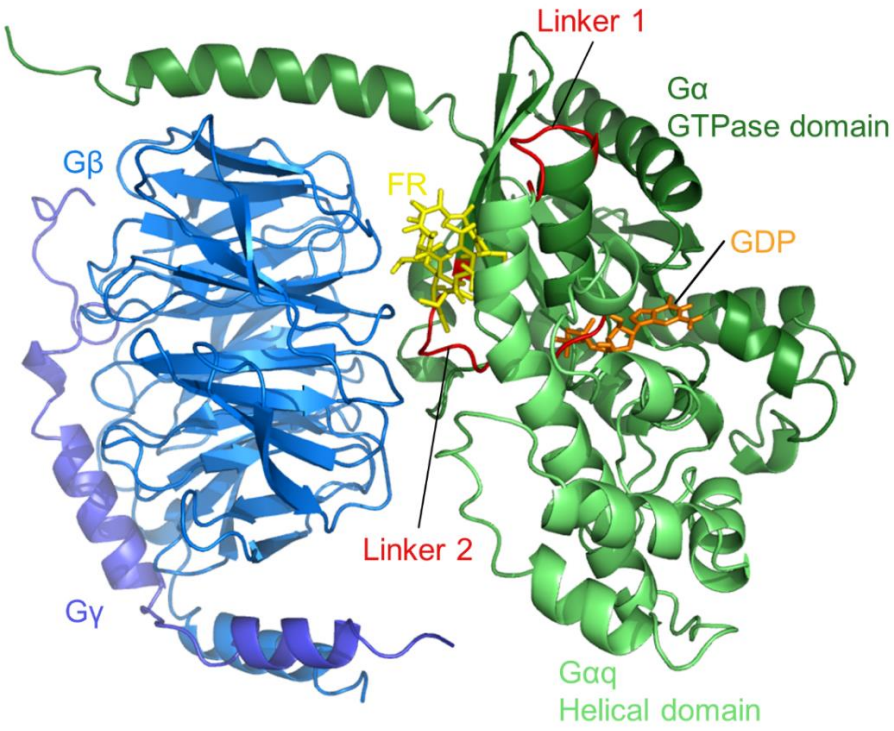


Figure 2.8: Docked pose of FR to the crystal structure of the heterotrimeric Gαqβ1γ2 protein in its GDP-bound state (PDB-ID:3AH8,[2] FR docked by Kuschak *et al.*[56] depicted by Jan H. Voss). Gαq consists of the GTPase (dark green) and the helical (light green) domains connected by two linker regions (red). Gβ and Gγ are blue and purple, respectively. GDP (orange) and FR (yellow) are shown as stick models.

2.6.2 Drug development

Due to their distinct mode of action, interest in FR and YM as lead structures or drugs themselves arose. These compounds are especially of interest for diseases where multiple GαqPCRs contribute to pathology.

2.6.2.1 Cardiovascular system

In a first pharmacological investigation, FR was found to decrease blood pressure in rats and to inhibit platelet aggregation *in vitro* and *ex vivo* in rabbits, demonstrating an effect on the cardiovascular system.[58] YM was also identified in a screening for new platelet aggregation inhibitors.[27] The antithrombotic and thrombolytic effects of YM in an electrically-induced carotid artery thrombosis model were examined in rats shortly thereafter.[55] In further studies, systemic administration of YM inhibited not only acute thrombosis but also neointima formation after vascular injury. In this study, however, YM was found to have a narrow therapeutic window, making its further systemic use questionable.[59] Thus, Uemura *et al.* investigated the local administration of YM on the experimental peripheral arterial disease in rats. YM was found to exert a more pronounced beneficial preventive effect in a severe peripheral arterial disease model in rats after *i.a.* administration than conventional drugs, without inducing hypotension. Also, it inhibited lesion progression in a laurate induced peripheral arterial disease model in rats.[60] In parallel, the group examined the effect of YM on platelet functions and thrombus formation under high-shear stress, showing it to be effective in this model as well.[61]

After YM was no longer available, a Japanese workgroup rediscovered FR in 2011 and verified its vasorelaxant effect on rat aortic arteries. In their study, FR showed an inhibitory effect on voltage-dependent and receptor-dependent Ca^{2+} influx, the latter playing a major role in the vasorelaxant activity of FR.[18] Shortly thereafter, Inamdar *et al.* characterized FR as a Gαq inhibitor in human platelets, showing inhibition of platelet aggregation as known for YM and mentioned above.[62]

A few years later, Meleka *et al.* tested FR and YM in parallel to block G protein-dependent vasoconstriction in mice. They concluded that Gαq/11 inhibitor ligands block vasoconstriction partly by directly inhibiting L-type calcium channels in vascular smooth muscle cells, which is in line with the early findings of Zaima *et al.*.[18,63] Additionally, they used FR to demonstrate the anti-hypertensive potential of chronically blocking Gαq/11 in a mouse model of established hypertension. Chronic administration of FR was sufficient to quickly and effectively reduce blood pressure to normal levels.[63]

2.6.2.2 Airway diseases

Common lung diseases like asthma are often caused or affected by aberrant activation of Gαq protein-dependent signalling. As various GαqPCRs are involved, targeting specific GPCRs has only limited effects. Carr *et al.* hypothesised that a compound inhibiting Gαq activation at the receptor or G protein level would be an advantageous asthma therapeutic, as Gαq-mediated airway smooth muscle (ASM) shortening is a primary contributor to bronchoconstriction. As predicted, FR was able to significantly

impact airway contraction in *ex vivo* human precision-cut lung slices. Also, FR inhibited synergistic ASM growth and synergistic Gβγ-dependent AKT activation.[64] One year later, a more detailed study, including *in vivo* experiments was published by Matthey *et al.*. The authors confirmed the selective Gαq inhibition of FR in mouse and human airway smooth muscle cells and proved it to be a strong bronchorelaxant in mouse, pig and human *ex vivo*. When applied in healthy and sensitized mice *in vivo*, they found that inhalation of FR prevented the elevation of airway resistance without acute effects on blood pressure or heart rate. Besides, FR also suppressed airway hyperreactivity and even early airway remodelling, but not inflammation in mouse models of airway hyperresponsiveness.[65]

2.6.2.3 Obesity

The research of Klepac *et al.* revealed the regulation of brown and beige adipocytes by Gαq signalling using FR. Treatment with FR enhanced the adipogenic and thermogenic potential of brown adipocytes, which indicated the involvement of Gαq. An autocrine loop of Gαq signalling or the presence of a constitutively active Gαq might be the reason for FR activity. These findings highlight the importance of Gαq in brown/beige adipogenesis. The authors concluded that the selective inhibition of Gαq signalling might be a novel approach to deal with obesity by enhancing the amount of brown/beige fat and thus increase energy expenditure.[66]

2.6.2.4 Cancer

The first investigations of FR as a potential cancer drug were performed by Schrage *et al.* in well-established B16 melanoma cells. FR showed no direct cytotoxicity and did not compromise mitochondrial metabolism, but inhibited cell proliferation without causing cell death. Additionally, FR forced melanoma cells into differentiation and inhibited cell migration. Interestingly enough, the inhibition of cell growth by FR was due to a G1 cell cycle arrest which did not lead to apoptosis.[1] The authors suggested that FR could not only be used for cancer treatment but could also prevent metastasis if Gαq is involved in this process. Later, more detailed investigations on the treatment of a specific form of eye cancer were performed.

Uveal melanoma (UM) is the most common form of intraocular cancer; its oncogenic drivers are predominantly mutated, constitutively active forms of Gαq or Gα11 like the Gαq Q209L mutant or the Gα11 mutants, Q209L and R183C.[67] In 2017, FR was first mentioned as a potential treatment for this disease.[68] Three workgroups then simultaneously investigated the therapeutic potential of FR and published their results in 2018 and 2019 leading to broad and detailed insights into this topic.[24,69,70] All studies proved FR to effectively inhibit Gαq Q209L, the oncogenic guanosine triphosphatase–defective Gαq mutant common in UM, and thus to hinder the proliferation and survival of UM cells. These studies and *in vivo* experiments for UM treatment are discussed in more detail in another recent review.[9] Remarkably, YM was also once tested against oncogenic $G\alpha_{11}$ mutants, Q209L and R183C in HEK293 cells, but showed only inhibition of the accumulation of IP_1 in expressed $G\alpha_{11}$-R183C, but not in Q209L.[16] A recent subsequent study of Onken and Cooper dealt with transendothelial migration (TEM),

a key step in the formation of metastases in UM. Inhibition of constitutively active Gαq/11 in UM cells by FR, led to a nearly complete loss of TEM activity in their assay system, confirming the promising value of FR as therapeutic agent for UM tumours.[71] Altogether, these studies give perspectives not only for future treatment options for UM but also for diseases associated with other constitutively active Gαq mutants.

2.7 FR and YM as Pharmacological Tools

During the last five years, FR has been proven to be a valuable tool for the investigation of the involvement of Gαq signalling in different processes by the ability to specifically inhibit these signalling pathways. This resulted in over 50 publications benefitting from the use of FR.[72–124] To give a few examples, Chang *et al.* investigated the adhesion GPCR GPR56/ADGRG1 as an inhibitory receptor on human natural killer (NK) cells. They used FR to test whether Gαq/11 activity is required for this process, but cytotoxicity was not restored in NK-92–GPR56 cells, proving that the signalling capacity of the GPR56-CD81 complex in NK cells does not rely on the engagement of Gαq proteins.[78] In another study, FR was used to show the specificity of Gαq coupling for the nucleotide receptor $P2Y_6$ after PGE_2-G activation.[85] During the investigation of *Pasteurella multocida* toxin (PMT), which is able to activate Gαq signalling and this way leading to the inhibition of osteoblast marker induction in a fibrodysplasia ossificans progressiva (FOP) model, FR was used to block the PMT effect, highlighting the importance of Gαq in this process.[100] Additionally, FR was used to study β-arrestin signalling in the absence of active G proteins. Here, Grundmann et al. combined genetic and pharmacological inhibition of G proteins to achieve a "zero functional G" state. Together with arrestin null cells, they investigated how the lack of G protein vs. the lack of arrestins does affect GPCR signalling. The authors concluded, that rather than arrestins, G proteins play a much more vital role as genuine drivers of GPCR-mediated signal transduction, which contradicts a long-established theorem.[90]

2.8 Conclusion of the review

The interest in the specific Gαq inhibitor natural products FR and YM has been constantly rising in the last decade, especially after the detailed pharmacological characterisation of FR in 2015. The extensive possibilities for the experimental use of this compound family range from pharmacological investigations of G proteins and GPCRs to the clinical use as a drug for Gαq induced diseases. Since G proteins show considerable sequence similarity, there have been efforts to generate specific inhibitors for other G protein families based on the FR/YM structure. While it was possible to transfer the FR binding site into Gαi[69], Gα16[125] and Gαs[126] to achieve inhibition of these proteins by FR, so far no FR inhibitors have been found able to block the wild type Gαi, Gα16 or Gαs proteins.[57]

Important current research goals include achieving a co-crystallisation of FR with the Gαqβγ protein complex, which might reveal new insights into the differences between FR and YM, e.g. regarding the very different kinetics.[56] Another open question is the site-specific application of FR or YM as a drug

to avoid general systemic responses like hypotension and bleeding. While there are some antibodies targeting GPCRs being evaluated as therapeutics,[127] different antibody-drug conjugates are investigated in cancer drug research in order to improve the therapeutic range of cytotoxic compounds.[128,129] The possibility to link the generally toxic FR molecule to a specific transport molecule for the effective targeting of disease state Gαq proteins may therefore be worth investigating. In addition, the use of aptamers might be another way for the targeted drug delivery of FR.[130] Taken together, the chromodepsins represent a small family of highly specialized natural products with an effective mechanism of action that affects many physiological processes. These complex metabolites are a good example that nature can provide bioactive molecules which can be used by humans as precise tools to tackle and to overcome pharmacological and therapeutical challenges.

3 Aim of the study

This thesis shall deal with the investigation of the biosynthesis of the cyclic depsipeptide FR900359 (FR). FR is a natural product with selective Gαq protein inhibitory activity, making it a valuable tool for the pharmacological research of intracellular G protein-coupled receptor signalling. FR is of bacterial origin and two different FR producers have been identified: The uncultivable, endosymbiont "*Candidatus* Burkholderia crenata", living in the leaf nodules of the plant *Ardisia crenata*, and *Chromobacterium vaccinii*, isolated from the soil of cranberry plants in the USA. Both bacterial genomes contain a highly similar biosynthetic gene cluster (BGC), *frs*, that was identified to be responsible for the production of FR.

FR biosynthesis is conducted by two nonribosomal peptide synthetase (NRPS) systems, FrsA and FrsD-G, assisted by the MbtH like chaperon protein FrsB, as well as the two tailoring enzymes FrsC and FrsH. In a biosynthetic model, the monomodular NRPS FrsA synthesizes the side chain of FR, while the FrsA thioesterase domain was hypothesised to transfer the side chain onto the macrocyclic intermediate FR-Core in an unusual intermolecular transesterification reaction.

This work focuses on the detailed investigation of FR side chain biosynthesis and its transfer and attachment onto FR-Core. Initial bioinformatic investigations of the two *frs* BGC will be performed, to predict the activity of the different NRPS modules. The heterologous expression of necessary constructs of the *C. vaccinii* BGC will be established in a suitable heterologous host. Then, the catalytic activity of the FrsA A, C, and TE domain will be tested in different assays to verify their predicted function. For the adenylation domain, an established γ-$^{18}O_4$-ATP-exchange assay will be used, whereas, for the starter condensation domain and the transesterifying thioesterase domain, new assays need to be developed or adapted from literature. To probe the transesterification of the side chain onto FR-Core, the cyclic intermediate needs to be isolated first in preparative amounts, e.g., from a *C. vaccinii* Δ*frsA* deletion mutant.

The substrate specificity of the domains involved in the side chain assembly will also be analysed to test the generation of new FR analogues with altered side chains. The bioactivity of these new derivatives will be tested in comparison to FR, to gain insight into the structure-activity relationships and possible evolutionary aspects.

Finally, the structure of FrsA and/or the $FrsA_{TE}$ domain shall be elucidated. This will provide an insight into the active sites and the binding of substrates and help to improve understanding of this unusual domain also to guide further biosynthetic engineering studies. To achieve this goal, X-ray crystallography and cryo-EM techniques will be evaluated.

4 Results and Discussion

4.1 Bioinformatic investigation of the *frs* BGCs

Before starting *in vitro* experiments we first performed a detailed bioinformatic analysis of both *frs* biosynthetic gene clusters (BGCs). The *frs* BGS of the new bacterial producer *C. vaccinii* (*cv_frs*) was sequenced and analysed by Dr. René Richarz. The detailed comparison of the *C. vaccinii* genes with the *frs* genes of "*Ca*. B. crenata" is listed in Table 4.1. The size of the BGCs and GC content had only slight differences, 58.5% GC on 35.8 kb for "*Ca.* B. crenata" and 66.9% GC on 35.9 kb for *C. vaccinii*.[33]

Table 4.1: Overview and comparison of genes and encoded proteins in the *frs* BGCs. Identities were calculated by using the EMBOSS needle alignment tool (EMBL-EBI).[33,131]

gene	*C. vaccinii* (nt)	"*Ca.* B crenata" (nt)	Identity nt (%)	*C. vaccinii* (aa)	"*Ca.* B crenata" (aa)	Identity aa (%)
frsA	3819	3768	70	1272	1255	71
frsB	219	219	72	72	72	75
frsC	987	987	68	328	328	72
frsD	3081	3078	70	1026	1025	70
frsE	9051	9048	70	3016	3015	71
frsF	7557	7560	73	2518	2519	75
frsG	9408	9411	72	3135	3136	73
frsH	1596	1599	77	531	532	85

The *frs* BGC encodes for two nonribosomal peptide synthetase (NRPS) systems, FrsA and FrsD-G. The biosynthetic principle of an NRPS is described in section 2.3 and depicted in Figure 4.1

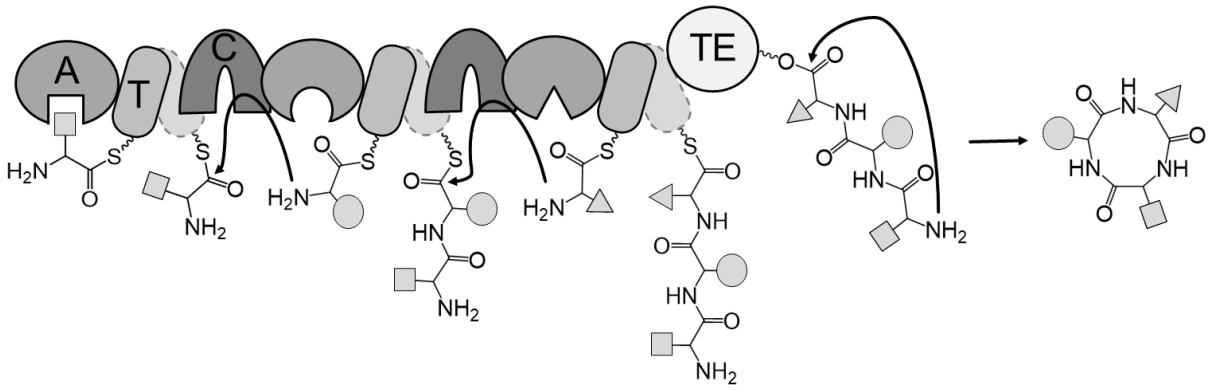


Figure 4.1: A hypothetical NRPS assembling a cyclic tripeptide, adapted from Stanisic *et al.*[39] The role of the T domain as flexible carrier domain is emphasized by showing it in two positions. . A = adenylation domain, C = condensation domain, T = thiolation domain, TE = thioesterase domain.

This work focuses on the first NRPS module FrsA, a monomodular NRPS that was hypothesised to contain an unusual TE domain (see Figure 2.5). The sequence of *frsA*$_{CAT}$ is almost identical (94.3%) to *frsD*, the first module of the heptamodular NRPS FrsD-G in both *frs* BGCs (Figure 4.2). The *cv_frs* nucleotide sequences encoding the A domains are 99.8% identical, differing only in 3 base pairs, however, still leading to a 100% identical amino acid sequence (see Figure 9.1). The C domains are 92.1% identical (see Figure 9.2), even though they are supposed to attach different acyl chains, propionyl for FrsA$_C$ and acetyl for FrsD$_C$. The overall comparison of FrsA to FrsD shows a 2640 bp sequence stretch that is 99.9% identical, encoding the complete A domain and large parts of the C and T domains (see Figure 4.2). Additionally, the protein sequence of the A domain of FrsA is 100% identical to the A domain of the module (7) of FrsG.[132] These results lead to the hypothesis, that duplication events occurred during the evolution of the *frs* BGC. We decided to investigate the C and the TE domain of FrsA more closely to get more evidence for this hypothesis.

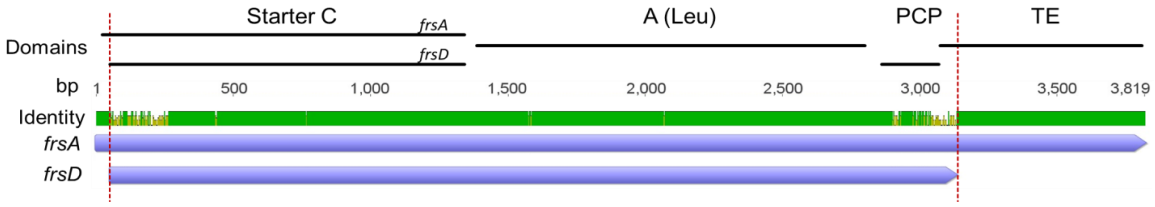


Figure 4.2: Nucleotide alignment of *frsA* and *frsD*. All predicted domains, identity and coverage (red lines) are indicated.[33]

4.1.1 The C$_{starter}$ domains of FrsA and FrsD

All C domains of the "*Ca.* B. crenata" *frs* BGC were analysed in a phylogenetic tree with other C domains.[32] C domains exist in different functional subtypes, the $^{L}C_{L}$, $^{D}C_{L}$, Cyc, E, dual E/C and C$_{starter}$ domains. The $^{L}C_{L}$ and $^{D}C_{L}$ domains catalyse the formation of a peptide bond between an L-amino acid to an L- or D-amino acid, respectively. Hererocyclization (Cys) domains catalyze the peptide bond formation followed by cyclisation of serine, cysteine, or threonine residues. Epimerization (E) domains invert the chirality of the last amino acid in a growing peptide and dual E/C domains induce both epimerization and condensation. A starter C (C$_{starter}$) can be present in the first module of an NRPS and acylates the first amino acid with a β-hydroxy-carboxylic acid, typically a β-hydroxy fatty acid.[44] The C domains of FrsA and FrsD cluster with these C$_{starter}$ domains.[32] This fits their proposed propionylation (FrsA) and acetylation (FrsD) function in the NRPS cluster. A coenzyme A activated acyl residue could be used as a substrate like demonstrated for the C$_{starter}$ domain in surfactin biosynthesis.[133] We aimed to investigate the phylogenetic origin of the FrsA and FrsD C$_{starter}$ domains from *bc_frs* and *cv_frs*. Therefore, we performed a BLAST search to collect the sequences of the most similar C$_{starter}$ domains. Removing all redundant sequences, yielded a set of 149 C$_{starter}$ sequences (see section 6.13.3 and Table 6.23). A phylogenetic analysis was performed, shown in Figure 4.3.

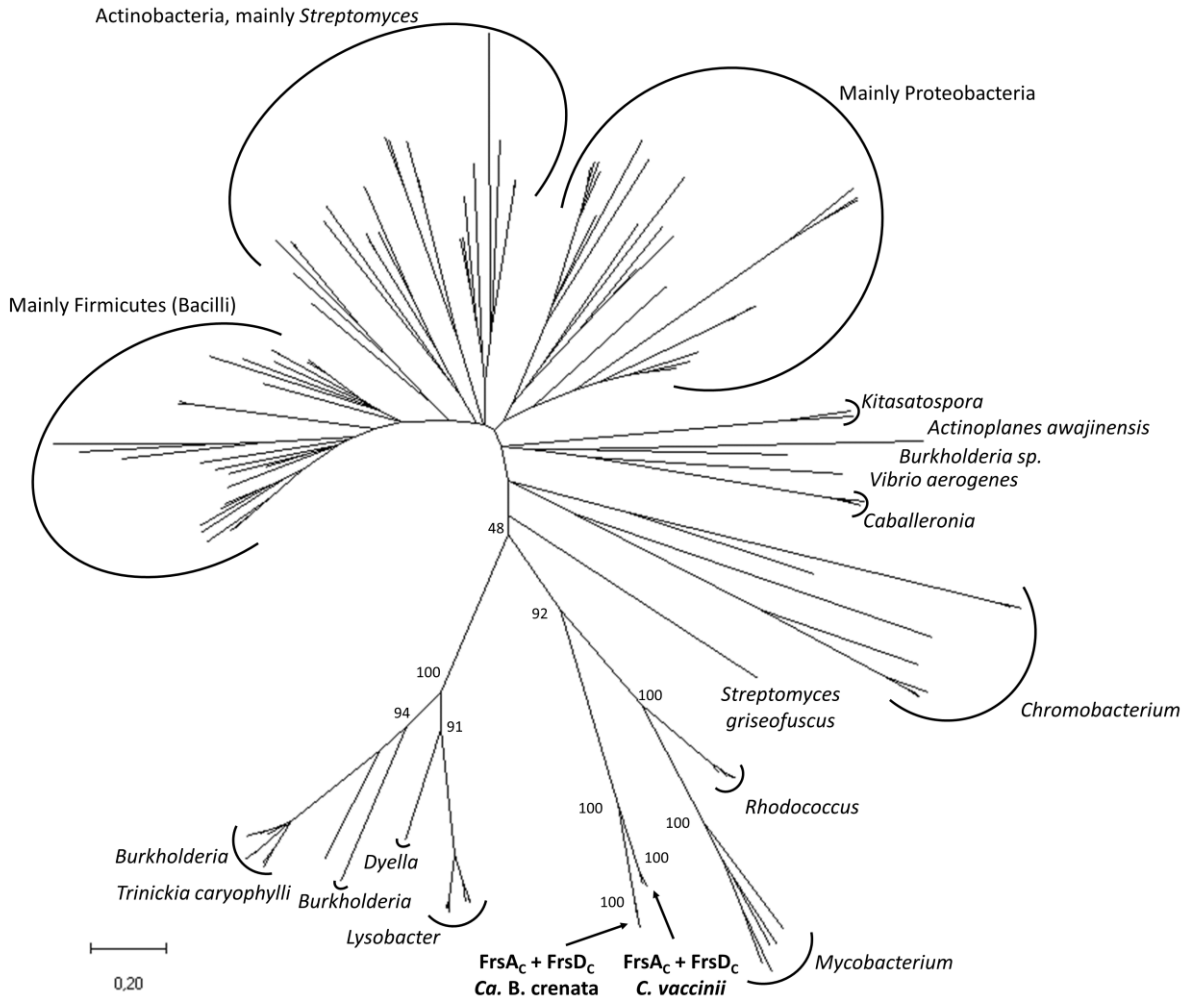


Figure 4.3: Phylogenetic tree of starter condensation domains. The $C_{starter}$ domains of the FR biosynthesis are indicated with arrows.[33]

The domains mostly clade according to their taxonomic origin. Niehs *et al*. recently investigated $C_{starter}$ domains with known substrates and found them to clade by taxonomy instead of substrate specificity. The authors hypothesised that the substrate specificity might be determined by a specificity-conferring code and cannot be distinguished in phylogenetic analysis.[134] Our results support this hypothesis, but interestingly the $C_{starter}$ domains from both *frs* clusters do not clade with others from *Burkholderia* or *Chromobacteria* taxa. Instead, the four domains form a deeply rooted clade, supporting our hypothesis of their close relationship and highlighting differences to other domains in the database.[33]

4.1.2 The TE domains of FrsA and FrsG

The general mechanism of action of TE domains is the same in fatty acid synthases, polyketide synthases and non-ribosomal peptide synthetases. The substrate, bound to the T domain, is transferred onto the hydroxy group of the active site serine of the TE domain and then released via nucleophilic attack. The release step occurs induced by an intramolecular O-, N-, or C-nucleophile, effecting macrolactonization, macrolactamization or Claisen-like condensation, respectively, or by the attack of an exogenous nucleophile like water leading to hydrolysis or transesterification, see Figure 4.4.[40]

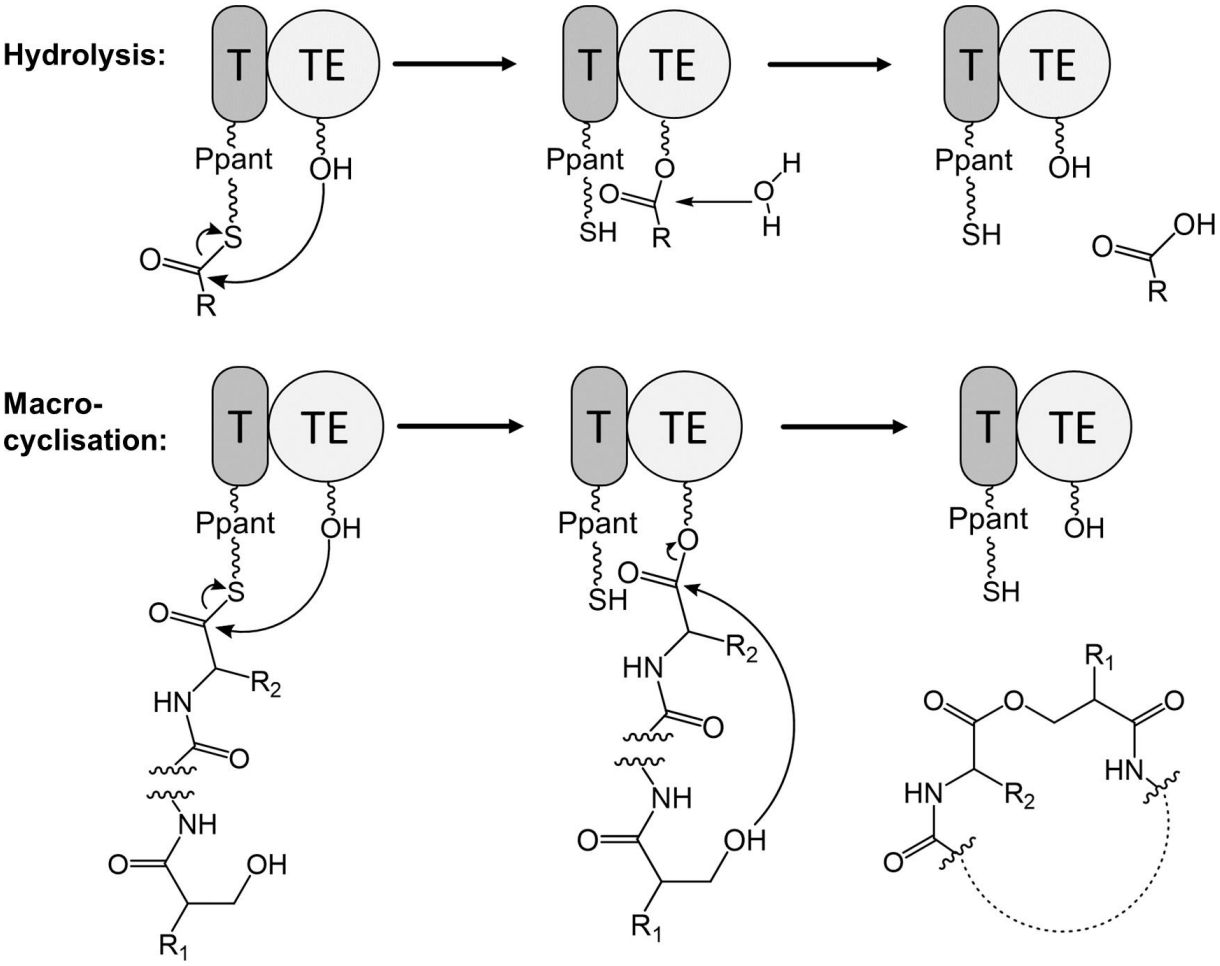


Figure 4.4: TEs catalyze substrate offloading from T domains of non-ribosomal peptide synthetases through hydrolysis or macrocyclization. The macrocyclization is depicted for the example of a serine residue, but it can also occur with the side chains of other amino acids. R_1 = peptide chain, R_2 = any amino acid side chain.

TE domains tend to share low sequence homologies across different taxa and generally exhibit broad substrate promiscuity.[40] In 2014, a phylogenetic investigation concluded, that TE domains, in general, do not cluster based on substrate specificity or function.[135] In 2018, Klapper *et al*. published a maximum likelihood tree of 27 TE-like domains from bacterial NRPS.[136] This tree also showed no as deeply rooted clades as the C domain tree (Figure 4.3), indicating a higher evolutionary distance between the TE domains. Nevertheless, some clades of TE domains with the same release mechanism were formed. We expanded the dataset from Klapper *et al*. by our TE domains and further on investigated TEs with different release mechanisms to calculate a new maximum likelihood tree with MEGA 6 (see Figure 4.5, section 6.13.3and Table 6.24).

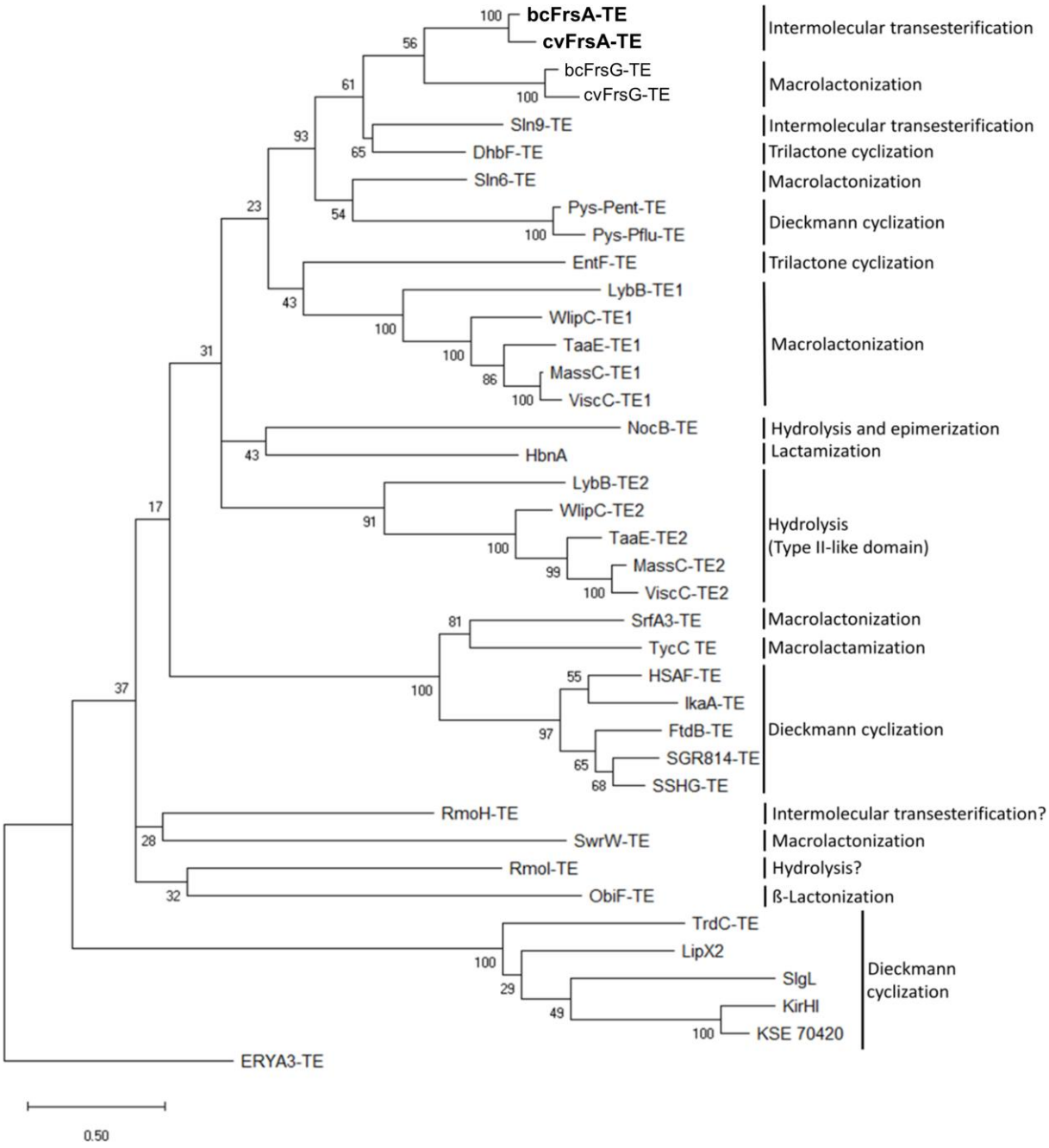


Figure 4.5: Phylogenetic tree of NRPS TE domains. For experimental details, see 6.13.3.

This analysis showed that the TE domains of FrsA are most closely related to the TE of FrsG from the same gene cluster, which is supposed to catalyse the classic macrolactonization. The transesterifying TE of Sln9[43] is located nearby, but not as close as other genes from the same organism. This is in line with the findings of Klepper *et al*. who could not find a motif for the TE domains that yield in Dieckmann cyclization products.[136] Specific functions of enzymes cannot always be found in a genomic motif.[135] In 2012 a study showed some TE domains to cluster, that produce the same ring size in the resulting peptide.[137] This could not be observed in our tree, as a lot of different offloading mechanisms, not only cyclisations, were present. The alignment of $FrsA_{TE}$ and $FrsG_{TE}$ from *cv_frs* showed a sequence identity of 41.7% and a similarity of 62.3% (see Figure 9.3), which is poor, regarding that they are the closest

phylogenetic neighbours. Both have the active site triad of serine, aspartic acid and histidine, which is common for most TE domains.[40] There are some NRPS TEs known to have a catalytic cysteine residue instead of serine like in the polymyxin synthetase but it is quite uncommon for type I TEs.[138] In Figure 4.6 the I-TASSER-calculated structure of $FrsA_{TE}$ (A) and an alignment of the structural models of $FrsA_{TE}$ and $FRsG_{TE}$ (B) is shown. The structures of the TE domains show a repeating β/α/β motif that forms a six-stranded parallel β-sheet with a left-handed helical twist and two α-helices forming a lid over the active site with the binding serine. This general structure is similar to the structures of reported type 1 TEs.[40] The alignment of $FrsA_{TE}$ and $FrsD_{TE}$ shows only some small variances, and the location of the amino acids in the active site is nearly identical.

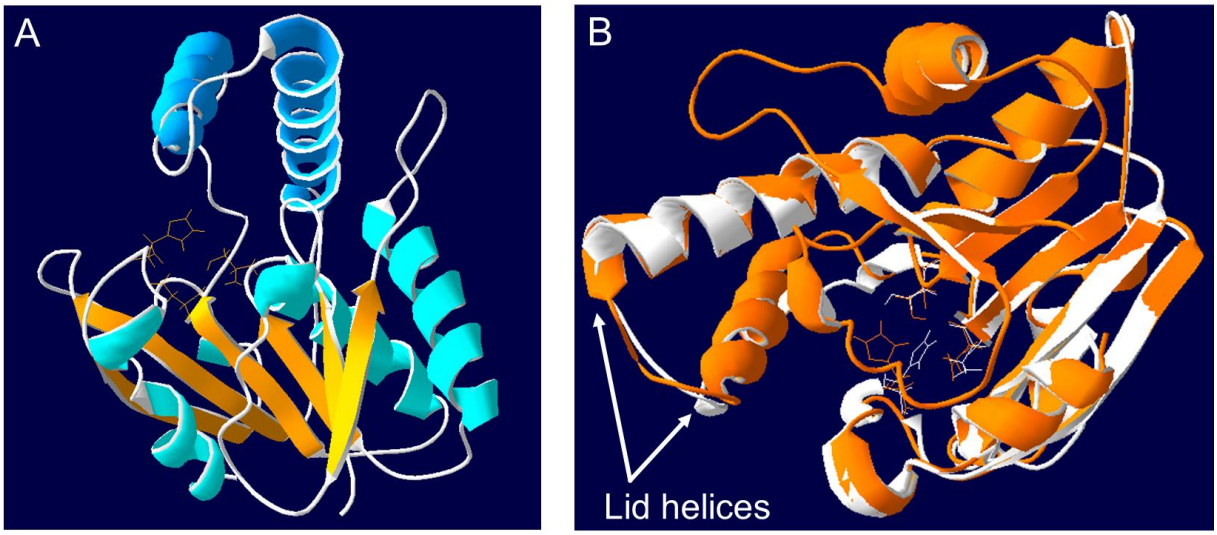


Figure 4.6: Structural models of the TE domain of FrsA and FrsG. A The I-TASSER model of $FrsA_{TE}$, β-sheets are displayed in orange, α-helices are displayed in cyan and the two helices of the lid in blue. The amino acids (Ser, Asp and His) of the active site are displayed in orange. **B** Alignment of the structural models of $FrsA_{TE}$ (orange) and $FrsG_{TE}$ (white) from *C. vaccinii*.

The highest similarity for existing structures of TE domains to the structure of $FrsA_{TE}$ calculated by I-TASSER is $NocB_{TE}$ (PDB: 6ojdA). The latter is a bifunctional domain that catalyses not only hydrolysis but also epimerisation in the biosynthesis of nocardicin.[139] Thus, it is a very special TE as well, but it does not clade very near to the *frs* TEs in the phylogenetic analyses. To verify the similarities and differences of these three domains in detail, the crystal structures of $FrsA_{TE}$ and $FrsG_{TE}$ would be needed.

In summary, predicting the function of TE domains from the primary structure TE is not as facile as for other domains. They vary in function and structure and do not have high sequence similarities, even when catalysing similar reactions. To get more insights into these domains in the *frs* BGC, we wanted to investigate them *in vitro* and therefore progressed with their cloning and heterologous expression.

4.2 Cloning and expression of *frs* genes

After the discovery of *C. vaccinii* as a cultivable producer of FR, opposed to the not cultivable endosymbiont "*Ca.* B. crenata", we decided to work with the *cv_frs* for all biosynthetic *in vitro* experiments of the *frs* BGC. For the detailed investigation of the biosynthetic modules FrsA and FrsD,

both modules were cloned and heterologously expressed in *E. coli* to perform *in vitro* reconstitution assays. Additionally, the modules were expressed in truncated single, di or three domain constructs. FrsB was coexpressed to assist in different assays and FrsH was expressed separately in cases where hydroxylation was needed. To compare the two TE domains, the TE of FrsG was also expressed in a single domain construct. All relevant constructs for this work are summarised in Table 4.2.

Table 4.2: Heterologously expressed proteins used in this work.

Protein	Calculated molecular weight	Coexpressed with:
FrsA	142.35 kDa	FrsB
$FrsA_{CAT}$	116.46 kDa	FrsB
$FrsA_{AT}$	67.76 kDa	FrsB
$FrsA_{A}$	59.17 kDa	FrsB
$FrsA_{TE}$	31.49 kDa	-
FrsB	8.19 kDa	-
FrsD	116.16 kDa	FrsB
$FrsG_{TE}$	30.98 kDa	-
FrsH	63.75 kDa	-

To generate new expression constructs, the gene sequences of the selected domains were amplified via PCR with specific primers (see Table 6.14) from the gDNA of *C. vaccinii.* By using selected restriction sites, the PCR products were ligated into the bacterial expression plasmid pET28a. These new plasmids, listed in Table 6.18, were transformed into *E. coli* α-Silver-Select and the correct sequence verified by Sanger sequencing. Afterwards, the plasmids were transformed into an *E. coli* expression strain and the expression was induced with IPTG and analysed. The proteins were all expressed with an N-terminal hexahistidine tag for Ni-NTA affinity purification. The first test expressions were used to see if the protein of the expected size is expressed in soluble form and in detectable amounts so that it can be purified by the chosen method. All fractions from the affinity chromatography were analysed via SDS-PAGE, for details, see section 6.6. In Figure 4.7, the SDS-PAGE of the first attempted expression of $FrsA_{AT}$, coexpressed with FrsB in BL21, is shown as an example for the protein purification. The protein band in the elution fractions between 60 and 70 kDa fits the calculated mass of $FrsA_{AT}$ with 67.76 kDa and an additional band at approximately 8 kDa matches the coexpressed FrsB which is co-eluted with the A domain even though it has no His6-tag. The band of $FrsA_{AT}$ is not visible in the non-induced sample, which was taken from the expression culture before adding IPTG and lysed separately. This reassures that the expression is dependent on the induction of the T7-polymerase, as expected in a DE3-expression strain.[140] The N-terminal His6-tagged $FrsA_{AT}$ yielded the highest amount of tagged protein in the first elution fraction, whereas most unspecific binding proteins are eluted in the two wash steps.

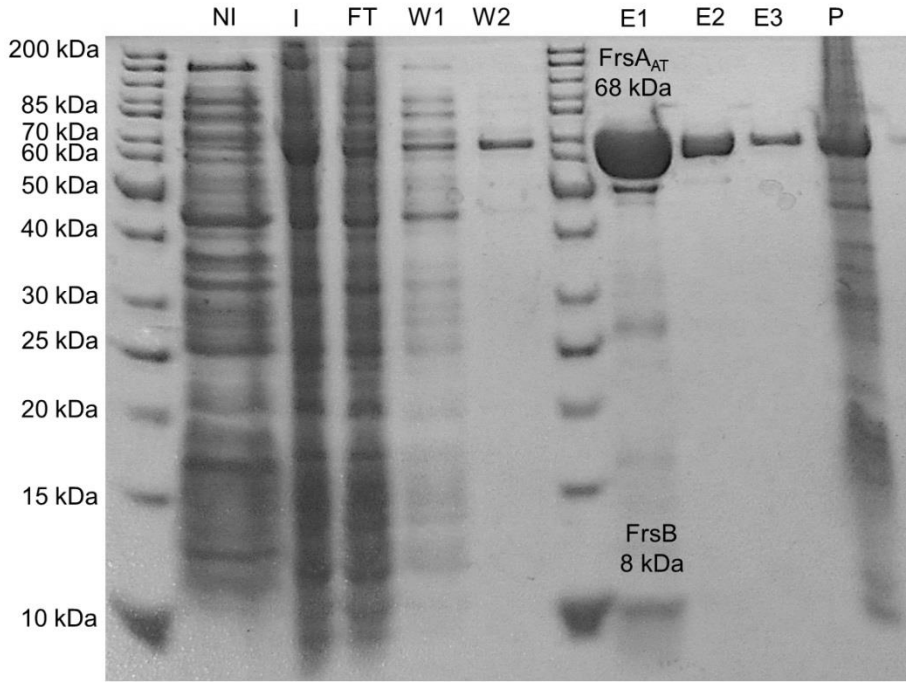


Figure 4.7: SDS-PAGE of FrsA$_{AT}$ test expression in BL21 coexpressed with FrsB. NI= non induced sample, I= induced sample, FT= flow through, W1= washing step 1, W2= washing step 2, E1= elution fraction 1, E2= elution fraction 2, E3= elution fraction 3, P= pellet.

If the protein was expressed in sufficient amount and purity, the combined elution fractions were used for further investigation. Otherwise, the expression conditions and the purification protocol were optimized, or a new expression construct needed to be designed.

In Figure 4.8, SDS-PAGE gels of all proteins used for bioassays are pictured with their respective calculated molecular masses. All proteins including an A domain were coexpressed with FrsB to ensure the solubility and activity of the protein. All constructs harbouring a T domain were conducted in the expression strain BAP1, which carries the *sfp* phosphopantetheinyl transferase gene to ensure *in vivo* phosphopantetheinylation of the T domain.[141] Interestingly, nearly all proteins migrate further than their calculated molecular weight. It has been discussed in different publications, that some proteins do not migrate at the estimated band hight due to their hydrophobicity.[142,143] Shirai *et al*. suggested, that SDS preferentially binds to the hydrophobic instead of the negatively charged regions of proteins, which may cause inconsistencies regarding theoretical mobility when comparing chemically diverse proteins.[142] As our proteins were proven to be expressed from the particular plasmid, and to be bioactive (see below), we did not investigate this unusual mobility any further.

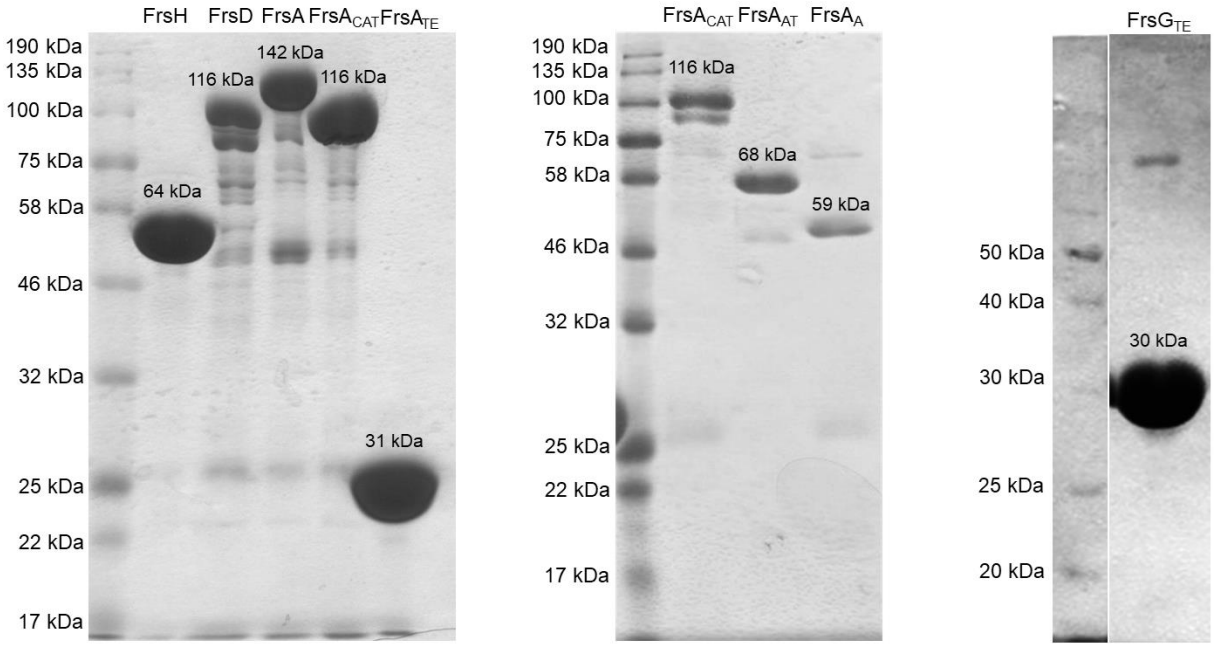


Figure 4.8: SDS-PAGEs of all purified proteins used in this work for bioassays. The names of the proteins are located on top of the lanes and the calculated molecular masses of the proteins are shown above the bands.

4.3 Activity tests of the FrsA A and C domains

The expressed and purified proteins were investigated for their *in vitro* activity. The chosen assays were designed to prove the bioinformatically proposed function of the single domains. Before the activity of the $C_{starter}$ domain can be assessed, the A domain needs to be proven to be active, as the C domain attaches the acyl residue onto the T domain bound amino acid.

4.3.1 A domain assays

The γ-$^{18}O_4$-ATP-exchange assay, chosen to test the activity of the A domain of FrsA, was published by Phelan *et al*. in 2009.[144] It utilises the equilibrium reaction of ATP and the specific amino acid to the amino acid adenylate and pyrophosphate (PPi) by the A domain. Under assay conditions, only labelled "heavy" γ-$^{18}O_4$-ATP is present, which is used by the A domain as substrate to activate the amino acid. This leads to the formation of labelled PPi and the adenylate. An excess of unlabelled PPi is added to the reaction, which reacts with AMP back to unlabelled γ-$^{16}O_4$-ATP. The ratio of unlabelled to labelled ATP is measured via MALDI-TOF-MS and is used to calculate the substrate conversion, described in section 6.11.1. The reaction scheme is shown in Figure 4.9.

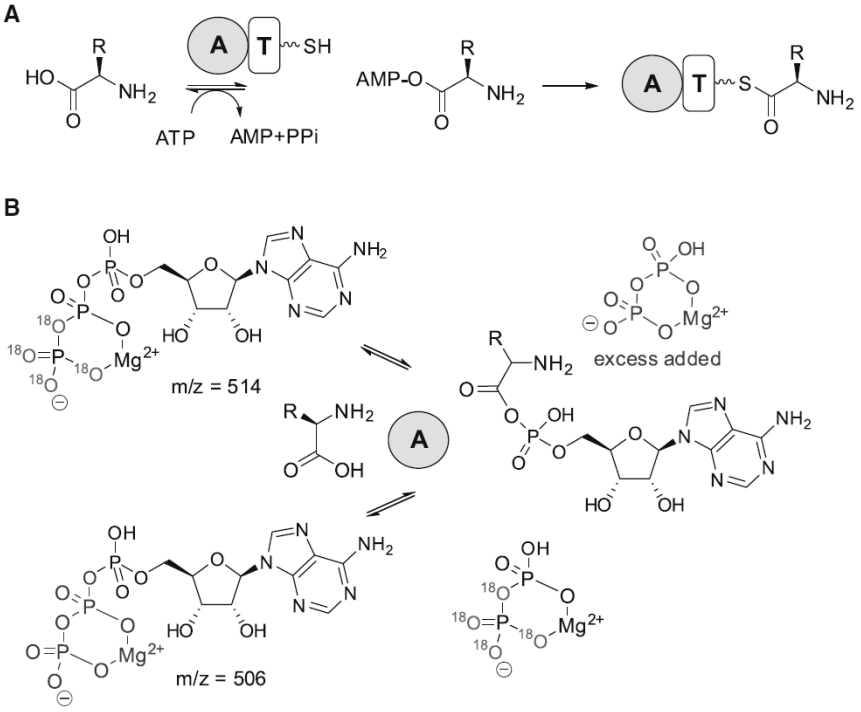


Figure 4.9:Principel of the γ-$^{18}O_4$-ATP-exchange assay. A. A domains in NRPS systems adenylate amino acids for subsequent thiolation reactions on T domains. **B.** The exchange reaction, performed in the absence of thiolation activity, measures equilibrium exchange of γ-$^{18}O_4$-ATP with $^{16}O_4$-pyrophosphate, taken from Phelan *et al*.[144]

The first A domain assays were performed by the intern student Tobias Götzen, supervised and with expression strains provided by Daniel Wirtz and me (see section 6.11.1). We aimed to investigate, if the size of the construct and the surrounding C domain and T domain, influence the A domains activity. Therefore, only the expected substrate L-leucine, the negative control without any amino acid and one further amino acid, L-isoleucine, were chosen for this reaction. The results are shown in Figure 4.10.

Clearly, the preferred activation of L-leucine was proven for all constructs, with a percental exchange of over 90% for each construct. Exchange rates in the negative controls were between 3.31% and 6.38% which is a common fluctuation for this assay and proved the assay to be working correctly. L-isoleucine was not activated in a significant amount by all tested constructs, even though it is structurally similar to L-leucine. While A domains normally act as gatekeepers, only activating one amino acid, standalone A domains are known to show some specificity also for structurally similar substrates.[145] The comparison of the three constructs showed no significant differences in the activity of the $FrsA_A$, hence the surrounding domains do not have a strong influence on substrate specificity in this experiment. In the next A domain assay, a bigger panel of structurally different amino acids were tested on the whole FrsA construct, which should be closest to the natural conditions of this reaction.

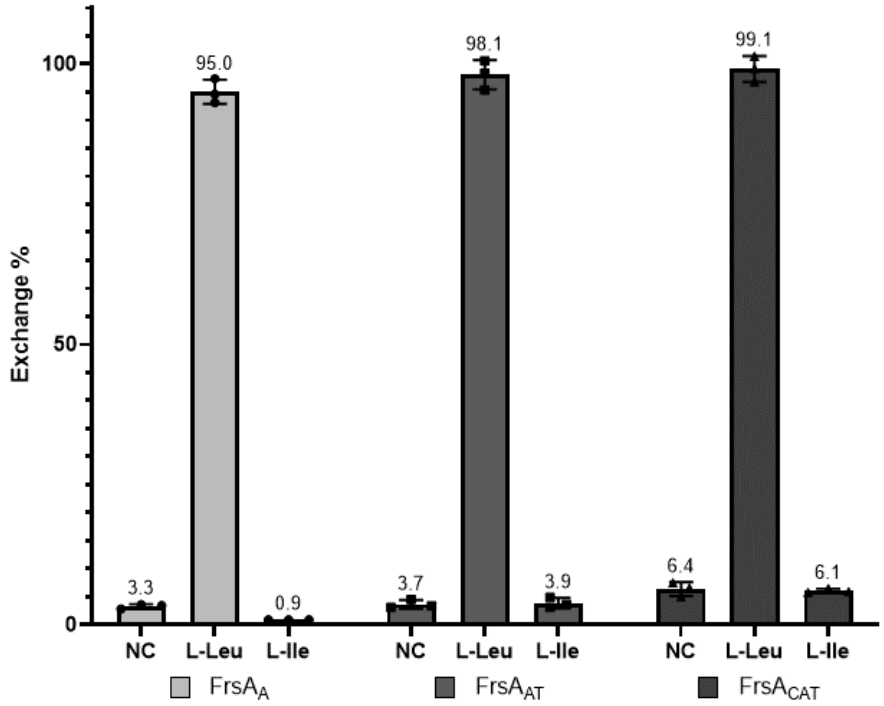


Figure 4.10: γ $^{-18}O_4$-ATP exchange assay results for the A domain of FrsA in three different constructs, $FrsA_A$, $FrsA_{AT}$ and $FrsA_{CAT}$, all coexpressed with FrsB. NC = negative control without any amino acid. Data are presented as mean values ± SD. All experiments were performed in technical triplicate.

The results of the A domain assay with the whole FrsA construct are pictured in Figure 4.11. The negative control showed an exchange of 1.08%, meaning no activation occurred without an amino acid, proving this assay can be used to compare the amino acid substrates. The L-amino acids alanine, glutamic acid, glycine, lysine, phenylalanine, serine, and threonine showed no ATP exchange. L-isoleucine and L-valine are activated with an exchange of under 10%, which is still in the range of unspecific activation. Interestingly, L-hydroxyleucine showed an exchange of 24% which is approximately one quarter of the exchange of L-leucine. As the module FrsA is supposed to incorporate *N*-propionylhydroxyleucine (*N*-Pp-Hle), direct activation of hydroxyleucine would be a possible way for FR biosynthesis as A domains can activate proteinogenic but also unusual amino acids.[146] The bioinformatic analyses, however, suggested activation of leucine and subsequent hydroxylation by FrsH (see Figure 2.5).[32] In conclusion, the results of this assay support this theory, which will be investigated further in section 4.3.2.

Besides L-leucine, D-leucine is also activated by FrsA, which is unusual for most A domains.[147] As D-amino acids rarely occur *in vivo*, it can be hypothesised that there was no evolutionary pressure to avoid activation of an amino acid that is not present in the natural environment, which might explain, why the activation of D-leucine is possible at all. It might be interesting to see if D-leucine is incorporated in the structure of FR in feeding experiments. Here, also the downstream enzymes would need to accept D-leucine and D-leucine containing precursor peptides as substrate.

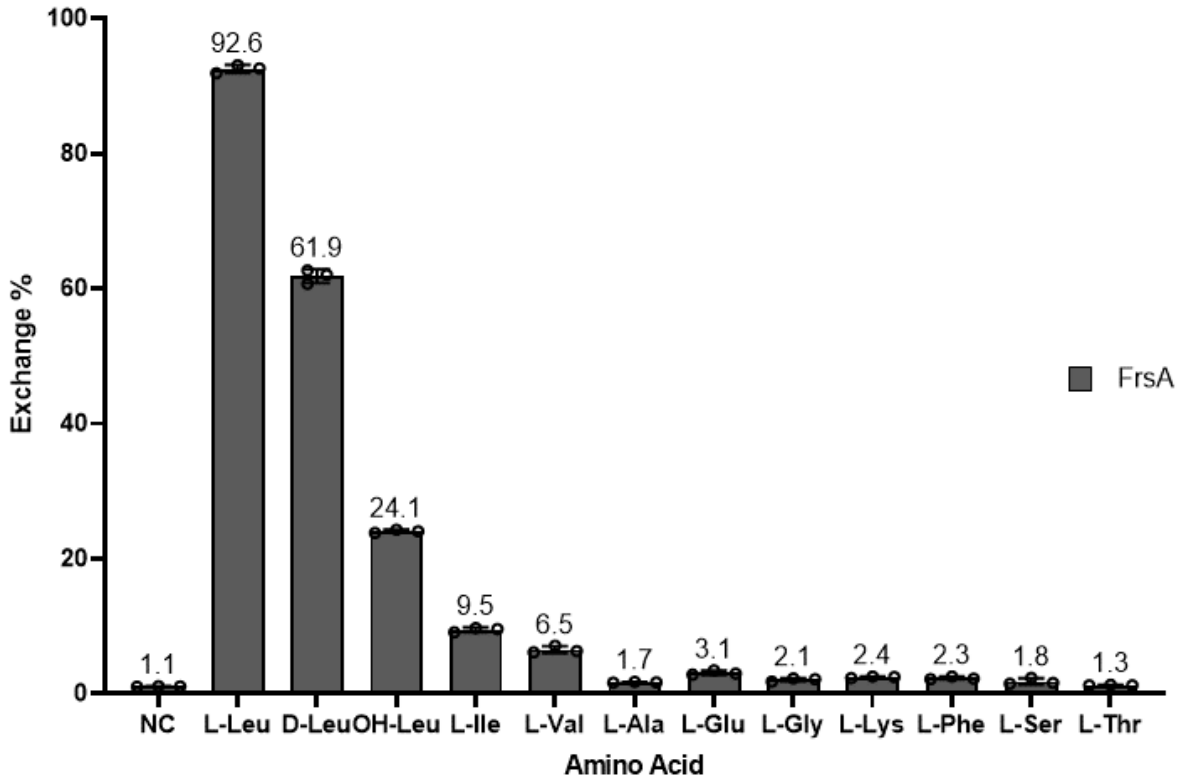


Figure 4.11: γ $^{-18}O_4$-ATP exchange assay results for the A domain of FrsA in the whole module construct coexpressed with FrsB. NC = negative control without any amino acid. Data are presented as mean values ± SD. All experiments were performed in technical triplicate.

All results of the A domain assays support the proposed function of the FrsA A domain. This should be also the same as for the A domains in FrsD and FrsG (7), as they are nearly or completely identical in sequence to $FrsA_A$ (see 4.1).

4.3.2 C domain assay

In 2010, the workgroup of Prof. Marahiel developed an *in vitro* assay to test the activity of a $C_{starter}$ domain in the biosynthesis of surfactin.[133] They first proved that 3-hydroxy myristic acid was activated by a fatty acyl CoA ligase, forming 3-hydroxymyristoyl-CoA thioester. Using the heterologously expressed first CAT module of SrfA, they characterised the transfer of the activated fatty acid onto the amino acid amino group. Based on this method, we developed an assay for the FrsA $C_{starter}$ domain. First, it had to be ensured, that the T domain is loaded with 4'-phosphopantetheine to bind the activated amino acid. This reaction is catalysed by 4'-phosphopantetheinyl transferases (PPTases).[148] We used the modified expression strain *E. coli* BAP1, which carries the Sfp PPTase from *Bacillus subtilis*[141] for *in vivo* phosphopantetheinylation to directly isolate the *holo*-protein. The overexpressed and isolated multidomain protein was then used for the bioassay, starting with the loading of the T domain with the amino acid. All substrates were added at the same time to the reaction (see section 6.11.2), but the following steps of the reaction are expected: As proven in section 4.3.1, L-leucine is the preferred

substrate for the A domain, so the reaction is started by ATP activation of L-leucine which is then tethered to the T domain. Then, the activated acyl residue propionyl-CoA is recognized by the donor site of the C domain catalysing the attack of the amino group of T domain bound L-leucine. After the reaction, the modified amino acid thioester was cleaved by alkaline hydrolysis and analysed by LC-MS. In Figure 4.12. the results of the first C domain assay are pictured. In the negative control, with heat-inactivated protein, no signal for the extracted mass of *N*-Pp-Leu is detectable. In the respective extracted ion chromatogram of the assay with active protein, there is a small peak with the fitting exact mass, but with extremely low intensity. As there was no synthetic standard available, we were not able to verify the retention time of *N*-Pp-Leu, leading to the conclusion that these results were not sufficient to clarify the activity of the $C_{starter}$ domain.

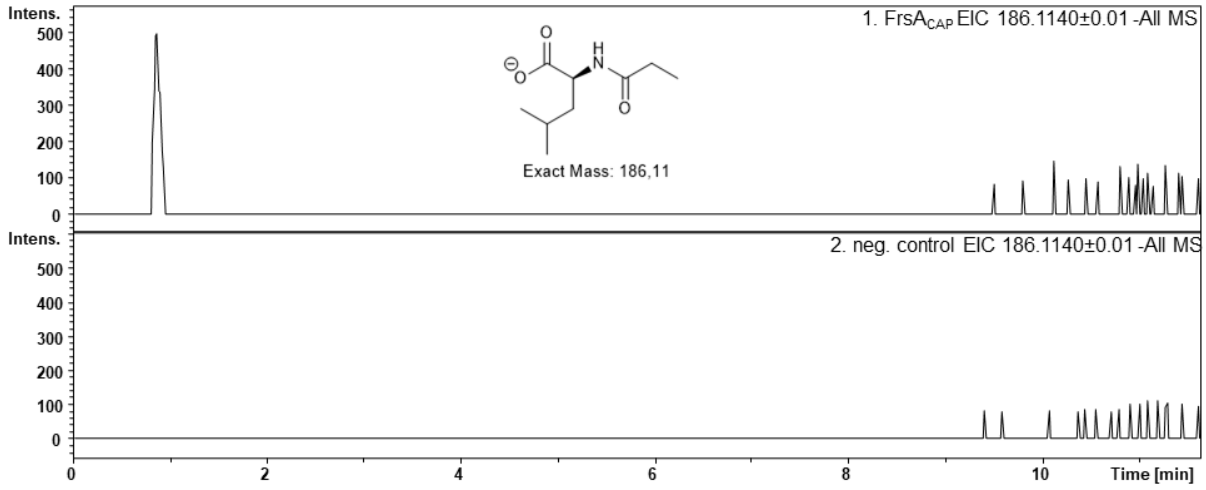


Figure 4.12: C domain assay with FrsA$_{CAT}$ coexpressed with FrsB. Extracted ion chromatograms (EIC) of *N*-Pp-Leu (*m/z* 186.1140) from HPLC-MS experiments. **1.** In vitro assay; **2.** Negative control with heat-inactivated protein.

In NRPS assembly lines, the succession of the single reaction steps is of utmost importance, as the substrate specificity of the involved domains can be very strict, meaning one part of the reaction can only take place when certain reactions were performed beforehand. During assembly of the FR side chain, there is another modification, which we hypothesised that it might occur before the attachment of the acyl residue: The hydroxylation at position 3 of L-leucine. A common feature of NRPS systems is the modification by tailoring enzymes *in trans* before the condensation domain catalyses the next step of the peptide synthesis.[149] In this case, the C domain acts as a gatekeeper to prohibit the conversion of wrong building blocks. The work of Daniel Wirtz from our workgroup focuses on the non-heme diiron monooxygenase FrsH which was supposed to hydroxylate leucine when bound to the T domain. We thus developed an adapted *in vitro* assay, including the activated FrsH. FrsH was heterologously expressed in medium supplemented with Fe(III) to ensure the formation of the diiron cluster in the active center of the enzyme. After purification, FrsH needed to be activated by transformation into the ferric state which was achieved by chemical reduction with sodium dithionite in the presence of

methylviologen as electron transmitter. Afterwards, FrsH was added to the C domain assay mixture (see 6.11.2).

This assay was expected to lead to the production of *N*-Pp-Hle (**21**), which was chemically synthesized, as authentic standard (see section 6.9). In this assay, the isomeric amino acids D-leucine and L-isoleucine were also tested as alternative substrates, the results are shown in Figure 4.13. For all substrates, the conversion to the side chain molecule **21**, or respective isomers with fitting exact mass and retention time was shown. As expected, the intensity of the product peak was highest for the natural substrate L-leucine. Still, the A and C domain and also FrsH show some substrate promiscuity *in vitro* through the formation of isomeric side chains. The sequence of these reactions remains enigmatic, therefore the experiment needs to be conducted in separate steps with each step analysed for the intermediates, which would be especially helpful for the more detailed investigation of the reaction mechanism and substrate specificity of FrsH.

The next aim of this study was to investigate the transfer of the side chain onto the anticipated substrate FR-Core, which should be discussed in the following sections.

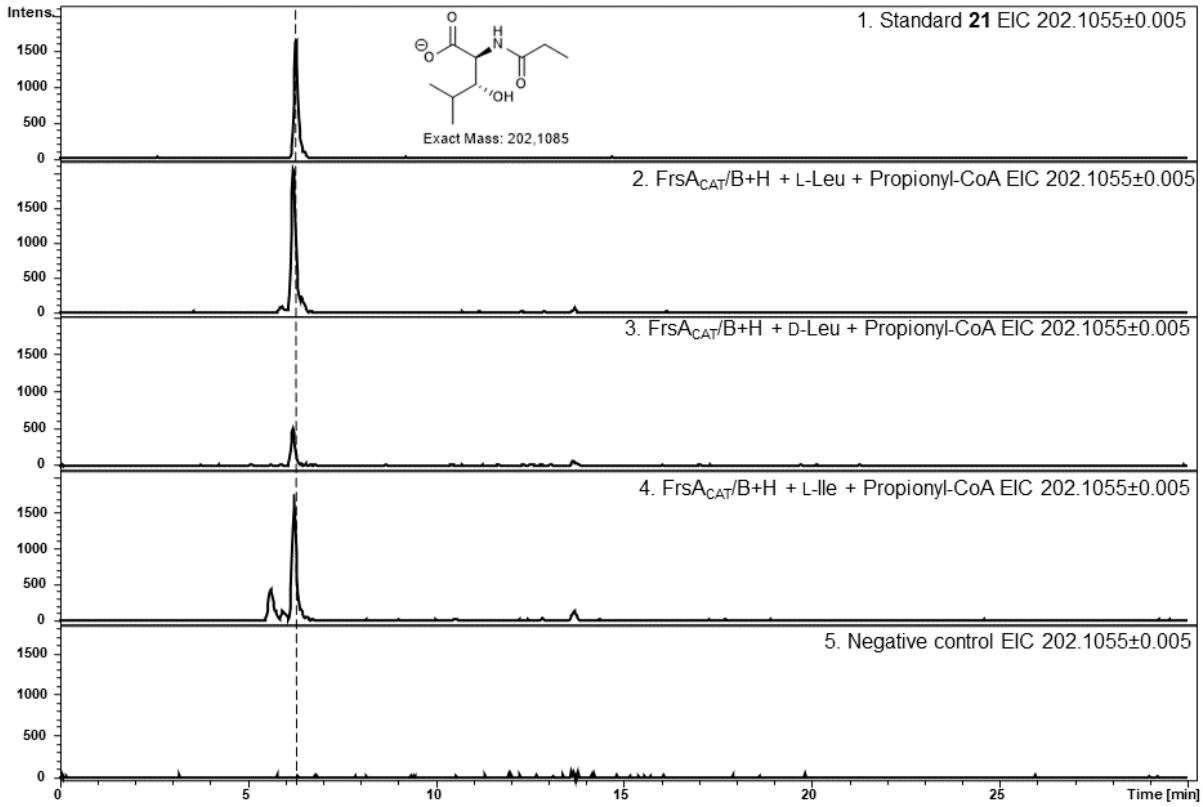


Figure 4.13: *In vitro* production of *N*-Pp-Hle (21). Extracted ion chromatograms of *N*-Pp-Hle (21) (*m/z* 202.108) from HPLC-MS experiments. **1.** Synthetic **21** (1 µg/ml) with depicted structure. Enzymatic assay with purified FrsA$_{CAT}$/FrsB, FrsH; **2.** incubated with L-Leu and propionyl-CoA, hydrolyzed with KOH; **3.** incubated with D-Leu and propionyl-CoA, hydrolyzed with KOH; **4.** incubated with L-Ile and propionyl-CoA, hydrolyzed with KOH. **5.** Negative control (NC) with heat-inactivated protein.

4.4 Transesterification assay

To investigate the function of the FrsA TE domain, we adapted the assay developed by Ray *et al*. in the Moore group, who investigated a transesterifying TE domain in salinamide biosynthesis. In this study, the TE domain of Sln9 was proven to be responsible for the installation of the "basket handle" (4-methylhexa-2,4-dienoyl)glycine moiety across a hexadepsipeptide core of salinamide A.[43] For the *in vitro* assay to prove the TE function, the authors explored the needed reaction substrates: The cyclic peptide substrate, Desmethylsalinamide E, was isolated from an Δ*sln4* MT mutant and substrate mimics for the T domain-bound "handle" were chemically synthesized. The recombinant hexahistadyl-tagged Sln9 TE domain was able to catalyse the transfer of the "handle" from the T domain mimic onto the free serin hydroxy group from the Desmethylsalinamide E.[43] We used this example to develop our own TE domain assay. Therefore, the substrates needed to be generated, which is described in the following section.

4.4.1 Synthesis and isolation of substrates for the TE domain assay

The *N*-Pp-Hle side chain, assembled by $FrsA_{CAT}$ is supposed to be transferred by $FrsA_{TE}$ onto the cyclic depsipeptide FR-Core. This side chain needed to be synthesized bound to a thioester, imitating the T-domain-bound state of the natural substrate. *N*-acetylcysteamine (SNAC) has been proven to successfully mimic the T domain-bound phosphopantetheine moiety.[150] The synthesis was developed and performed in cooperation with Dr. Jim Küppers from the group of Prof. Gütschow of the Institute of Pharmaceutical & Medicinal Chemistry, University of Bonn.

The other needed suspected substrate was the cyclic peptide biosynthetic intermediate FR-Core, which is produced only in traces by "*Ca*. B. crenata" and *C. vaccinii*. To isolate and characterise this compound, an Δ*frsA* mutant of *C. vaccinii* was generated by Dr. René Richarz of our working group.

4.4.1.1 Synthesis of *N*-Propionylhydroxyleucine-SNAC (22)

In cooperation with Dr. Jim Küppers, we developed a synthetic route towards the side chain precursor molecule incorporating the *N*-acetylcysteamine thioester.

Initially, the carboxylic acid functionality of propionic acid was activated via the "mixed anhydride method",[151] and coupled to the amino group of (2*S*,3*R*)-β-hydroxyleucine (**20**), yielding 762 mg (75%) of *N*-propionyl-3-hydroxyleucine (*N*-Pp-Hle, **21**), the free side chain of FR, as a white solid (see section 6.9.1). The final thioester derivative **22** was assembled through a carbodiimide-promoted coupling reaction with *N*-acetylcysteamine, which serves as an imitator of the phosphopantetheinyl moiety of the T domain. It was obtained as a clear oil with a yield of 82 mg (9%) (see section 6.9.2). The reaction scheme is shown in Figure 4.14.

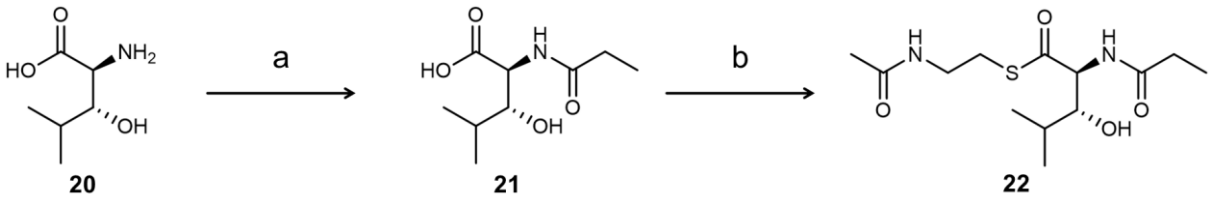


Figure 4.14: Synthesis of the *N*-Pp-Hle SNAC thioester 22. Reagents and conditions: (a) propanoic acid, $ClCO_2i$-Bu, NMM, THF, 1M NaOH, -10 °C (0.5 h) to rt (24 h); (b) N-acetylcysteamine, DCC, HOBt × H_2O, MeCN, rt, 24 h, N_2.

The substances were isolated via column chromatography and analysed using ^{1}H and ^{13}C NMR data to confirm the structure and purity (see Figure 9.8 to Figure 9.11). The synthetic intermediate *N*-Pp-Hle was also used as the standard for the free side chain in the C domain assay in section 4.3.2.

4.4.1.2 Generation of *C. vaccinii* Δ*frsA/vioA* double knock-out

We had hypothesised that FrsA produces the side chain in parallel to the biosynthesis of FR-Core by FrsD-G and subsequently couples it to FR-Core to yield FR. The detection of a molecule with *m/z* 817 during MS/MS networking with *Ardisia* extracts and its subsequent detailed MS/MS analysis indicated this molecule to be a biosynthetic FR intermediate without side chain, present only in traces.[32] The same mass signal was also found in traces in extracts of *C. vaccinii*. We named this hypothetical intermediate FR-Core. According to our biosynthetic hypothesis, the deletion of *frsA* would stop the assembly of the side chain, but not of the core molecule and should thereby lead to an accumulation of FR-Core. This would also further prove that the side chain is assembled by FrsA. Additionally, the production of the FR-Core in larger scale would enable preparative isolation and complete structure elucidation by NMR, as FR-Core was only analysed by LC-MS/MS so far.[36]

For the construction of *C. vaccinii* deletion mutants, a strategy employed also for the investigation of FK228 (Romidepsin) biosynthesis[152] was adapted by our lab. Here, the respective knock-out vectors contain the FRT cassette from pPS858, which has been developed for the construction of *Pseudomonas* deletion mutants (see section 6.5.8 and 6.5.9).[153] In Figure 4.15 a scheme of the deletion strategy is shown.

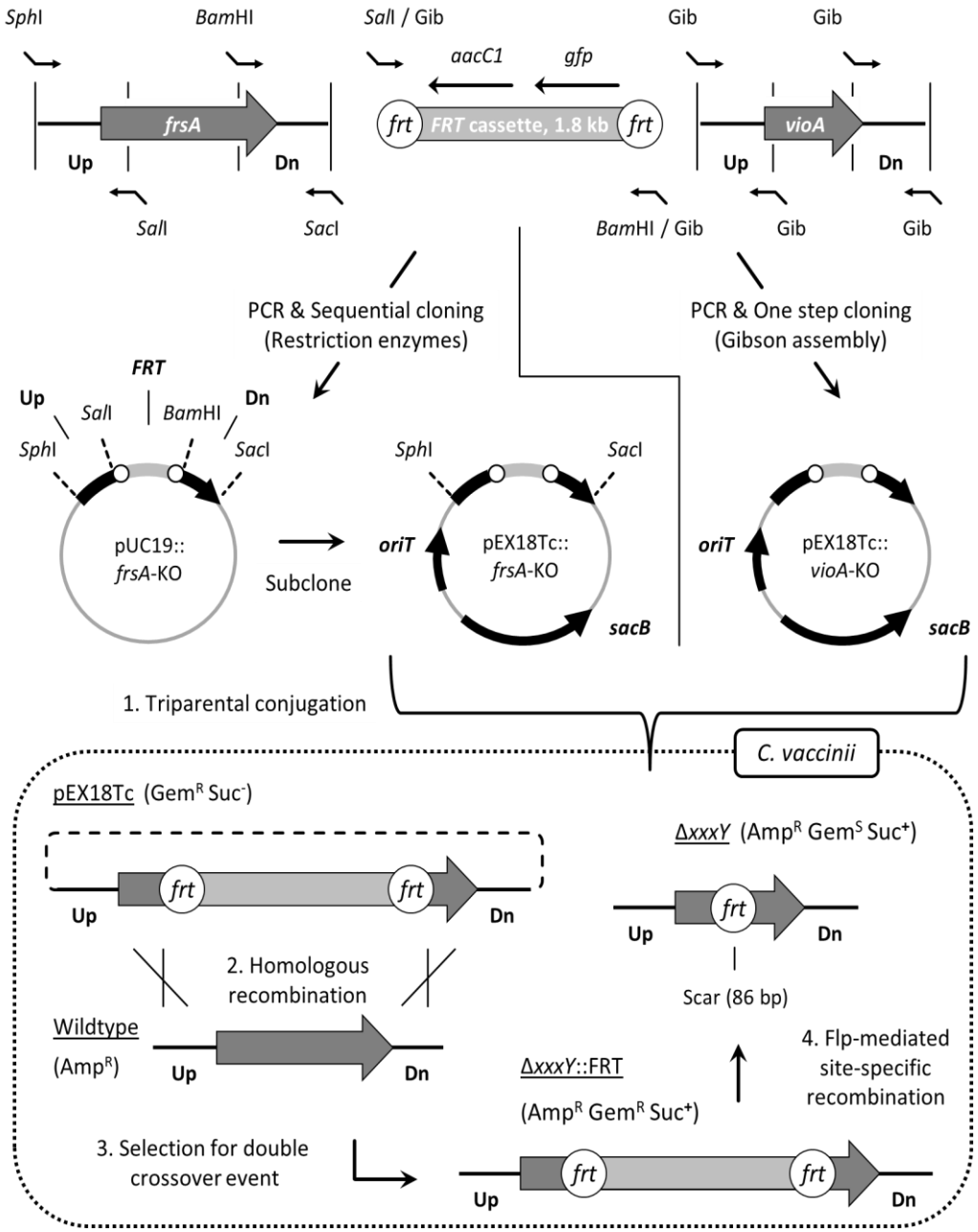


Figure 4.15: Schematic representation of the deletion strategy used for *frsA* and *vioA*, taken from Hermes *et al.*[33] This involves the construction of the knock-out vectors pEX18Tc::*frsA*-KO (sequential cloning) and pEX18Tc::*vioA*-KO (Gibson assembly) (upper half) and their use for generation of the *C. vaccinii* Δ*frsA* and Δ*vioA* deletion mutants (lower half). This involves: 1. Transfer of the knock-out vectors by triparental conjugation, 2. Exchange of the targeted gene with the *FRT* cassette (*aacC1* and *gfp* flanked by two *frt* sites) by homologous recombination, 3. Selection for an exchange by double homologous recombination as well as loss of the knock-out vector and 4. Removal of the *FRT* cassette by Flp-mediated site-specific recombination between the two *frt* sites. Restriction enzyme cutting sites used for cloning as well as overhangs necessary for Gibson assembly (Gib) are indicated. For further details see respective sections in manuscript and Materials and Methods. (Δ*xxxY*::FRT = deleted gene with integrated FRT cassette; Δ*xxxY* = deleted gene with scar; Amp^R = Ampicillin resistant; Gem^R / Gem^S = Gentamicin resistant / sensitive; Suc^+ / Suc^- = Sucrose unsusceptible / susceptible).

As described in Hermes *et al.*, in a first step, a 3,708 bp internal region of *frsA* was replaced with the FRT cassette yielding the strain *C. vaccinii* Δ*frsA*::FRT. To avoid any polar effects on downstream genes, the 3'-end of *frsA* (51 bp) was left intact. Then, the FRT cassette was removed from the genome by Flp-mediated site-specific recombination, leaving only an 86 bp scar at the site of integration. This resulted in the strain *C. vaccinii* Δ*frsA*. Integration and loss of the FRT cassette during this process were

verified by site-specific amplification of the *frsA* locus (Figure 4.16). Analysis of the butanolic extract of *C. vaccinii* Δ*frsA* by HPLC-MS revealed the absence of FR, while the production of FR-Core appeared to be enhanced (Figure 4.17).[33]

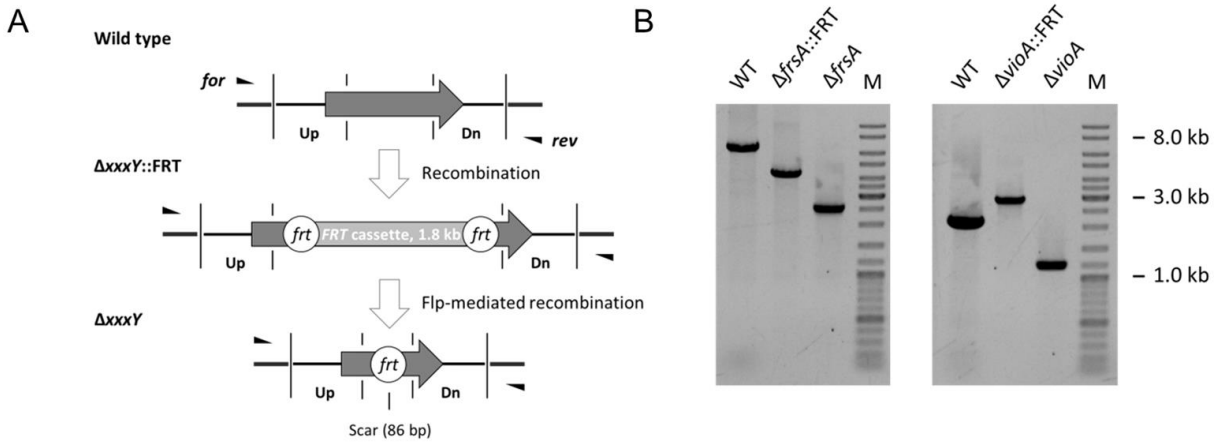


Figure 4.16: Verification of *C. vaccinii* deletion mutants, taken from Hermes *et al.*[33] A. PCR-based genotype verification applied during the construction of markerless *C. vaccinii* deletion mutants. All primers used for amplification (for/rev) bind to the regions outside the sequences used for knock-out vector construction (Up/Dn), which are shown here in red. This ensures a modification at the right locus. **B.** Q5-PCR of the *C. vaccinii* wild type as well as the different deletion mutants (Δ*frsA*::FRT, Δ*frsA*, Δ*vioA*::FRT, Δ*vioA*) with the respective verification primers (PCR-frsA_for/rev or PCR-vioA_for/rev). PCRs performed with the same primer pairs for the *C. vaccinii* Δ*frsA*/Δ*vioA* double mutant yielded identical results as for the single knock-out mutants. (Δ*xxxY*::FRT = deleted gene with integrated FRT cassette; Δ*xxxY* = deleted gene with scar).

We were able to isolate FR-Core (**8**) from *C. vaccinii* Δ*frsA* as described in 4.4.1.3, but the yields appeared insufficient for our needs. So, we decided to test if the production of FR-Core is further enhanced by interrupting the production of the purple pigment violacein (**23**). A violacein deficient *C. vaccinii* mutant (*C. vaccinii* Δ*vioA*) had already been generated in our lab, by inactivating *vioA* catalyzing the first step of violacein biosynthesis.[154] The knock-out mutant showed no production of **23** and enhanced production of FR (Figure 4.17 A). Therefore, we utilised the *C. vaccinii* Δ*frsA* mutant to construct a double mutant by replacing a 932 bp region of *vioA* with the FRT cassette with the same procedure as described above. The detailed method is described in section 6.5.8 to 6.5.9 and the verification of the mutants is depicted in Figure 4.16. Both strains with deleted production of **23** showed a white phenotype instead of the purple wild type phenotype shown in Figure 4.17 B.[34]

The three deletion mutants (Δ*frsA*, Δ*vioA*, Δ*frsA*/*vioA*) and the wild type of *C. vaccinii* were cultivated and extracted under the same conditions to compare their substance profile. This was accomplished by the master student Goran Grujicic under my supervision. The chromatograms of these extracts are pictured in Figure 4.17 A; the mass of FR (*m/z* 1002.5) is absent in both Δ*frsA* mutants, as is **23** in the Δ*vioA* mutants. The production of FR-Core in the double knock-out mutant is significantly enhanced in comparison to the single Δ*frsA* mutant. The same is true for the FR production in the Δ*vioA* mutant compared to the wild type. Consequently, construction of the double mutant strain allowed the isolation of FR-Core in a larger scale.

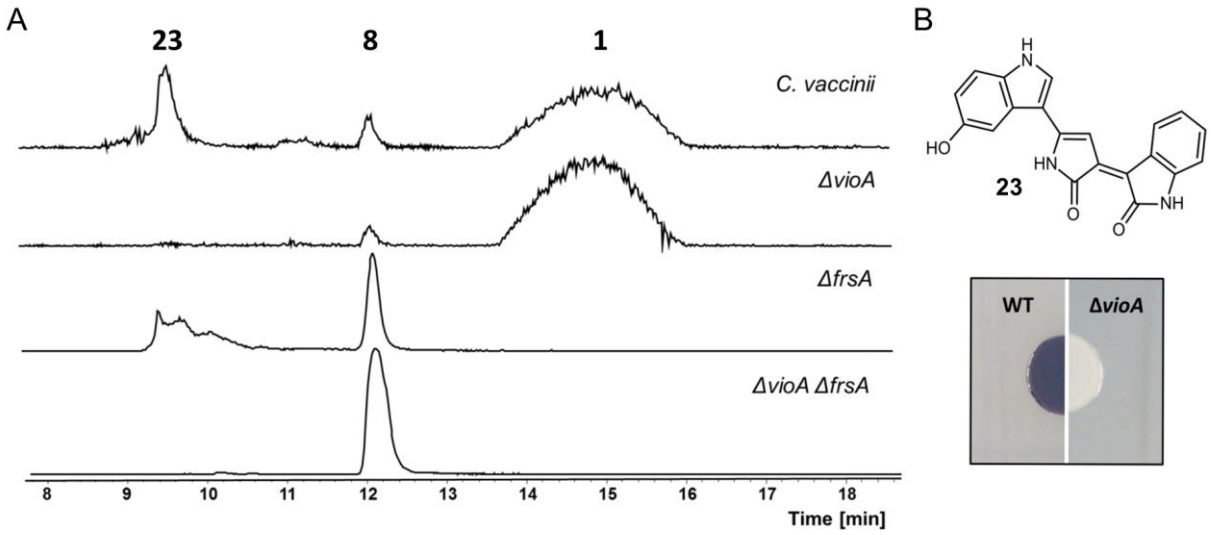


Figure 4.17: Characterisation of *C. vaccinii* knock-out mutants, taken from Hermes *et al.*[33] A. Extracted ion chromatograms of FR (*m/z* 1002.539), FR-Core (*m/z* 817.430) and violacein (*m/z* 344.101) for butanolic extracts of *C. vaccinii* and the Δ*vioA*, Δ*frsA* and Δ *frsA/vioA* deletion mutant strains from HPLC-MS experiments. **B.** Structure of violacein (**23**), Colony phenotype of wild type (WT) *C. vaccinii* changes, after deletion of *vioA*, from purple (production of **23**) to white.

4.4.1.3 Isolation and structure elucidation of FR-Core

FR-Core was first isolated from *C. vaccinii* Δ*frsA*, and a second time in higher amounts from *C. vaccinii* Δ*frsA/vioA* by the master student Goran Grujicic under my supervision. The isolation protocol was the same for both batches (see section 6.10.2). The butanolic crude extract of 4 l culture was first fractionated by flash chromatography on a C-18 reversed-phase column. Final purification was done by HPLC with a semi-preparative RP-18 column. The elution system was isocratic by 19% water and 81% methanol. The pure compound was isolated as a white powder and analysed via high-resolution MS and one- and two-dimensional NMR studies, for details, see section 6.10.2.

The molecular formula of the isolated compound was determined to be $C_{40}H_{61}N_6O_{12}$ based on HR-ESI-MS (calculated *m/z*: 817.4342; observed *m/z*: 817.4359) for $[M+H]^+$. The fragmentation during MS-MS experiments gave a similar pattern as the fragmentation of FR (see Figure 4.18), indicating the only difference between the molecules is the lacking side chain in FR-Core.

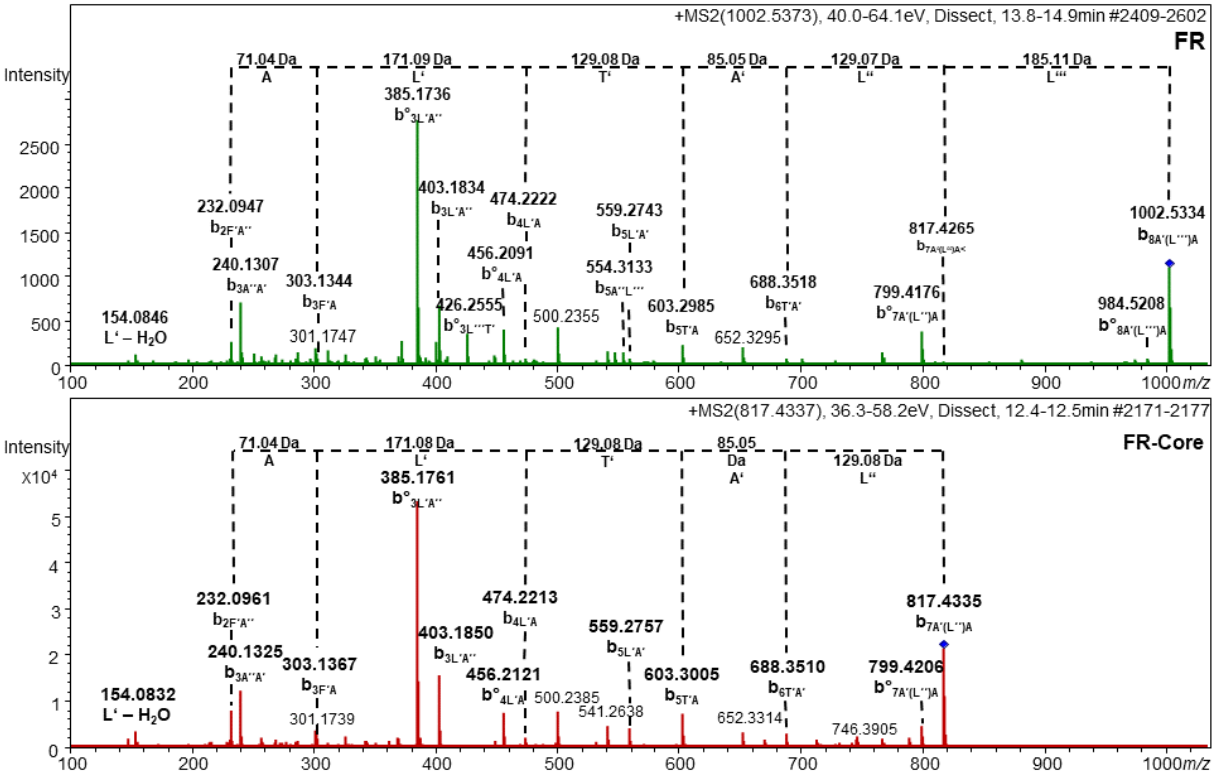


Figure 4.18: MS/MS spectra of FR (*m/z* 1002.54) and FR-Core (*m/z* 817.43). Fragmentation is labeled following the nomenclatur system by Ngoka et al. based on Biemanns modifications of Roepstorffs nomenclature in one-letter amino acid code.[155] b° = b-ion with loss of water. L' = N-acetylhydroxyleucine, A = alanine, A' = N-methylalanine, T' = N,O-dimethylthreonine, T = threonine, L'' = hydroxyleucine, L = leucine, F' = phenyllactic acid, A'' = N-methyldehydroalanine, L''' = N-propyonylhydroxyleucine.

NMR spectral data were extensively analysed in cooperation with Dr. Stefan Kehraus. The absence of all NMR signals of *N*-Pp-Hle in both the ^{1}H NMR and ^{13}C NMR spectra as well as an upfield shifted ^{1}H resonance of H-33 (δ 3.42) provided further evidence for this hypothesis. Finally, analyses of the 1D and 2D NMR data (^{1}H, ^{13}C, ^{1}H-^{1}H-COSY, ^{1}H-^{13}C-HSQC, ^{1}H-^{13}C-HMBC, ^{1}H-^{1}H-ROESY, see Figure 9.12 to Figure 9.17) unambiguously proved the structure of FR-Core (see Figure 4.19).[33] The complete assignment of hydrogen and carbon atoms is listed in Table 4.3.

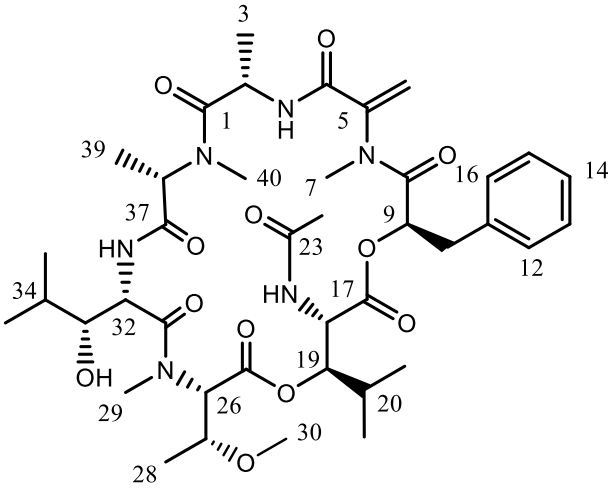


Figure 4.19: Chemical structure of FR-Core (8). Carbon atoms are numbered.

Table 4.3: ^{1}H and ^{13}C NMR spectroscopic data of FR-Core (see Figure 4.19) in acetonitrile-d_3 (^{1}H: 600 MHz; ^{13}C: 150 MHz).

Residue[a]	**No C/H**	**δ_C, mult**	**δ_H (J [Hz])**
Ala	1	172.5, C	–
	2	44.3, CH	5.11 (dq, 9.1, 6.7)
	2-NH	–	7.12 (d, 9.1)
	3	16.7, CH_3	1.19 (d, 6.7)
N-Me-Dha	4	162.2, C	–
	5	142.1, C	–
	6a	122.8, CH_2	a 5.62 (br s)
	6b	–	b 3.60 (br s)
	7	36.6, CH_3	2.89 (s)
Pla	8	168.5, C	–
	9	71.6, CH	5.56 (dd, 4.3, 10.5)
	10a	38.5, CH_2	a 3.15 (dd, 4.3, 12.5)
	10b	–	b 3.10 (dd, 10.5, 12.5)
	11	136.3, C	–
	12/16	130.5, CH	7.24[b]
	13/15	128.9, CH	7.29[b]
	14	127.5, CH	7.27[b]
N-Ac-β-OH-Leu	17	169.7, C	–
	18	53.1, CH	4.91 (dd, 2.2, 9.6)
	18-NH	–	6.88, (d, 9,6)
	19	78.3, CH	5.41 (br d, 10.0)
	20	30.4, CH	1.79 (m)
	21	18.6, CH_3	0.83 (d, 6.8)
	22	18.1, CH_3	0.87 (d, 6.8)
	23	170.5, C	–
	24	22.4, CH_3	2.02 (s)
N,*O*-Me_2-Thr	25	168.1, C	–
	26	67.9, CH	3.55 (d, 9.8)
	27	74.1, CH	3.95 (dq, 9.8, 5.9)
	28	17.8, CH_3	1.26 (d, 5.9)
	29	40.0, CH_3	3.23 (s)

	30	56.5, CH_3	3.27 (s)
β-OH-Leu	31	170.5, C	–
	32	50.3, CH	4.95 (dd, 6.9, 10.4)
	32-NH		6.94, (d, 10.4)
	33	77.1, CH	3.42 (br t, 5.8)
	34	28.8, CH	1.93 (m)
	35	20.3, CH_3	1.02 (d, 6.7)
	36	15.6, CH_3	0.93 (d, 6.7)
N-Me-Ala	37	169.6, C	–
	38	61.4, CH	3.61 (q, 6.7)
	39	12.3, CH_3	1.38 (d, 6.7)
	40	38.0, CH_3	3.17 (s)

[a] Residues: Ala = alanine, *N*-Me-Dha = *N*-methyldehydroalanine, Pla = 3-phenyllactic acid, *N*-Ac-*β*-OH-Leu= *N*-acetyl-3-hydroxyleucine, *N*,*O*-Me_2-Thr = *N*,*O*-dimethylthreonine, *β*-OH-Leu = 3-hydroxyleucine, *N*-Me-Ala = *N*-methylalanine,. [b] overlapping resonances.

4.4.2 Transesterification assay with the synthesized substrate 22 and FR-Core

After gaining preparative access to FR-Core and synthesis of the side chain precursor **22**, the anticipated transfer of the side chain from the T domain mimic onto the free hydroxy group of the core molecule by the FrsA TE domain could now be investigated *in vitro*. We overexpressed the $FrsA_{TE}$ in *E. coli* BL21 with a hexahistidine tag and purified it using affinity chromatography (see Figure 4.8). Afterwards, the pure protein was incubated with **22** and FR-Core at 22 °C for 5 h and the assay extracted with CH_2Cl_2 for LC-MS analysis (see section 6.11.3). If the formation of FR takes place, we would expect the *m/z* of 1002.53 for the protonated FR to appear. In comparison, the negative control with heat-inactivated protein should only contain FR-Core, but no FR. Figure 4.20 shows the extracted mass traces for FR in the assay and its negative control. Indeed, as hypothesised, we observed the transfer of the *N*-Pp-Hle onto FR-Core. The turnover was not complete, as there was still a lot of FR-Core left after the assay was stopped. Still, we could confirm the proposed function of $FrsA_{TE}$ and were, therefore, able to characterize the second intermolecular-acting TE domain within a bacterial NRPS.

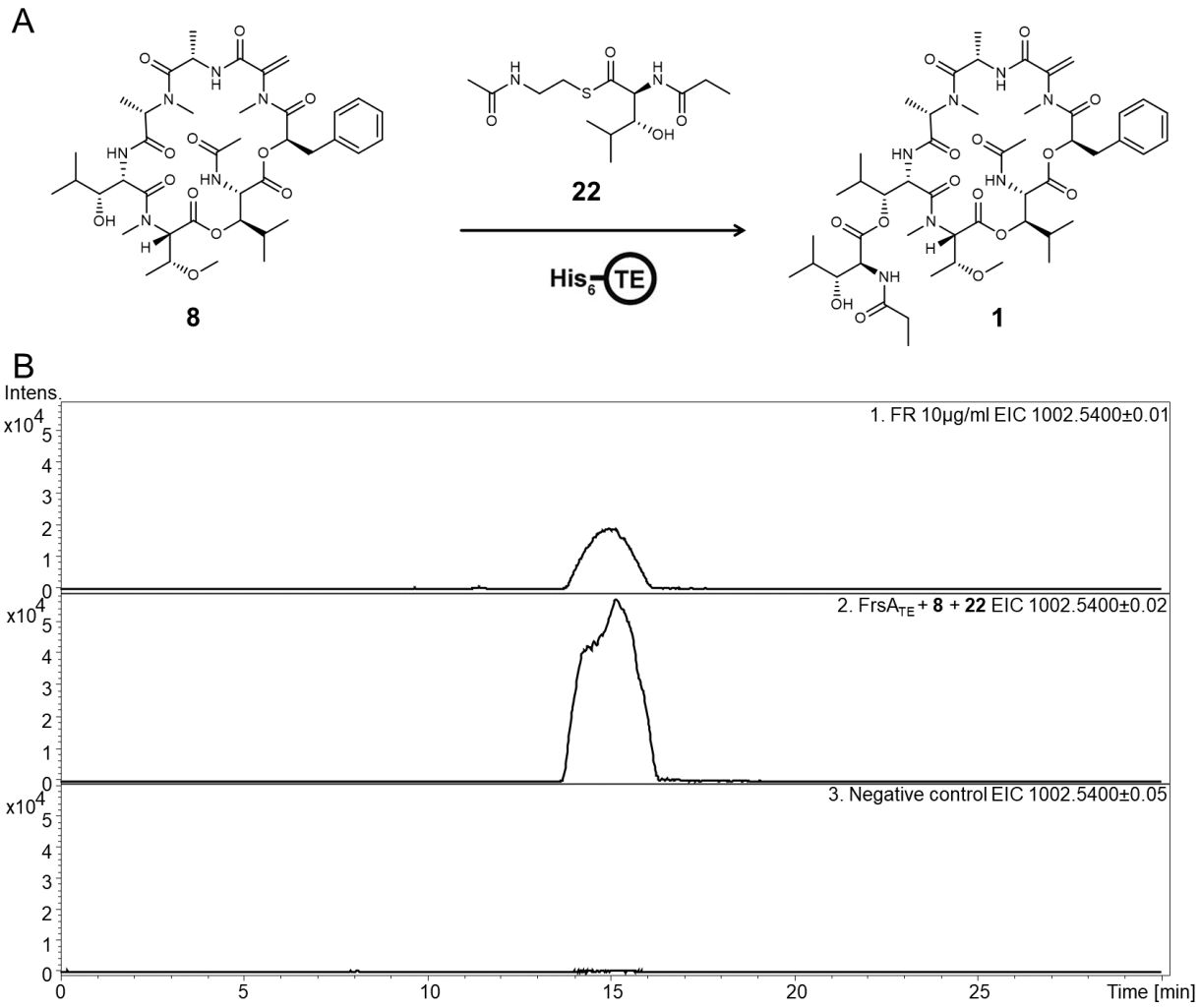


Figure 4.20: ***In vitro*** **assay with FrsA$_{TE}$. A.** Reaction scheme of intermolecular transesterification of **22** onto FR-Core to yield FR. **B.** Extracted ion chromatograms of FR (*m/z* 1002.54) from HPLC-MS experiments; **1.** FR standard (10 µg/ml); **2.** purified FrsA$_{TE}$ incubated with FR-Core and **22**; **3.** negative control with heat-inactivated protein. His$_6$-TE = hexahistidine tagged thioesterase domain.

We also employed the whole FrsA construct, coexpressed with FrsB, in this assay to test if further domains of the protein complex have any influence on the turnover. It might have been expectable to have a higher turnover than with the standalone TE domain, as the complete module is closer to the natural situation, but instead, a lower turnover was observed (Figure 4.21). This might be explained by a higher netto concentration of the TE domain when expressed alone as the same mg/ml amount was used for both assays. For better comparison, the assays would have to be conducted with the same molarity. Another explanation could be the accessibility of the binding sites in the TE domain, which might be better when no other domains are present. As the substrate **22** is not bound to the T domain of FrsA in this assay, the SNAC residue could also interfere with other domains, thus hindering its access to the TE domain.

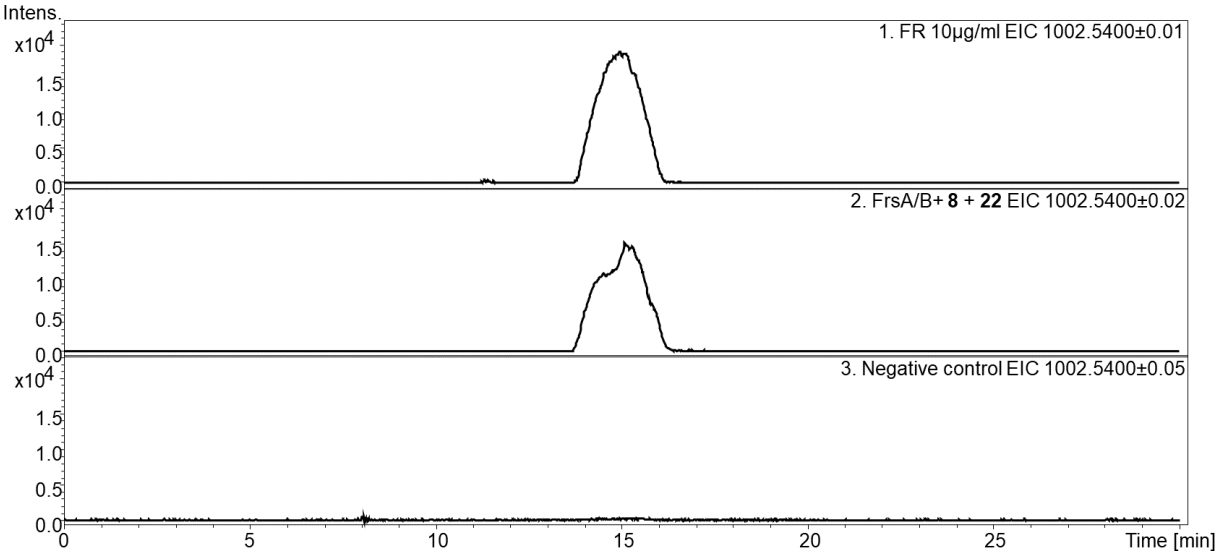


Figure 4.21: *In vitro* assay with FrsA. Extracted ion chromatograms of FR (*m/z* 1002.54) from HPLC-MS experiments; **1.** FR standard (10 µg/ml), **2.** purified FrsA/B, incubated with FR-Core and **22**, **3.** negative control with heat-inactivated protein.

4.4.3 *In vitro* assembly of the side chain and transfer to FR-Core

After the functionality of the TE domain had been proven, we wanted to test the complete assembly and transfer of the side chain *in vitro*, using the whole FrsA module coexpressed with FrsB and the separately expressed FrsH. The assay combines the steps of the C domain assay and the transesterification assay: first, the A domain uses ATP to activate L-leucine, which is bond to the T domain. Then the hydroxylation catalysed by activated FrsH takes place and the C domain connects the propionyl-CoA to the amino group of leucine. Finally, the TE domain transfers the whole side chain onto FR-Core (see reaction scheme in Figure 4.22 A). This experiment also led to the production of FR (Figure 4.22 B), confirming *in vitro* activity of the whole enzymatic machinery.

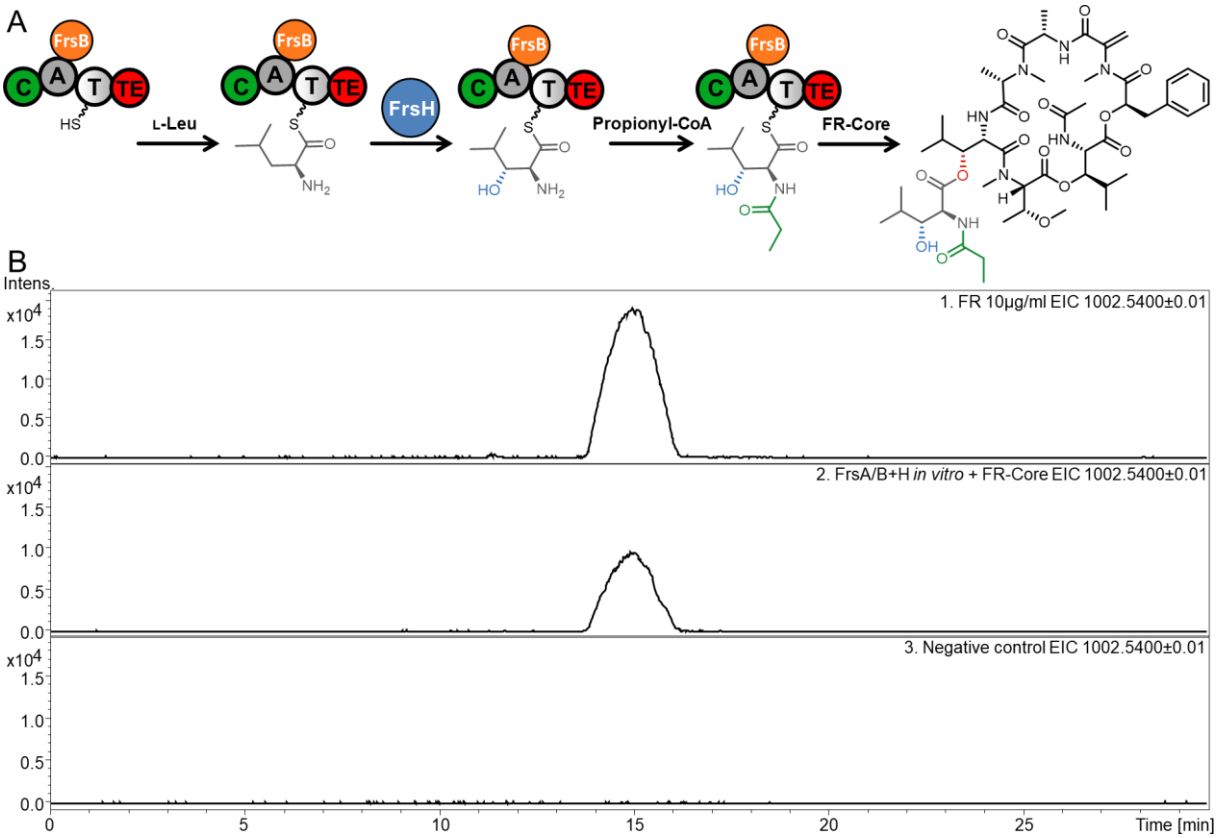


Figure 4.22: *In vitro* side chain assembly and transfer assays with FrsA/B and FrsH. A. Reaction scheme of *N*-Pp-Hle formation and intermolecular transesterification onto FR-Core to yield FR; **B.** Extracted ion chromatograms of FR (*m/z* 1002.54) from HPLC-MS experiments; **1.** FR standard (10 µg/ml); **2.** Purified FrsA/B, FrsH incubated with propionyl-CoA, L-Leu and FR-Core; **3.** negative control with heat-inactivated protein. A = adenylation domain, C = condensation domain, T = thiolation domain, TE = thioesterase domain.

Taken together these three different *in vitro* assay experiments not only confirmed the hypothesised biosynthesis of the side chain of FR and the extraordinary transesterification catalysed by the TE domain, but they also open the door for further experiments to probe the substrate specificity of the FrsA domains. The promiscuity of the A and C domain has already been investigated in the previous assays (compare section 4.3.1 and 4.3.2) and was now used to generate FR derivatives *in vitro*. While there are different ways to engineer NRPS assembly lines by domain and module swapping, the method of precursor directed biosynthesis uses inherent substrate promiscuity of biosynthetic enzymes to generate altered natural product structures.[156] For the success of this approach, not only the activating enzyme needs to accept the altered substrate, but also the downstream enzymes have to tolerate the altered structure of the growing peptide chain substrate.

4.4.4 *In vitro* generation of FR analogues

The previous experiments revealed some substrate promiscuity for the A and C domain of FrsA, and the variety of FR derivatives isolated from *A. crenata* and *C. vaccinii* also indicate the acceptance and incorporation of structurally different substrates *in vivo*.[11,36] One natural FR derivate with an altered side chain has been isolated: FR-2 carries an acetyl moiety instead of the propionyl residue of *N*-Pp-Hle.[11] As acetyl-CoA is a very common product of glycolysis and a widely used precursor for C_2 building

blocks in different biosynthesis pathways,[157] it is most likely easily accessible for FR biosynthesis as well. As the $C_{starter}$ domains of FrsA and FrsD are 92.4% identical, while FrsD is supposed to catalyse the assembly of *N*-acetyl-hydroxyleucine (*N*-Ac-Hle), this also suggests the possible use of an activated acetyl residue, i.e., acetyl-CoA as an alternative substrate for the C domain.

We, therefore, performed the *in vitro* side chain assembly and transfer assay as described above but with different CoA substrates. As a positive control, we used propionyl-CoA, which leads to the already observed FR (*m/z* 1002.54) production. We exchanged propionyl-CoA for acetyl-CoA in order to produce FR-2 (*m/z* 988.53). Additionally, we tested butyryl-CoA to determine if a longer acyl chain is also accepted to form a new, unnatural "FR-butyryl" (*m/z* 1016.55). As negative controls, heat-inactivated proteins were used.

The results of the assays are pictured in Figure 4.23. The positive control with propionyl-CoA led to the production of FR and confirmed the results in section 4.4.3. The *in vitro* production of FR-2 was also successful since the mass and retention time fitted to the isolated standard of FR-2 (**5**). This proves the promiscuity of the $C_{starter}$ domain and might explain the detected high amounts of FR-2 in *C. vaccinii* cultures, as acetyl-CoA is easily accessible under laboratory conditions and hence can compete with propionyl-CoA. Interestingly, also the assay with butyryl-CoA yielded a novel signal for the calculated *m/z* of 1016.55. Its retention time was slightly higher than that of FR and in line with the earlier FR-2, fitting to the longer acyl chain. In Figure 4.24 the MS/MS fragmentation patterns of the novel compound and FR are depicted and compared. From *m/z:* 799.42 (loss of side chain, and loss of H_2O, see Reher *et al*.),[36] the compounds have the same fragmentation pattern, indicating, that the additional methylation of FR (M+14 Da) is in the side chain and thus stems most likely from the incorporation of a butyryl group from the precursor butyryl-CoA in position (1) instead of the propionyl residue in FR like reported for FR-3 in position (2).[36]

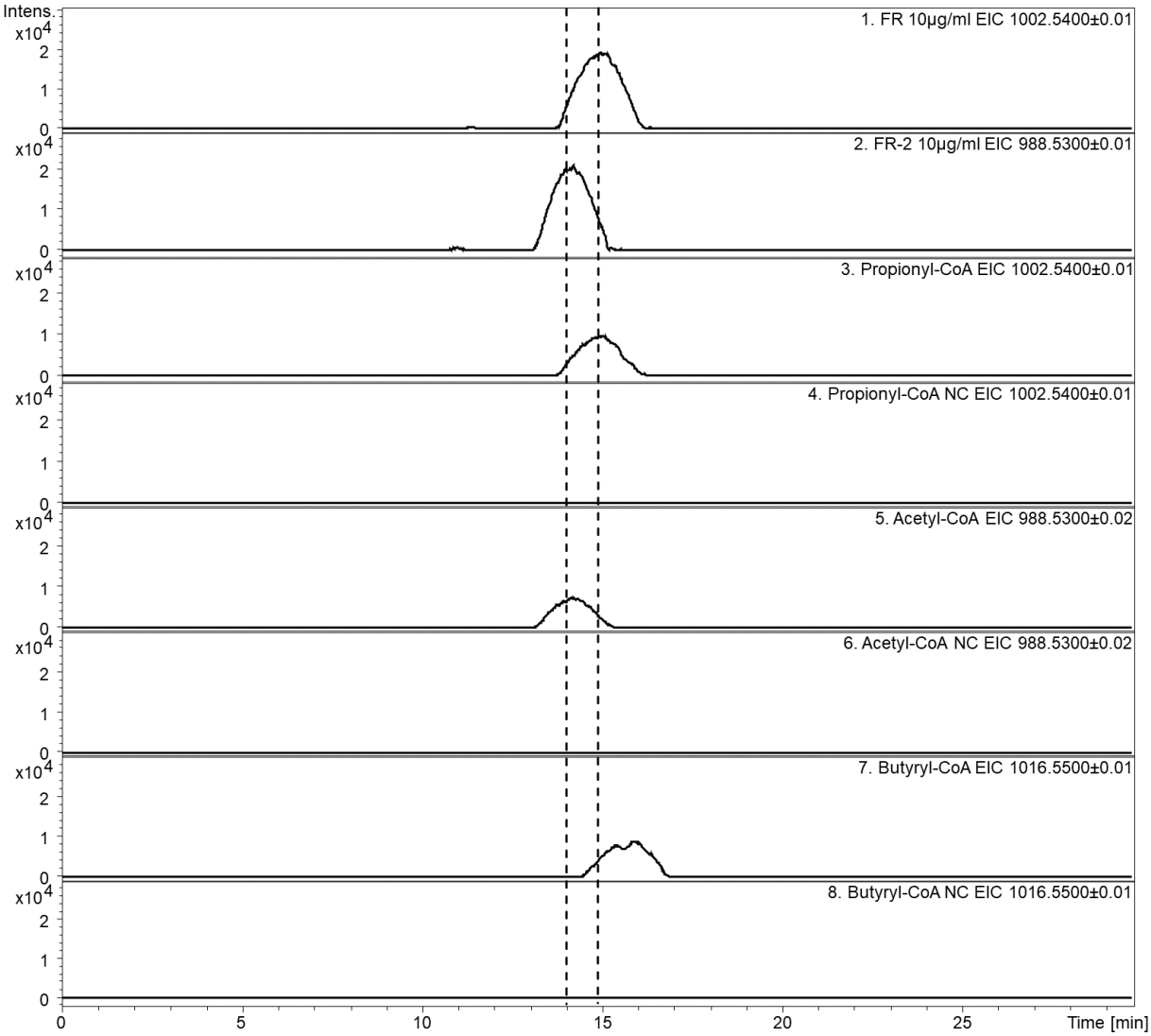


Figure 4.23: Different Acyl-CoAs for *in vitro* side chain assembly and transfer assays with FrsA/B and FrsH. Extracted ion chromatograms of FR (*m/z* 1002.54), FR-2 (*m/z* 988.53) and FR-butyryl (*m/z* 1016.50) from HPLC-MS experiments; **1.** FR standard (10 µg/ml); **2.** FR-2 standard (10 µg/ml); **3.** Purified FrsA/B, FrsH incubated with propionyl-CoA, L-Leu and FR-Core, and **4.** negative control with heat-inactivated protein; **5.** Purified FrsA/B, FrsH incubated with acetyl-CoA, L-Leu and FR-Core, and **6.** negative control; **7.** Purified FrsA/B, FrsH incubated with butyryl-CoA, L-Leu and FR-Core, and **8.** negative control.

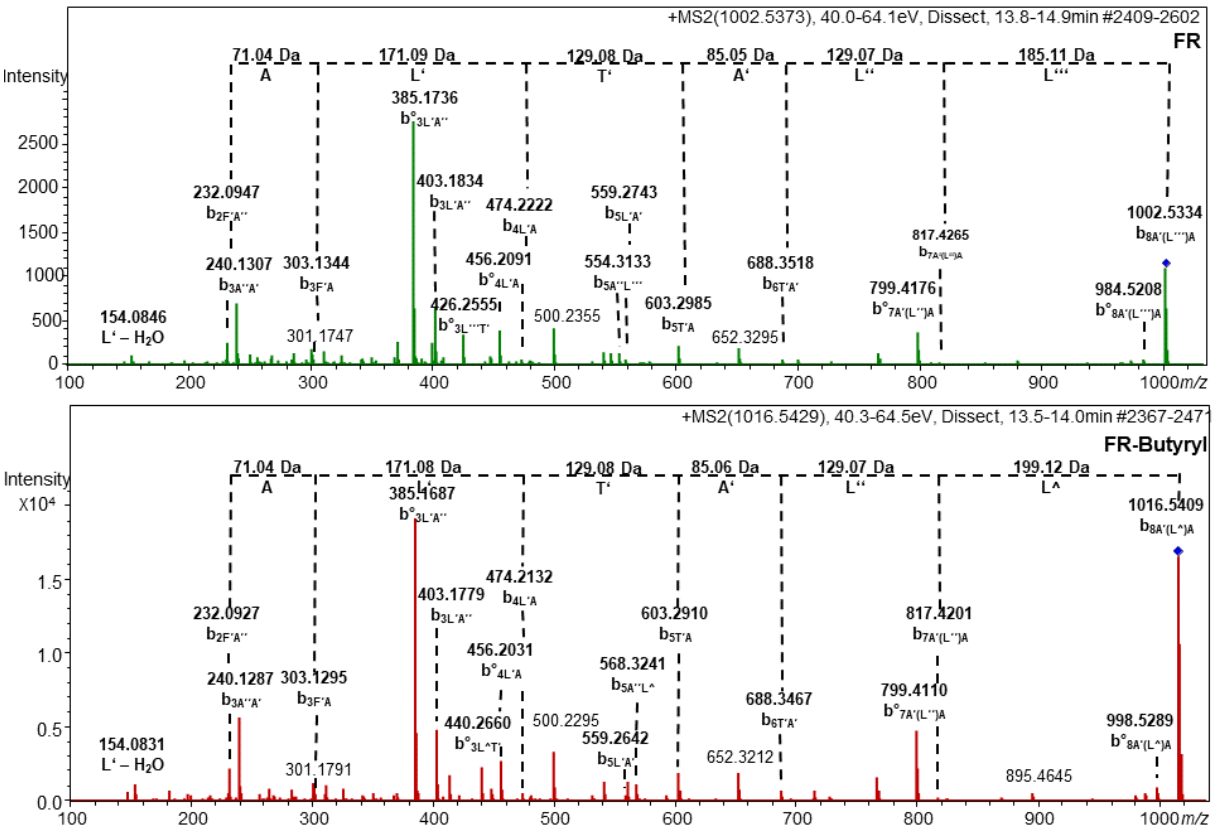


Figure 4.24: MS/MS spectra of FR (*m/z* 1002.54) and FR-butyryl (*m/z* 1016.55). Fragmentation is labeled following the nomenclatur system by Ngoka et al. based on Biemanns modifications of Roepstorffs nomenclature in one-letter amino acid code.[155] b° = b-ion with loss of water. L’ = N-acetylhydroxyleucine, A = alanine, A’ = N-methylalanine, T’ = N,O-dimethylthreonine, T = threonine, L’’ = hydroxyleucine, L = leucine, F’ = phenyllactic acid, A’’ = N-methyldehydroalanine, L’’’ = N-propyonylhydroxyleucine, L^ = N-butyrylhydroxyleucine.

These data indicated the *in vitro* synthesis of a new FR derivative with an *N*-butyryl-3-hydroxyleucine side chain, but the *in vitro* assay yielded insufficient amounts of the compound for structure elucidation. Upscaling of the *in vitro* assay was hampered by the low conversion rate of FR-Core to FR-butyryl and the limited supply of FR-Core. Thus, the next approach to gain access to FR-butyryl was the precursor-directed biosynthesis which is presented and discussed in 4.5.

We also tested if the substrate promiscuity of the A domain may lead to the integration of structurally different amino acids in the side chain assembly. For this assay, L-leucine, propionyl-CoA and FR-Core were used as a positive control to obtain FR and heat-inactivated proteins were used for the negative controls. The tested amino acids D-leucine and L-isoleucine are stereoisomers of L-leucine, so the resulting mass of the products would be expected to be the same as for FR, only the retention times might vary. As shown in Figure 4.25, the positive control generated FR and all negative controls showed no FR as expected. For the D-leucine, no signal for FR production was detected. This could imply, that the TE domain might act as a gatekeeper, a mechanism reported for a noncanonical TE domain in nocardicin biosynthesis, which only catalysed offloading of the NRPS when the β-lactamisation took place before.[158] An analogous phenomenon was reported for the TE domain in the biosynthesis of the glycopeptide antibiotic teicoplanin, where the TE domain is selective for cross-linked aglycones with

an unusual extended N-terminal linker region.[159] In general, most TE domains are known so exert low substrate selectivity for the loading stage, giving way for pathway evolution, which is required to access new chemical diversity and maintain long term evolutionary fitness.[40] In the case of $FrsA_{TE}$, there is some substrate promiscuity visible, but the stereochemistry of the assembled side chain might be crucial for the activity of the TE domain. In contrast to the D-leucine, the assay with L-isoleucine resulted in a peak with the expected mass, so the steric organisation of the lipophilic side chain of the amino acid does not seem to be as influential as the first stereocenter for the activity of the TE domain. Interestingly, FrsH, which catalyses the hydroxylation at position 3 of leucine, also seems to be able to hydroxylate a tertiary carbon atom at position 3. Of course, these findings would need to be verified by more detailed structure elucidation of the product. However, similar to the FR-butyryl, this compound could not be isolated from the *in vitro* assay in sufficient amounts.

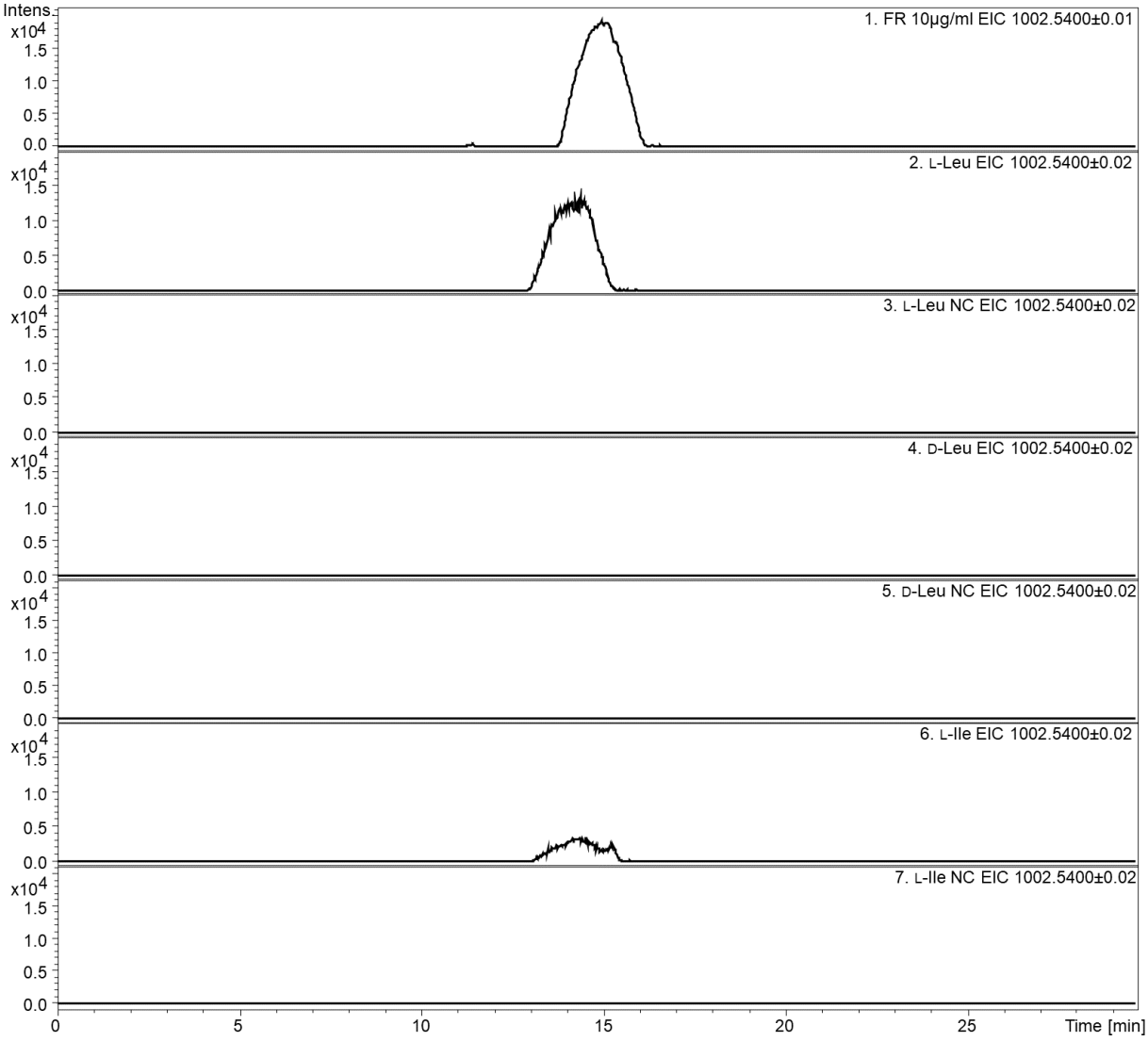


Figure 4.25: Different amino acids for *in vitro* side chain assembly and transfer assays with FrsA/B and FrsH. Extracted ion chromatograms of FR (*m/z* 1002.54) from HPLC-MS experiments; **1.** FR standard (10 µg/ml); **2.** Purified FrsA/B, FrsH incubated with propionyl-CoA, L-Leu and FR-Core, and **3.** negative control with heat-inactivated protein; **4.** Purified FrsA/B, FrsH incubated with propionyl-CoA, D-Leu and FR-Core, and **5.** negative control; **6.** Purified FrsA/B, FrsH incubated with propionyl-CoA, L-Ile and FR-Core, and **7.** negative control.

We also tested the substrate specificity of $FrsA_{TE}$ concerning the acceptor substrate. Our previous results strongly suggested FR-Core to be the natural substrate of $FrsA_{TE}$, implying this step takes place after the cyclisation catalysed by $FrsG_{TE}$. The linear depsipeptide product of FrsD-G might however also be a possible substrate, but as this substrate could not be obtained in the time of this thesis, we were not able to test it in the assay. Instead, we tested the minimal substrate **20** (L-Hle) to see if the TE domain might recognize this substrate and transfer the side chain onto the hydroxy group of Hle. The side chain harbours an additional free hydroxy group, so this reaction could take place repeatedly, resulting in oligomers. We analysed the LC-MS data from this assay for all possible masses, but no formation of any product was observed (see Figure 4.26 and Supplemental Figure 9.24).

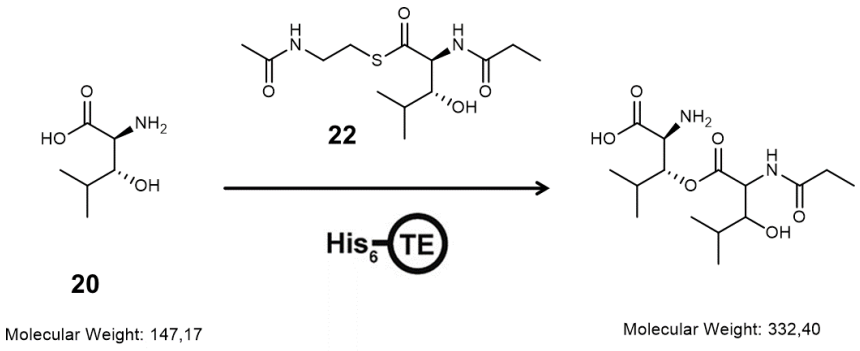


Figure 4.26: Reaction scheme of the hypothetical intermolecular transesterification of 22 onto L-Hle (20).

It was also tested whether another cyclic depsipeptide containing a Hle moiety could serve as a substrate for $FrsA_{TE}$. The antibiotic natural product lysobactin (also known as katanosin B, **24**) is commercially available and has a Hle moiety in its backbone. We performed the transesterification assay with lysobactin as substrate instead of FR-Core. The proposed reaction is shown in Figure 4.27. In this experiment also no formation of a new product could be observed (see Supplemental Figure 9.25), which is however not surprising, as the three-dimensional structure of lysobactin is vastly different to the one of FR-Core.

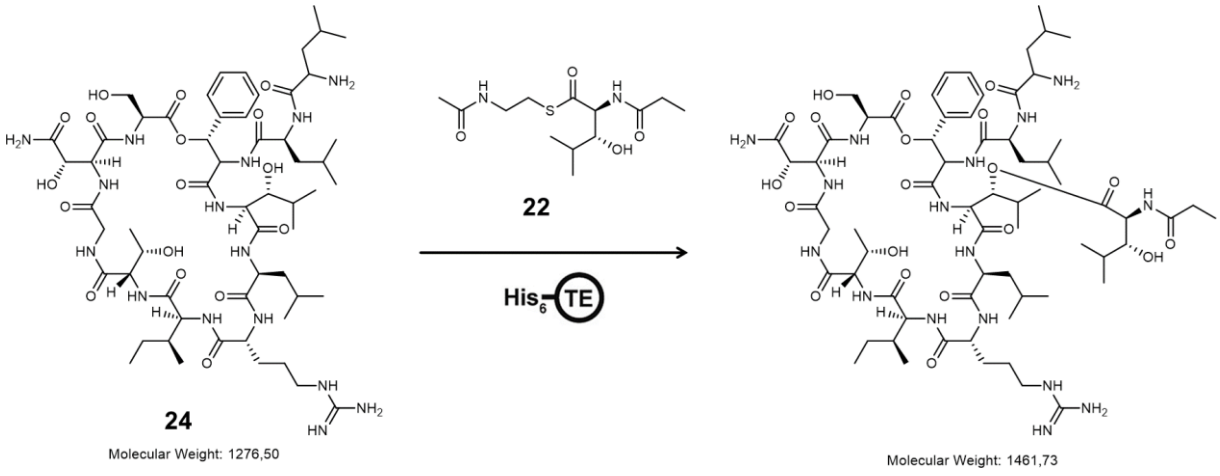


Figure 4.27: Reaction scheme of the hypothetical intermolecular transesterification of 22 onto lysobactin (24).

The next step for this project would be the generation or isolation of the linear depsipeptide analogue to FR-Core and to use it as substrate for the assay, either as an open molecule, as SNAC or bound to the FrsG T domain. If the peptide is synthesized as CoA-thioester, the phosphopantetheinyl transferase Sfp of *Bacillus subtillis* could be used to load the substrate onto the expressed T domain,[150] an established method for TE domain analyses that was also used for $NocB_{TE}$ studies.[158] The T domain bound peptide would be the substrate closest to nature and could be also used to verify the function of $FrsG_{TE}$. Another way to get hands on this linear molecule would be the cultivation of a $frsG_{TE}$ knock-out mutant. Without $FrsG_{TE}$ the biosynthesis should stop at the stage of the linear heptapeptide which might be released from the assembly line as such by other mechanisms like TE-type II mediated hydrolysis.[160] Otherwise, the side chain transesterification would take place independent from the cyclisation and yield a branched, linear FR molecule, which would disprove the hypothesis of FR-Core to be the natural substrate. Additionally, a systematic evaluation of other, structurally diverse cyclic or linear peptide substrates is expected to shed more light on $FrsA_{TE}$ acceptor specificity.

So far, our results support the hypothesis of the transesterifying TE domain and revealed some promiscuity for the composition of the side chain. This opens up new possibilities for biosynthetic engineering as these noncanonical intermolecular transesterification reactions can create depsipeptide bonds at polyketide or peptide side chain hydroxy functions. This biosynthetic principle can lead to the generation of potent natural products like salinamide A, which is produced by $Sln9_{TE}$-catalysed transesterification in a marine *Streptomyces* bacterium and has strong antibiotic properties.[43] Besides $Sln9_{TE}$ and $FrsA_{TE}$, whose functions are now well investigated, there is another recently identified TE domain in the PKS/NRPS hybrid biosynthesis cluster of the cytotoxic necroximes that might have a similar function.[161] Investigations of the *nec* BGC also revealed two TE domains. $NecH_{TE}$ is supposed to catalyse the cyclisation of the oxime-substituted benzolactone enamide core molecule Necroxime C and D and the monomodular NRPS NecA assembles a peptidic side chain, which is presumably transferred by $NecA_{TE}$ to yield Necroxime A or B. A Δ*necA* mutant abolished the production of Necroxime A and B which supports the theory of the NecA function, but the TE domain was not investigated in detail.[161] Detailed comparative structural analyses of these three noncanonical TE domains could be a great basis for further exploitation efforts of this new type of enzyme for chemoenzymatic purposes. Our structural investigations of $FrsA_{TE}$ are still ongoing and discussed in section 4.8.

4.4.5 Comparison of $FrsA_{TE}$ and $FrsG_{TE}$

To test the similarity of the two *frs* TE domains, from FrsA and FrsG, we performed the transesterification assay with $FrsG_{TE}$. The bioinformatic analyses in section 4.1 show phylogenetic relationships even though the proposed activity is quite different. FrsG is supposed to catalyse the offloading and cyclisation of linear FR-Core. To examine, if $FrsG_{TE}$ is also capable to catalyse the intermolecular transesterification, we incubated $FrsG_{TE}$ with $FrsA_{CAT}$, FR-Core and all necessary

substrates for the side chain assembly (see Figure 4.28). As expected, this did not result in the formation of FR while the same set up with the standalone $FrsA_{TE}$ led to the production of the complete depsipeptide. So, in this case, the clustering in the phylogenetic tree does not indicate the same mechanism of release for the domains.

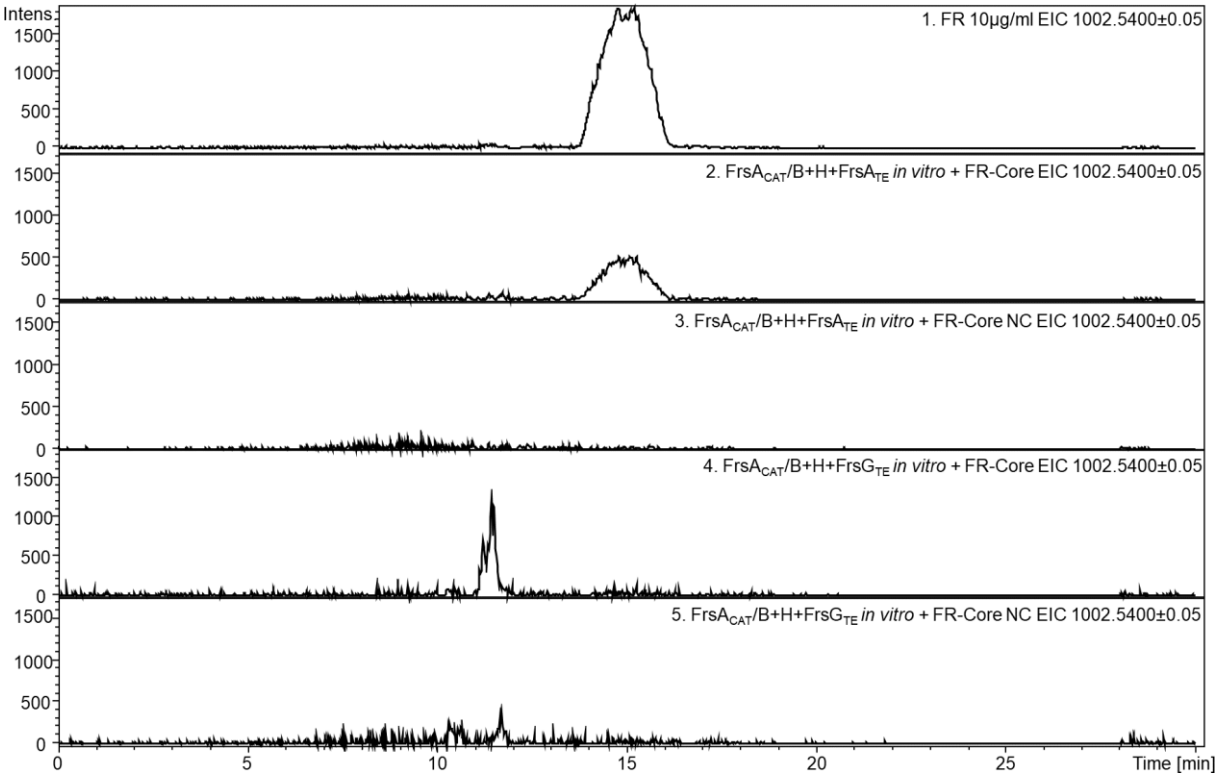


Figure 4.28: *In vitro* side chain assembly and transfer assays with $FrsA_{CAT}$/B, FrsH and standalone $FrsA_{TE}$ or $FrsG_{TE}$. Extracted ion chromatograms of FR (*m/z* 1002.54) from HPLC-MS experiments; **1.** FR standard (10 µg/ml); **2.** Purified $FrsA_{CAT}$/B, FrsH and $FrsA_{TE}$ incubated with propionyl-CoA, L-Leu and FR-Core; **3.** negative control with heat-inactivated protein; **4.** Purified $FrsA_{CAT}$/B, FrsH and $FrsG_{TE}$ incubated with propionyl-CoA, L-Leu and FR-Core; **5.** negative control with heat-inactivated protein.

4.5 Precursor-directed biosynthesis of FR-5 (19)

The bioassays described in section 4.4.4 did not yield sufficient amounts of the new FR derivative "FR-butyryl" for structure elucidation. We thus thought about using the FR producing *C. vaccinii* strain for the production of this new compound *in vivo*. There are different approaches to utilise the substrate promiscuity of PKS or NRPS systems to generate new analogues, nicely reviewed by Ladner and Williams.[156] For example, precursor feeding was successful for the pacidamycin pathway, where different tryptophan analogues were incorporated into the natural product by supplementation of the amino acids to the wild type bacterium *Streptomyces coeruleorubidus*.[162] The scheme in Figure 4.29 shows the principle of precursor-directed biosynthesis: by addition of non-native or non-natural building block to the cultivation of a producer organism the production of a natural product analogue can be induced parallel to the production of the natural product.

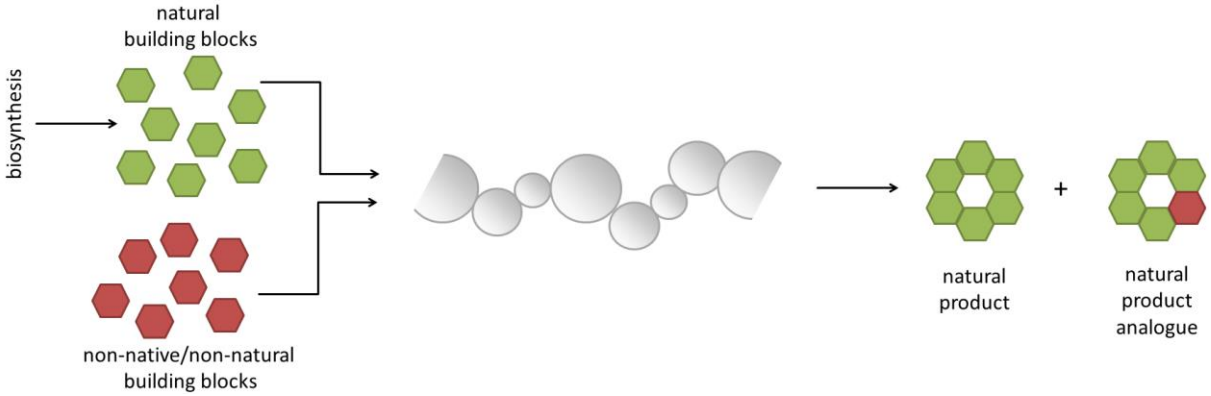


Figure 4.29: Schema of precursor-directed biosynthesis. Non-native or non-natural building blocks are installed through inherent promiscuity of a module and/or domain and usually results in a mixture of the natural product and non-natural analogue. Adapted from Ladner and Williams.[156]

The direct precursor for the production of FR-butyryl would be butyryl-CoA, which was a sufficient substrate for the *in vitro* synthesis of the side chain. As butyryl-CoA is an expensive compound and high amounts of the precursor were needed for the feeding experiments, we decided to feed butyrate to *C. vaccinii*, assuming the bacterium might convert it to butyryl-CoA prior to integration into the peptide. This approach was successful in the production of new avermectin derivatives, where two different CoAs were the natural substrates and feeding of different carboxylic acids led to the production of new analogues.[163] In this case, the biosynthesis of the natural precursor acids was disrupted in a mutant strain of *Steptomyces avermitilis*, so the approach would be a precursor-directed mutasynthesis, but we assumed that the activation of the carboxylic acids should work in a wild type strain as well. The advantage of mutasynthesis is, that no natural product is produced besides the new natural product analogue. As the natural precursor, in our case, is propionyl-CoA, which is present in different pathways within the bacterial cell, its production cannot easily be turned off.[164] So, we still expect some amount of FR in our feeding studies. Butyrate or butyric acid, on the other hand, is not an artificial building block. This reduces the potential of toxic effects during feeding. We supplemented the medium with a high concentration (20 mM) of butyrate to put the production of butyryl-CoA in favour over the production of other CoAs. For the feeding experiments, we used the chemically defined M9 minimal medium to reduce the influence of other potential precursors. Wiebke Hanke had investigated the production of FR in rich LB medium in comparison to M9 minimal medium. Interestingly, the production of FR and its derivative FR-2 is enhanced in M9 medium. The production of FR-2, which is in average lower than that of FR in LB medium, is in M9 medium drastically increased, the yields of FR-2 are approximately 2.5 times higher than those of FR (see Supplementary Figure 9.26). This might be due to the fact, that the availability of acetyl-CoA in minimal medium is much higher than that of propionyl-CoA. The first experiments were performed in small scale to test if the feeding of butyrate leads to the production of FR-butyryl. Besides, we investigated the effects of the feeding solution on the growth of *C. vaccinii* and the production rate of FR and FR-2.

4.5.1 Feeding experiments with butyrate

To test the effect of 20 mM sodium butyrate on the cultivation of *C. vaccinii* wt in M9 minimal medium, three repeats of *C. vaccinii* with feeding solution, one flask with *C. vaccinii* without feeding solution and one flask with only blank medium including feeding solution were cultivated for two days. The cultures were extracted after 48 h with n-butanol, as the production of FR and its derivatives has shown to reach its maximum between 36 and 48 h of cultivation (Supplementary Figure 9.26). The crude extracts were analysed via LC-MS and the extracted ion chromatograms (EIC) for the masses of FR (*m/z* 1002.5), FR-2 (*m/z* 988.5) and FR-butyryl (*m/z* 1016.5) were compared (see section 6.11.4). As expected, no peaks for any of the three substances were detected in the blank medium (Supplementary Figure 9.27). The chromatograms of the control without butyrate feeding solution are shown in Figure 4.30. These results are in line with the data of Wiebke Hanke: In M9 medium, the production of FR-2 is higher than the production of FR. The EIC of the calculated FR-butyryl *m/z* value (1016.5) for the protonated derivative, gives an exceedingly small signal with a retention time slightly after FR. We did however not expect such a signal with the mass of FR-butyrate in the control solution, but there are two known FR derivatives with the same *m/z* of 1016.5, FR-3 (**3**) and FR-4 (**6**), as described in section 2.2.5. Both compounds were isolated in minor amounts from in the leaves of *A. crenata*,[36] but their production in *C. vaccinii* was not investigated so far. These traces could thus be FR-3 or FR-4, traces of endogenously produced FR-butyryl, as butyryl-CoA might be produced by *C. vaccinii* during cultivation in minor amounts as well.

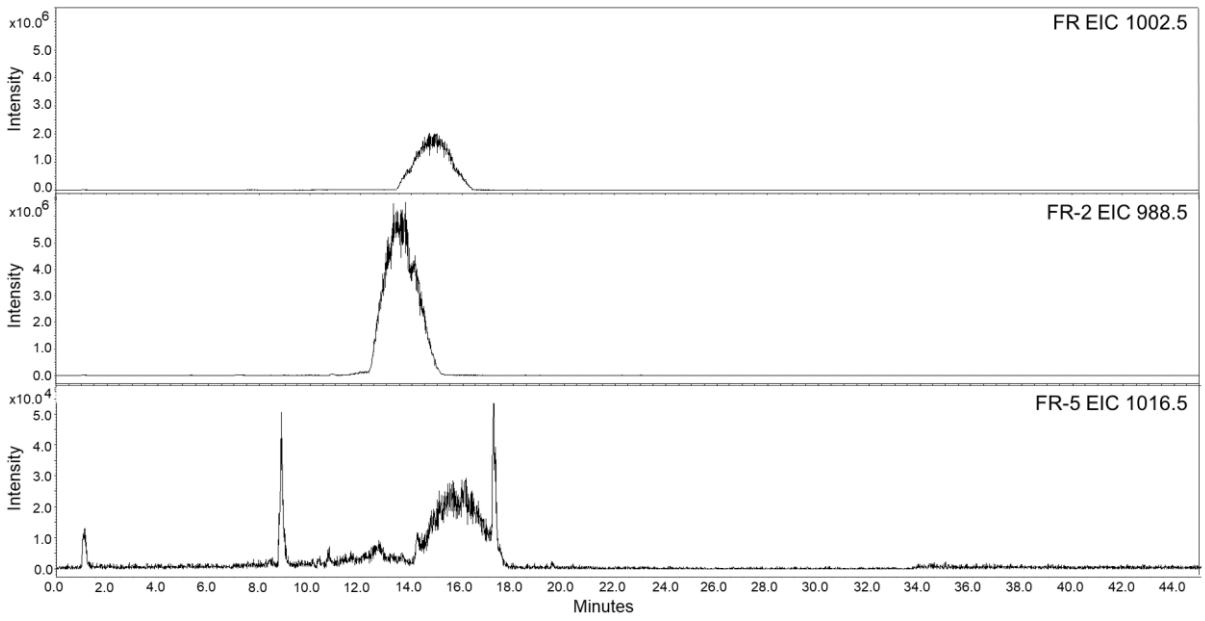


Figure 4.30: Extracted ion chromatograms of the *C. vaccinii* culture in M9 medium without 20 mM butyrate feeding solution. Top: EIC of FR (*m/z* 1002.5); **middle:** EIC of FR-2 (*m/z* 988.5); **bottom:** EIC of FR-butyryl (*m/z* 1016.5).

Another possibility might be a further, undescribed FR derivative with one additional methylene group in any position. Nevertheless, the addition of the feeding solution drastically changed the situation, as shown in Figure 4.31. While there are decreased but still high amounts of FR-2, the production of FR

was significantly decreased. On the other hand, there is a strong signal with the *m/z* of 1016.5, with notably higher intensity than the signal of FR-2. This experiment gives strong evidence for the formation of FR-butyryl, as the addition of butyrate is the only change compared to the culture conditions of the control in Figure 4.30.

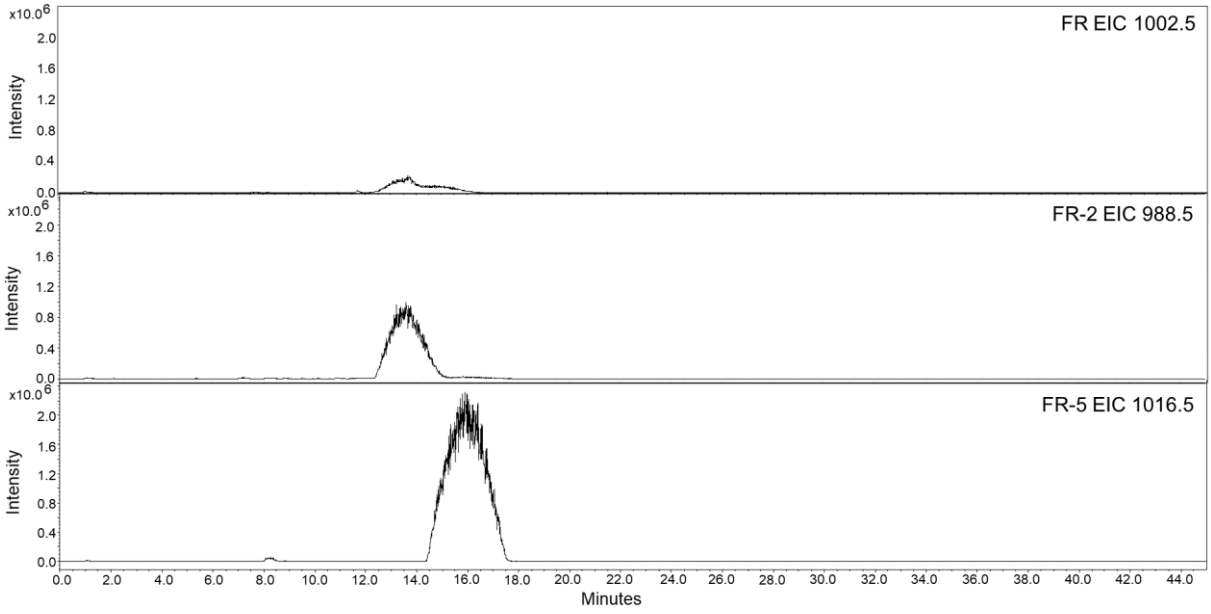


Figure 4.31: Extracted ion chromatograms of the *C. vaccinii* culture in M9 medium with 20 mM butyrate feeding solution. Top: EIC of FR (*m/z* 1002.5); **middle:** EIC of FR-2 (*m/z* 988.5); **bottom:** EIC of FR-butyryl (*m/z* 1016.5).

Interestingly, there was also a change in the growth of the culture observable: the control turned dark purple due to the production of violacein (**23**) (see Figure 4.17), which is normal for a 48 h *C. vaccinii* culture, but the three repeats with the feeding solution reached only a light violet state. Accordingly, the extracts of the feeding repeats contained visibly lower amounts of the pigment **23**. As the optical densities of the cultures were not recorded, there could be two explanations for the observed decrease in violacein production: on the one hand, butyrate might somehow interfere directly with the production of **23,** or that the culture growth of the bacteria is slowed down due to the feeding which would lead to a later start of production of **23**, that is known to be regulated by quorum sensing.[165,166] The latter seems to be the more likely explanation as the production of **23** has been investigated in detail and no influence of carbon acids was reported.[167]

The crude extract obtained from the feeding experiment was also analysed with high-resolution LC-MS, where the exact masses and fragmentation confirmed the identity of the peaks. Also, the mass traces of FR, FR-2 and FR-butyryl could be compared directly, as shown in Figure 4.32. This result confirmed, that FR-butyryl is the main part of the FR derivatives in the crude extract of the feeding experiment and the amount is comparable to the amount of FR in the control. These results indicate sufficient production of the new compound for upscaling, isolation and subsequent structure elucidation. The next section deals with the isolation of FR-butyryl which is from now on termed FR-5.

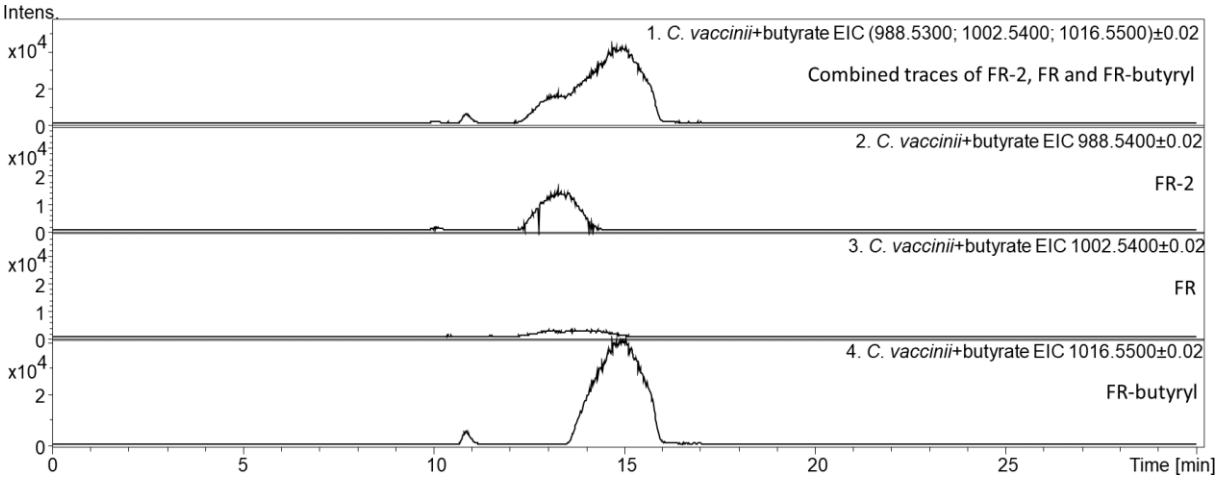


Figure 4.32: LC-MS data of the *C. vaccinii* culture in M9 medium with 20 mM butyrate feeding solution. 1. Combined EIC of FR (*m/z* 1002.54), FR-2 (*m/z* 988.54) and FR-butyryl (*m/z* 1016.55); **2.** EIC of FR-2 (*m/z* 988.54); **3.** EIC of FR (*m/z* 1002.54); **4.** EIC of FR-butyryl (*m/z* 1016.55).

4.5.2 Isolation and structure elucidation of FR-5 (19)

For the preparative isolation of FR-5 *C. vaccinii* was cultivated for 48 h in 4.5 L M9 medium supplemented with 20 mM butyric acid and subsequently extracted with *n*-butanol (see 6.10.3). Analogously to the isolation of FR-Core, the crude extract was fractionated via flash chromatography and the final purification was done by semi-preparative HPLC, (see section 4.4.1.3). 10 mg of the pure compound were isolated as white powder and analysed via high-resolution MS and one- and two-dimensional NMR studies. Based on HR-ESI-MS (calculated *m/z*: 1016.5556; observed *m/z*: 1016.5507) for $[M+H]^+$, the molecular formula was determined to be $C_{50}H_{77}N_7O_{15}$. The MS/MS fragmentation spectrum is shown and discussed in section 4.4.4 and gave a pattern, comparable to the fragmentation spectrum of FR (see Figure 4.24). This proved, that the only difference between the molecules is the changed acyl residue of the side chain. The structure of the molecule (see Figure 4.33) was confirmed by detailed NMR analysis, performed by Dr. Stefan Kehraus. The complete assignment of the hydrogen and carbon atoms is listed in Table 4.4.

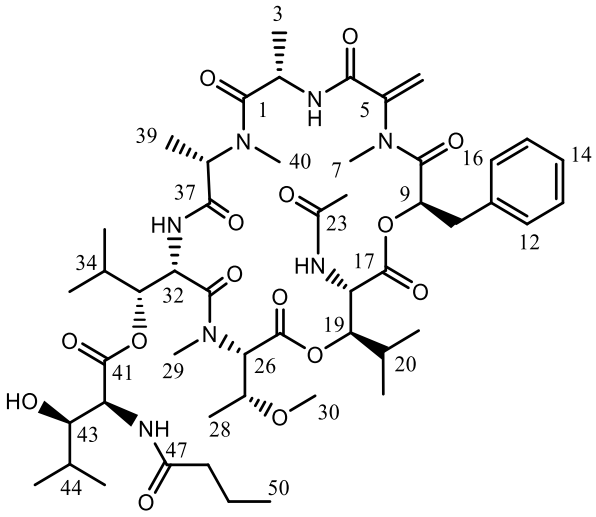


Figure 4.33: Chemical structure of FR-5 (19). Carbon atoms are numbered.

Table 4.4: ^{1}H and ^{13}C NMR spectroscopic data of compound FR-5 in $CDCl_3$ (^{1}H: 300 MHz; ^{13}C: 75 MHz).

Residue[c]	No C/H[b]	δ_C[a], mult	δ_H[a] (mult, *J* [Hz])	COSY	HMBC
Ala	1	172.5, C	–		
	2	45.7, CH	4.90 (m)	3, 2-NH	1, 3, 4
	2-NH	–	8.53 (d, 9.1)	2	2, 4
	3	18.0, CH_3	1.38 (d, 6.7)	2	1, 2
N-MeDha	4	163.9, C	–		
	5	145.3, C	–		
	6	106.7, CH_2	a 5.31 (brs)	6b	5
		–	b 5.07 (brs)	6a	5
	7	36.2, CH_3	3.13 (s)		5, 8
D-Pla	8	167.7, C	–		
	9	72.6, CH	5.20 (dd, 4.2, 8.3)	10a, 10b	8, 10, 17
	10	36.6, CH_2	a 3.08 (dd, 4.2, 14.8)	9, 10b	9, 11, 12/16
		–	b 2.97 (dd, 8.3, 14.8)	10a	9, 11, 12/16
	11	136.0, C	–		
	12/16	129.6, CH	7.24[d]	13/15	10, 14
	13/15	128.6, CH	7.27[d]	14, 12/16	11
	14	126.9, CH	7.23[d]	13/15	12/16
N-Ac-β-OH-Leu	17	169.2, C	–		
	18	50.3, CH	5.24 (brd, 10.0)	18-NH, 19	17, 19
	18-NH	–	7.55, (d, 10,0)	18	18, 23
	19	77.7, CH	5.10 (brd, 10.0)	18, 20	20, 25
	20	28.8, CH	1.86 (m)	19, 21, 22	
	21	18.9, CH_3	1.01 (d, 6.8)	20	20, 22
	22	18.8, CH_3	0.85 (d, 6.8)	20	20, 21
	23	171.4, C	–		
	24	22.5, CH_3	2.21 (s)		23
N-MeThr(*O*Me)	25	166.5 C	–		
	26	64.4, CH	4.05 (d, 9.6)	27	25
	27	72.3, CH	3.74 (m)	26, 28	26, 28
	28	16.3, CH_3	1.16 (d, 5.8)	27	26, 27
	29	28.7, CH_3	2.68 (s)		26, 31
	30	57.2, CH_3	3.40 (s)		27
β-OH-Leu	31	171.2, C	–		
	32	46.6, CH	5.35 (d, 9.9)	32-NH, 33	31, 33
	32-NH		6.74, (d, 9.9)	32	32, 37
	33	77.0, CH	5.30, (d, 10.0)	32, 34	41
	34	30.5, CH	1.70 (m)	33, 35, 36	35, 36
	35	19.4, CH_3	1.08 (d, 6.7)	34	33, 34, 36
	36	18.3, CH_3	0.82 (d, 6.7)	34	33, 34, 35
N-MeAla	37	169.9, C	–		
	38	56.4, CH	4.70 (q, 6.8)	39	37, 39
	39	14.3, CH_3	1.37 (d, 6.8)	38	37, 38
	40	31.4, CH_3	2.87 (s)		1, 38
N-But-β-OH-Leu	41	170.2, C	–		
	42	56.8, CH	4.55 (brd, 7.8)	42-NH	41
	42-NH	–	7.17 (d, 7.8)	42	42, 47
	43	78.2, CH	3.71 (m)	43-OH, 44	41, 44
	43-OH	–	6.87 (d, 4.2)	43	

44	30.0, CH	1.96 (m)	43, 45, 46	45, 46
45	20.5, CH_3	1.15 (6.7)	44	43, 44, 46
46	18.5, CH_3	0.85 (d, 6.7)	44	43, 44, 45
47	173.9, C	–		
48	37.4, CH_2	2.47, m	49	
49	19.1, CH_2	1.68, m	48, 50	50
50	13.8, CH_3	0.95 (t, 7.5)	49	48, 49

[**a**]Assignments are based on extensive 1D and 2D NMR measurements (HMBC, HSQC, COSY). ^{13}C-NMR spectra were recorded at 75 MHz. [**b**]Numbers according to Supplementary Figure 24. [**c**] Residues: Ala = alanine, *N*-MeDha = *N*-methyldehydroalanine, D-Pla = D-3-phenyllactic acid, *N*-Ac-β-OH-Leu= *N*-acetylhydroxyleucine, *N*-MeThr(*O*Me) = *N*,*O*-dimethylthreonine, β-OH-Leu = β-hydroxyleucine, *N*-MeAla = *N*-methylalanine, *N*-But-β-OH-Leu = *N*-butyryl-β-hydroxyleucine.[d] overlaying resonances

The successful precursor-directed biosynthesis of FR-5 leads the way for analogous experiments. Regarding the results obtained with L-isoleucine in section 4.4.4 (Figure 4.25), feeding of ^{13}C marked L-isoleucine to *C. vaccinii* might be a possibility to verify if this amino acid is actually incorporated into the molecule *in vivo*. The elevated mass of the isotopic label would confirm the presence of a new FR derivative even in trace amounts.

4.6 Bioactivity of FR-Core and FR-5

The structure of a secondary metabolite is highly relevant for its bioactivity, and this is especially true for FR.[11] Numerous structure-activity relationship (SAR) studies have been conducted with FR derivatives and are discussed in detail in section 2.4. Taken together they demonstrate, that even small structural changes in the pharmacophore drastically decrease the inhibitory activity of FR derivatives towards Gαq.[11] The new analogues obtained in this study, FR-Core and FR-5, were investigated for their bioactivity as well. For both molecules, the Gαq inhibition capacities were tested in cooperation with the workgroup of Prof. Kostenis in our institute. Julian Patt and Judith Alenfelder performed the dynamic mass redistribution (DMR) experiments. For competitive Gαq binding studies and molecular docking with FR and FR-Core, we cooperated with Jan Hendrik Voß and Vigneshwaran Namasivayam from the lab of Prof. Christa Müller in the Institute for Pharmaceutical & Medicinal Chemistry of the University of Bonn. The results gave insights about the binding affinity of FR-Core compared to FR. Additionally, we investigated the ecological role of FR and FR-Core by testing its effect against nymphs of the stinkbug *Riptortus pedestris*, which was conducted by Dr. Tsubasa Ohbayashi and Dr. Peter Mergaert from the Institute for Integrative Biology of the Cell of the University of Paris-Saclay. These experiments are described and discussed in the following sections. The used methods are described in our recent publication.[33]

4.6.1 Dynamic mass redistribution (DMR)

As described in Hermes *et al*., we first measured the Gαq inhibitory activity of FR-Core in comparison to FR in CRISPR-Cas9 genome-edited HEK293 cells deficient in Gαq and Gα11, using real-time live-cell phenotypic biosensing, based on dynamic mass redistribution (DMR).[168,169] In our experimental setup, endogenous or overexpressed Gαq-coupled receptors (herein: carbachol-activated muscarinic M3 receptor) showed activity in DMR assays only upon re-expression of Gαq. We first examined FR and observed the full inhibition of Gαq-activity with an IC_{50} value of 0.45 µM, which correlates well with previously determined inhibitory activities in related assays.[125] Next, we examined FR-Core and

observed significant differences in Gαq-inhibition capabilities. It was about 16-fold less potent than FR (Figure 4.34, Table 4.5).[33] This shows the importance of the side chain for the activity of FR which might be due to the structural influence of the side chain in the binding to the Gαq protein. The side chain was also identified as part of the pharmacophore of FR,[11] so it is not surprising, that its absence reduces the activity.

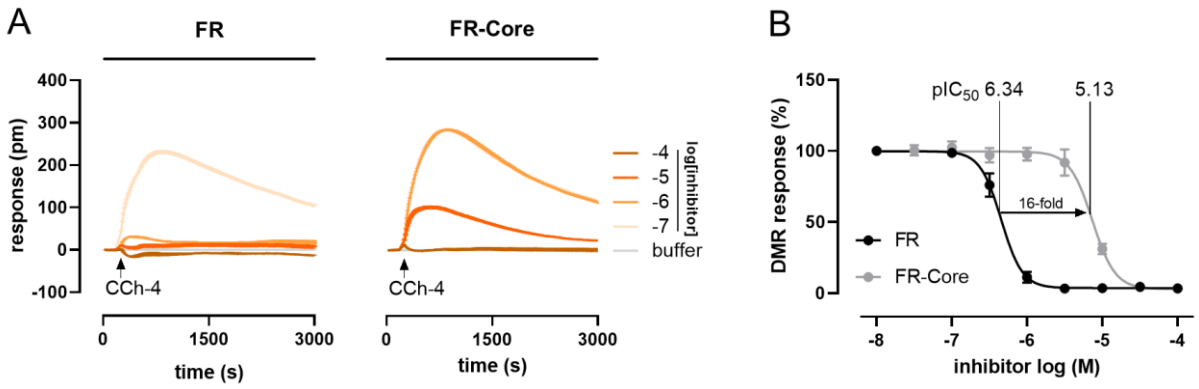


Figure 4.34: Pharmacological characterisation of FR-Core on Gαq-mediated signalling. A. Concentration-dependent inhibition of cell responses induced with carbachol (CCh) [100 µM] by **FR** and **FR-Core** in HEK293 Gαq/Gα11-null cells transfected to express wild type Gαq. Data shown are representative real-time recordings (mean + s.e.m., technical triplicates) of at least four independent experiments. **B.** Concentration-dependent inhibition of activated Gαq proteins by **FR** and **FR-Core** as determined by label-free whole cell DMR biosensing. DMR recordings are representative (mean + s.e.m.) of at least four independent biological replicates conducted in triplicate.[33]

The activity of the butyryl analogue FR-5 was tested analogously. It showed a sevenfold reduced Gαq inhibition capacity compared to FR (Figure 4.35). This shows that the elongation of the acyl residue in the side chain by one methylene group has a considerable negative effect on Gαq inhibition.

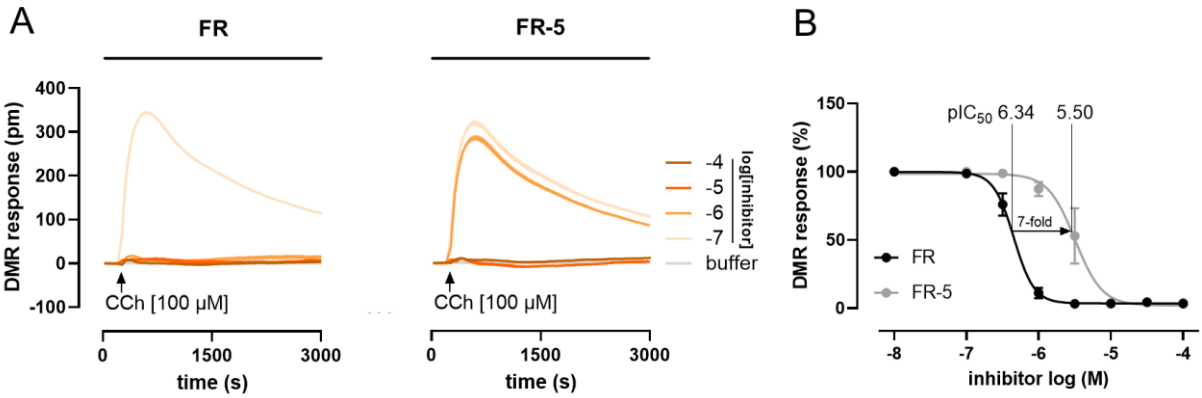


Figure 4.35: Pharmacological characterisation of FR-5 on Gαq-mediated signalling. A. Concentration-dependent inhibition of cell responses induced with carbachol (CCh) [100 µM] by **FR** and **FR-5** in HEK293 Gαq/Gα11-null cells transfected to express wild type Gαq. Data shown are representative real-time recordings (mean + s.e.m., technical triplicates) of at least four independent experiments. **B.** Concentration-dependent inhibition of activated Gαq proteins by **FR** and **FR-5** as determined by label-free whole cell DMR biosensing. DMR recordings are representative (mean + s.e.m.) of at least four independent biological replicates conducted in triplicate.[33]

Interestingly, the shorter acyl residue in the side chain of the natural FR derivate FR-2 had a weaker effect on the IC_{50} value, shown in Table 4.5.[11] This indicates, that the acyl chain length might have been optimized during evolution to generate maximal affinity towards its target. Further variations of the

length of the acyl chain in the side chain to generate FR derivatives with enhanced activity are probably not expedient. A crystal structure of FR cocrystallised with a Gαq protein might give more insights about the spatial possibilities of the molecule. So far, only a crystal structure of YM, bound to a Gαq protein is available. Analogously to FR-2, YM has an acetyl residue in the side chain.[2] This makes this region especially interesting for the comparison of FR and YM, also in consideration of the molecular docking studies of Kushak *et al*. with FR and YM radiotracers. They discussed that FR has two more lipophilic residues, the propionyl at the side chain (1) and the Hle at position (2), in contrast to YM. These "handles" anchor FR in the binding pocket like a dowel forming a latch, while YM lacks those anchor points and can therefore more readily be released.[56] FR-5, with butyryl instead of acetyl or propionyl, would have one even more lipophilic residue. But the increased chemical space seems to be disadvantageous for inhibition and has a strong negative influence, explaining the lower IC_{50} value of FR-5 compared to FR.

Table 4.5: Quantification of FR, FR-Core and FR-5 inhibitory activities at wild type Gαq in HEK Gαq/Gα11-null cells. IC_{50} values were determined by nonlinear regression on concentration-effect data and represent the mean of 'n' independent biological replicates performed as technical triplicates.[33] Data of FR-2 and YM for comparison from Reher *et al.*[11]

#	**pIC_{50}± s.e.m.**	**IC_{50} [µM]**	**n**
FR (**1**)	6.34±0.03	0.45	11
FR-Core (**8**)	5.13±0.04	7.34	4
FR-5 (**19**)	5.50±0.06	3.18	4
FR-2 (**5**)		1.79[11]	
YM (**2**)		1.55[11]	

4.6.2 Competitive binding studies and molecular docking

To compare the binding affinity of FR and FR-Core to Gαq, competitive binding studies were performed. Competition binding assays with FR and FR-Core against the radiolabeled FR-derivative [^{3}H]PSB-15900,[19] performed with human platelet membrane preparations, revealed a 207-fold decrease in binding affinity for FR-Core compared to FR. Here, the measured *p*IC_{50}-value decreased from 7.88 ± 0.09 for FR to 5.56 ± 0.04 for FR-Core (Figure 4.36).

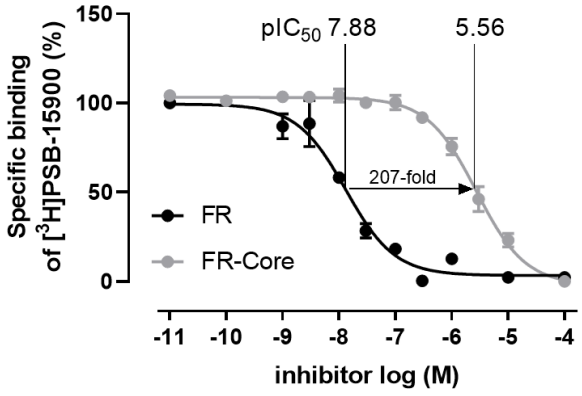


Figure 4.36 Competition binding experiments of FR and FR-Core versus the FR-derived radiotracer [^{3}H]PSB-15900 at human platelet membrane preparation (50 µg protein per vial), incubated at 37 °C for 1 h.[33]

To get insights into the molecular interaction of FR-Core to its Gαq binding pocket, a docking-based model of the Gαq protein in complex with FR-Core based on the co-crystal structure of Gαq-βγ bound with YM (PDB ID: 3AH8, resolution: 2.9 Å)[2] was generated and compared with the model created for FR. We found that the macrocyclic core structures likely display identical orientation, and the isopropyl groups of both molecules are anchored inside the binding pocket through hydrophobic interactions. In both structures, *N*,*O*-dimethylthreonine and D-phenyllactic acid are predicted to form hydrogen bond interactions with the side chain of R60 and the main chain of P185, respectively (Figure 4.37). In addition, the side chain **21**, which is absent in FR-Core, is proposed to form strong interactions (two hydrogen bonds) with the backbone of E191. The intramolecular interactions formed between the side chain and *N*-methylalanine likely stabilise the binding conformation of FR inside the pocket. All these additional interactions are believed to be the reason for the much higher binding affinity of FR.[33]

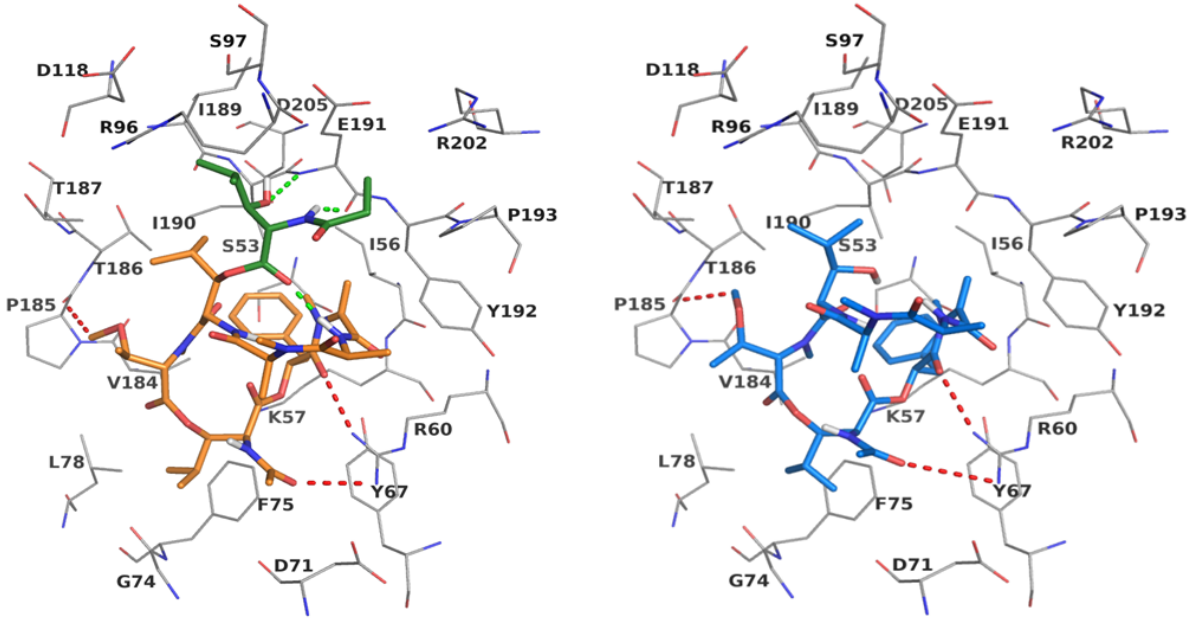


Figure 4.37: Molecular Docking of FR and FR-Core in the binding pocket of the Gaq protein. Docked poses of FR (left, represented in sticks and coloured in orange, the *N*-Pp-Hle group present only in FR is coloured in green) and FR-Core (represented in sticks and coloured in blue) in the binding pocket of the Gaq protein shown as line representation. Some of the interactions common for FR and FR-Core are indicated by red dotted lines, and the interactions specific for FR are shown as green dotted lines. Oxygen atoms are coloured in red, nitrogen atoms in blue and polar hydrogen atoms in white.[33]

4.6.3 Insect toxicity assays

The ecology of FR has only been started to be investigated. Crüsemann *et al*. had tested the compound *in vivo* against mammals and insects with quite some effect, suggesting FR might be used as a defence against herbivores.[32] To check if the lethal effect on insects is strongly influenced by the side chain, we fed FR and FR-Core parallel to the nymphs of the stinkbug *Riptortus pedestris* and investigated the survival rate over twelve days. While high concentrations (0.2 µg/µl) of both metabolites killed all

insects after nine days, at 0.04 µg/µl, only feeding of FR had lethal effects, while FR-Core did not affect the animals (Fig. 4c), demonstrating improved *in vivo* toxicity of FR compared to FR-Core.[33]

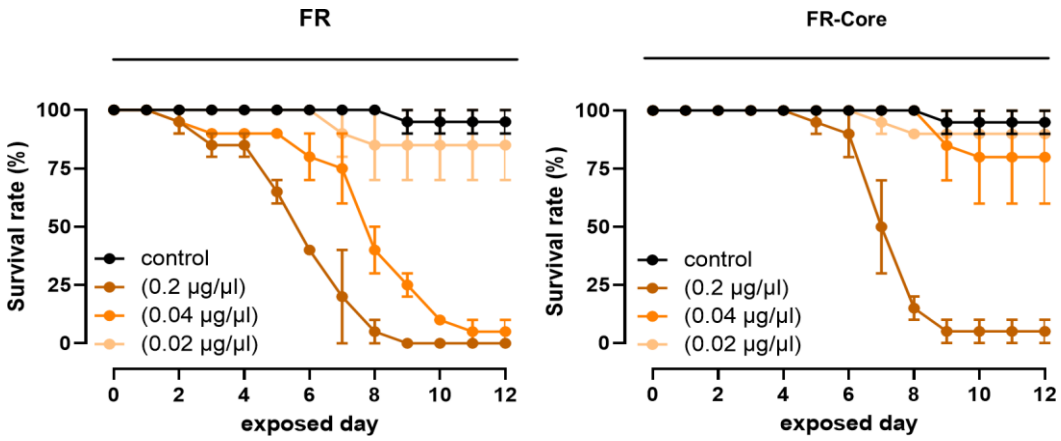


Figure 4.38 Exposure of nymphs of a stink bug (*Riptortus pedestris*) to different concentrations of FR (left) and FR-Core (right), the survival rate over 12 days was measured.[33]

4.6.4 Summary of the bioactivity tests

The new FR analogues were analysed concerning the activity and affinity to the Gαq protein and its effect on insects. The added structure of the side chain in FR compared to FR-Core showed a huge impact on the bioactivity of this natural product which also highlights the importance of FrsA for inhibitor biosynthesis. Without the side chain, the binding affinity to the target Gαq protein was significantly decreased, probably due to missing intermolecular interaction which might be the reason for the loss of activity of FR-Core compared to FR. The DMR assay, as well as the insect feeding assay, show a significantly decreased activity of FR-Core, confirming the improvement of the molecule by addition of the side chain. The comparison of the different acyl residues in the side chain, using the new FR-5 derivative, revealed only C_2 and C_3 as suitable lengths for effective Gαq inhibition. This position in the side chain is slightly variable, but the further side chain is a crucial part of the pharmacophore of FR, introduced 2018 by Reher *et al.*,[11] strongly enhancing the inhibitory activity of FR compared to FR-Core.

4.7 Comparison of FrsA and FrsD

The bioinformatic analyses of the *frs* BGCs in section 4.1 and the detailed investigations of Dr. Isabella Schamari revealed a very high sequence identity of the modules FrsA and FrsD.[132] The amino acid sequence of the CAT domains are 92.3% identical, only varying at the N-terminus of the C domain and the C-terminus of the T domain (see Figure 9.4). The proposed building blocks of the modules are *N*-Pp-Hle for FrsA and *N*-Ac-Hle for FrsD, which was proven for FrsA in section 4.3.2. The next chapter of this work will deal with the comparison of the two NRPS modules in activity and substrate specificity to highlight their similarities and gain insight into possible evolutionary processes.

4.7.1 A domain activity of FrsD

As the A domains of FrsA and FrsD are nearly identical, one would expect similar results for the activation of amino acids by FrsD as in the assay for FrsA in section 4.3.1. We performed the same γ-$^{18}O_4$-ATP exchange assay with FrsA and FrsD in parallel to get comparable results and to see if the slightly differing surrounding domains might influence the activity or specificity of the A domains. The results are shown in Figure 4.39. The absolute substrate conversion rate with the natural substrate L-leucine is roughly the same for FrsD and FrsA. Interestingly, the turnover of the isomers D-leucine and L-isoleucine is significantly decreased in comparison to FrsA. This indicates a broader substrate specificity for $FrsA_A$ than for $FrsD_A$, which can be only due to differences in the surrounding domains, as the amino acid sequence of FrsA and FrsD is, except for one amino acid, identical.

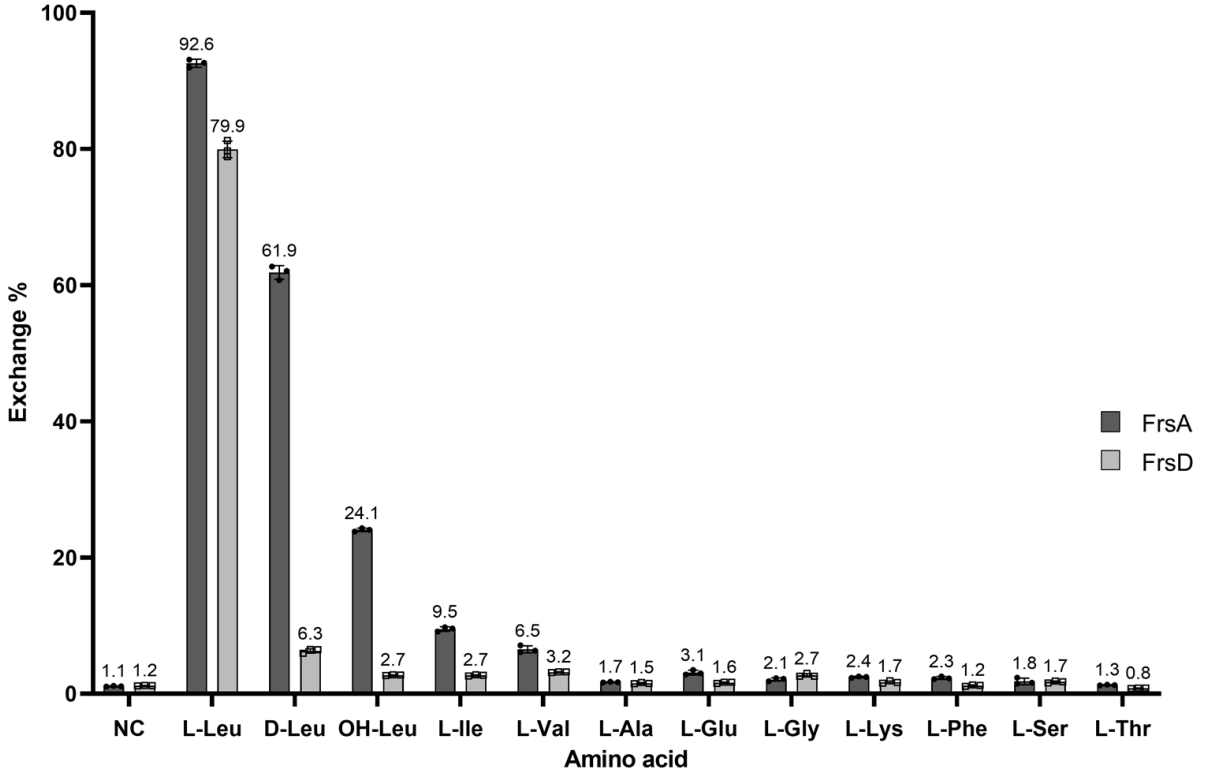


Figure 4.39: γ-$^{18}O_4$-ATP exchange assay results for the A domain of FrsA and FrsD in the whole module construct, both coexpressed with FrsB. NC = negative control without any amino acid. Data are presented as mean values ± SD. All experiments were performed in technical triplicate.

For a more detailed investigation of the enzyme kinetics, a different *in vitro* assay is needed. We aimed to establish the NADH/pyrophosphate (PPi) detection assay reported from Kittilä *et al*., which would allow the determination of the kinetic constants K_M and k_{cat}.[170] However we had difficulties with high background signals and were not able to obtain reproducible results, a problem with this assay that was mentioned in recent literature.[39] So, we decided not to perform further A domain assays, as the C and TE domain Assay in section 4.3.2 and 4.4.3 had proved the activation of L-leucine and in smaller amounts of D-leucine and L-isoleucine in accordance with this assay for FrsA.

4.7.2 C domain assay of FrsD

After the activity of the A domain was confirmed, we performed the C domain assay with FrsD in direct comparison to FrsA. The A domain of FrsD is supposed to activate L-leucine, which is hydroxylated by FrsH. Subsequently, the $C_{starter}$ domain performs the *N*-acylation with acetyl-CoA. In the transesterification assay FrsA was proven to accept not only propionyl-CoA and acetyl-CoA as a substrate, so we tested both substrates for FrsD. The assay was performed as described in section 4.3.2 and 6.11.2 with FrsA or FrsD coexpressed with FrsB and activated FrsH added to the reaction mixture. The results are shown in Figure 4.40. Both enzymes catalyse the production of *N*-Pp-Hle, and the retention time fits the standard in Figure 4.13. The assays supposed to yield *N*-Ac-Hle, also show small peaks in the extracted ion chromatogram of the calculated mass. Since no synthetic standard of this compound is available, we could not prove the identity of the peak, but its retention time is a bit lower than for *N*-Pp-Hle which would be expected for the shorter acyl chain on a C_{18} column. For both assays, the intensity of the peaks for FrsD is lower than for FrsA, but as the intensities are overall very low, this result cannot be taken as a significant statement for the activity.

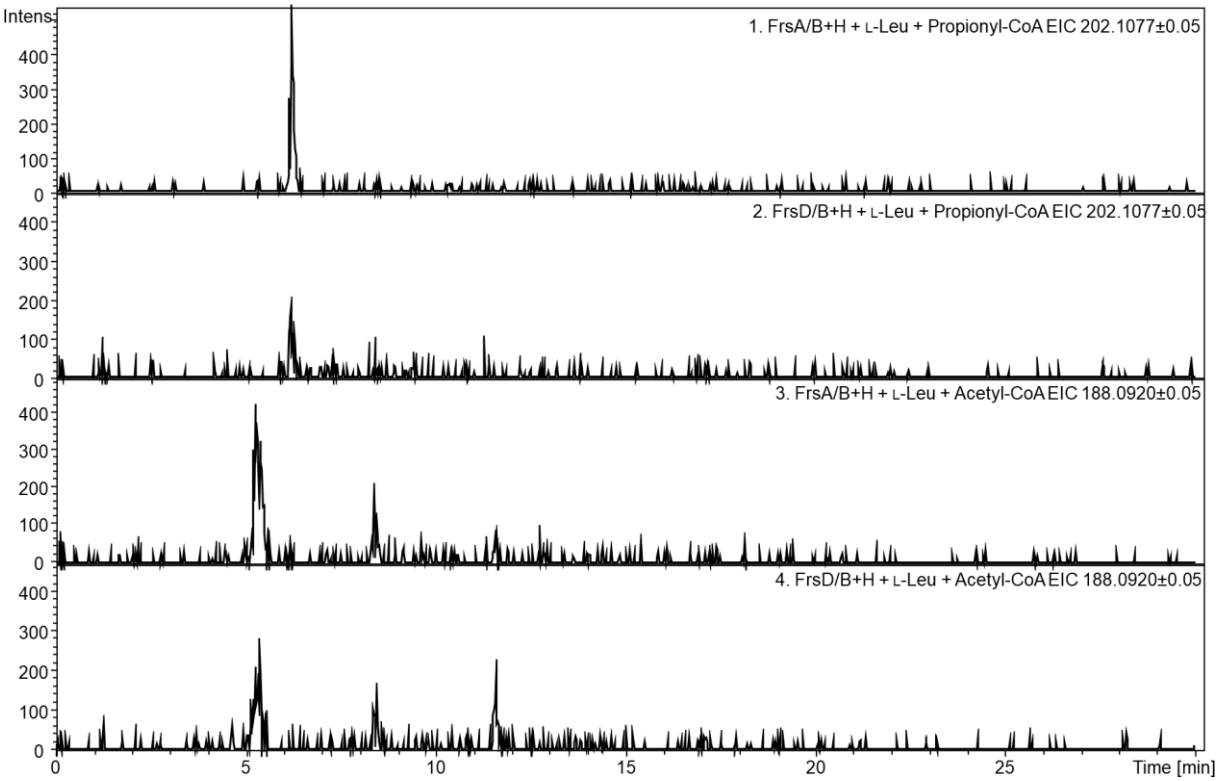


Figure 4.40: Extracted ion chromatograms of *in vitro* production of *N*-Pp-Hle (21) (*m/z* 202.108) and of *N*-Ac-Hle (*m/z* 188.092) from HPLC-MS experiments. Enzymatic assays with purified FrsA or FrsD, FrsB, FrsH incubated with L-Leu and acyl-CoA, hydrolyzed with KOH; **1.** FrsA with propionyl-CoA; **2.** FrsD with propionyl-CoA; **3.** FrsA with acetyl-CoA; **4.** FrsD with acetyl-CoA.

4.7.3 Assay with FrsD and $FrsA_{TE}$

Based on these results, we speculated, if FrsD with the *in vitro* assembled side chain could serve as a substrate for the TE domain of FrsA to catalyse the intermolecular transesterification to FR-Core. In section 4.4.5 and Figure 4.28, it was shown, that the $FrsA_{CAT}$ tridomain, in combination with the

standalone $FrsA_{TE}$, could perform the transesterification of FR-Core to FR. We thus planned the analogous transesterification assay with FrsD and $FrsA_{TE}$ to assemble and transfer the side chain *in vitro*. We performed the assay with propionyl-CoA as well as with acetyl-CoA to investigate the substrate preferences of FrsD. The data is shown in Figure 4.41. Both assays led to the production of FR or FR-2, respectively, but FR was produced in quite small amounts. The results indicate, that $FrsD_C$, has a strong preference for acetyl-CoA over propionyl-CoA as substrate, despite the high sequence identity with FrsA. That could be an explanation for the different ratios of the FR derivatives detected in both natural sources *A. crenata* and *C. vaccinii*. FR-2 is found in extracts of both producers,[11] and *C. vaccinii* even produces it in higher amounts than FR under laboratory conditions (see Figure 9.26). The derivative FR-3, which has a propionyl residue in position (2), incorporated by FrsD, is only present in trace amounts in *A. crenata*[36] and could not be detected in *C. vaccinii* so far (unpublished results of Wiebke Hanke). This relation indicates a higher substrate specificity of $FrsD_C$ compared to $FrsA_C$, as FR-3 is produced when $FrsD_C$ incorporates propionyl-CoA instead of acetyl-CoA. However, as FR-3 was also isolated from a didemnid ascidian, here termed sameuramide A,[19] it would be very interesting to compare the sameuramide BGC with the *frs* BGC. Here it may be assumed that the propionyl residue must be preferred in both starter C domains, as FR was not reported to be found in this organism. Unfortunately, no genetic information of the producer has yet been published. The same is true for the well-investigated YM, which has the acetyl residue in position (1), which is one of two changed positions to the FR structure. YM is produced by *Chromobacterium* sp. QS3666, but up to date, investigations on its biosynthesis were not published.[27] Its equivalent to FrsA would be responsible for the side chain biosynthesis and could give insights to the different substrate preference.

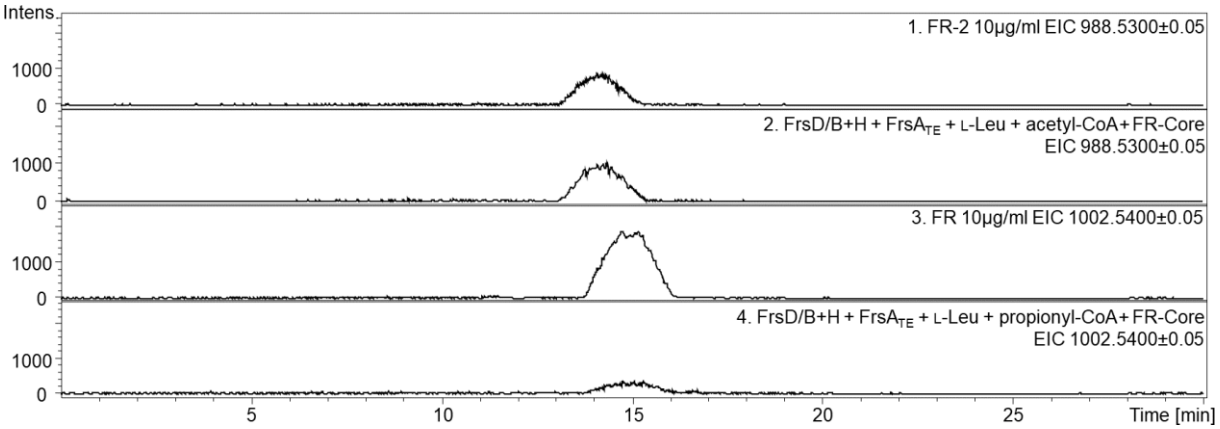


Figure 4.41: *In vitro* side chain assembly and transfer assays with FrsD/B, FrsH and standalone $FrsA_{TE}$. Extracted ion chromatograms of FR-2 (*m/z* 988.53) and FR (*m/z* 1002.54) from HPLC-MS experiments; **1.** FR-2 standard (10 µg/ml); **2.** Purified FrsD/B, FrsH and $FrsA_{TE}$ incubated with acetyl-CoA, L-Leu and FR-Core; **3.** FR standard (10 µg/ml); **4.** Purified FrsD/B, FrsH and $FrsA_{TE}$ incubated with propionyl-CoA, L-Leu and FR-Core.

4.7.4 Investigations on the evolution of FrsA

In our recent publication,[33] we discuss an evolutionary theory for the formation of the *frs* BGC: The evolution of natural product BGCs is characterized by both, evolution of enzyme promiscuities and genetic events such as horizontal gene transfer, duplications, recombination and gene expansions.[171,172]

In both BGCs, the regions encoding the Hle activating A domains of *frsA*, *frsD* and *frsG* are identical in addition to large parts of $frsA_C$ and $frsD_C$ (see section 4.1). The close evolutionary relationship of all $C_{starter}$ domains is further supported by a detailed phylogenetic analysis (Figure 4.3). It is thus conceivable that evolution of FR side chain biochemistry involves at least one duplication event, leading to biosynthesis of another acylated Hle moiety. Thioesterase phylogenetic analysis suggests close evolutionary relationship of the transesterifying $FrsA_{TE}$ and the macrocylizing $FrsG_{TE}$, all together pointing at the evolution of all parts of *frsA* from ancestors within the BGC. It has been suggested, that some specialized microbial natural products may have evolved from ancestral metabolites, that had once been the end product of a biosynthetic pathway.[53,54] These ancestor molecules may have served as templates for structural variations based on evolutionary processes.[54] Bioinformatic studies, such as the reconstruction of the evolutionary history of the large natural product families of glycopeptides[173] and type II polyketides,[174] support this theory.

In our example, the side chain attachment boosts bioactivity of the depsipeptide scaffold, as shown by comparative activity and affinity tests of FR and FR-Core on Gαq proteins and insect toxicity assessments. This improvement is supported by the accompanied docking studies revealing that the extension of chemical space generates additional interactions with the target, thereby highlighting the enormous importance of the side chain for effective Gαq inhibition by FR.

Based on our analyses, it is tempting to speculate that the intermediate macrocycle FR-Core is the product of the ancestral BGC *frsB-H* that was extended by duplication, gene expansion and further domain evolution resulting in biosynthesis and incorporation of the acylated Hle side chain, thereby encoding a metabolite with improved Gαq inhibition potency (see Figure 4.42). This hypothetical model would be consistent with the recently published dynamic chemical matrix evolution (DCME) hypothesis published by Chevette *et al*. for natural products evolution: “A graphical, virtual representation of how the chemical matrix changes over time in response to evolutionary forces, integrating structure, biomolecular activity, and relative fitness. It incorporates negative selection, in addition to positive selection and neutral evolution, as an often-overlooked evolutionary force for exploration of the chemical diversity throughout evolutionary dynamics of BGCs and their biosynthetic sub-clusters”.[53] Evolutionary relationships of modular pathways have e.g. been analyzed for pyrrolamides in *Streptomyces*,[175] actin-binding macrolides from various organisms,[176] and lipopeptides from *Pseudomonas*.[177] This model could be also fitting for FR, considering improved Gαq inhibition properties of the resulting metabolite as fitness advantage for the producer and trait for positive selection of the altered BGC, which however has yet to be investigated experimentally. Other scenarios for *frsA* evolution, such as the reverse duplication of a previously existing *frsA* to yield *frsD*, its *de novo* generation or its horizontal acquisition from another organism have thus at this stage also be taken into account and cannot be excluded with certainty.

The global BGC analysis in Hermes et al. revealed not closely related NRPS system or fragments thereof in the sequence databases, coining the two *frs* BGCs a rare, small and so far uniform gene cluster family (GCF) that was likely shaped by intra-BGC evolution. Sequencing and comparative analyses of more Gαq inhibitor depsipeptide BGCs, such as those of YM and sameuramide, that await discovery, is expected to reveal more details on the evolution of this fascinating family of natural products.[33]

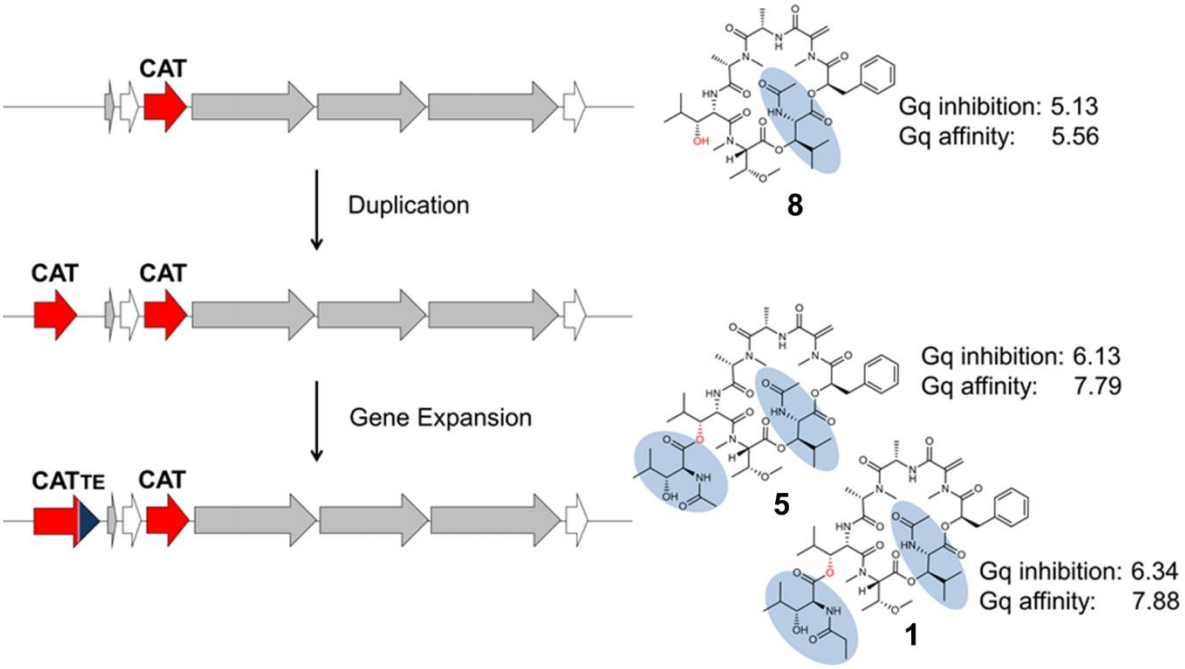


Figure 4.42: Hypothetical model for evolution of FR side chain biosynthesis based on the assumption that *frsA*$_{CAT}$ was generated by a duplication event of the highly similar *frsD* (>94% identity). For each compound, pIC50 values are indicated. The exact order and timing of the steps cannot be determined.[33]

4.8 Crystallisation experiments

The three-dimensional structure of a protein provides crucial insights into its active site and spatial arrangements which are relevant to understand the function of the enzyme. Amongst other methods, X-ray crystallography is one of the most powerful, well-established and accurate technologies for protein structure elucidation.[178] As discussed before, TE domains are diverse in sequence and function and not many crystal structures thereof have been published to date. Some examples for PKS TE structures are PksA TE (PDB ID code 3D4H) from aflatoxin biosynthesis[179] and DEBS TE (PDB ID code 1KEZ) from 6-deoxyerythronolide B biosynthesis.[180] From NRPS pathways, FenTE (PDB ID code 2CB9) of the fengycin biosynthesis[181], SrfC TE (PDB ID code 2VSQ) of the surfactin biosynthesis,[182] and NocTE (PDB ID code 6OJC) from nocardicin biosynthesis[139] have been crystallised. So far, no structure of an intermolecular transesterifying TE domain is reported, which is why we aimed to achieve this goal. We, therefore, cooperated with the workgroup of Prof. Gebhard Schertler from the Paul Scherrer Institute (PSI) in Switzerland, which has an excellent protein crystallisation facility. The PSI is located at the Swiss Light Source (SLS), which is a state-of-the-art Synchrotron, hosting several beamlines specialized for Protein X-ray crystallography. Jonas Mühle planned and performed all crystallisation trials.

Before starting crystallisation trials, the protein of interest needed to be isolated in high concentration and purity and, if procurable, without any unnatural additions like histidine tags that might have a negative influence on diverse properties of the protein.[183–185] To achieve this, a new expression construct of $FrsA_{TE}$ with a protease recognition site between the protein and the histidine tag, that can be used to cleave the histidine tag after the first purification step, was designed. An additional purification step utilising a size-exclusion column (SEC) for the fast protein liquid chromatography system (FPLC) followed before sending the protein to our collaborators.

Additionally, we analogously prepared the monomodular NRPS FrsA for cryo-electron microscopy (cryo-EM) experiments, an alternative method to X-ray crystallography for protein structure elucidation.[186] This approach was also performed in cooperation with Prof. Gebhart Schertler and Jonas Mühle from the PSI and with the Scientific Center for Optical and Electron Microscopy (ScopeM) of the ETH Zürich.

4.8.1 Purification of $FrsA_{TE}$

A new expression construct was created by cloning the TE domain of FrsA into the pHis8-TEV plasmid. This plasmid was constructed by Dr. René Richarz and encodes for the specific cleaving site of the Tobacco Etch Virus protease (pTEV) directly behind the N-terminal His8 tag. The cloning procedure was performed analogously to the one described in section 4.2 and 6.5.7. In this case, the restriction sites *Bam*HI and *Hind*III were used and the genomic DNA of *C. vaccinii* was the PCR template. The resulting plasmid was isolated and after verification of the nucleotide sequence *via* Sanger sequencing, it was transformed into the expression host *E. coli* BL21 (DE3).

Overexpression of $FrsA_{TE}$, induced with IPTG in TB medium, and subsequent purification via Ni-NTA affinity chromatography, yielded high amounts of protein, approximately 100 mg from 1 l culture, which was sufficient for protease digestion (see section 6.6). The protease pTEV (plasmid kindly provided by Dr. René Richarz) was expressed and purified in the same manner with an His6 tag. Both proteins were rebuffered in a suitable system for proteolysis, mixed and incubated overnight at 4 °C. Afterwards, all remaining proteins with a His tag were eliminated via inverse Ni-NTA chromatography (see section 6.6.5). To obtain highly pure protein, we subsequently performed FPLC purification via an SEC column (see section 6.6.6). The chromatogram in Figure 4.43 shows only slight impurities which could be well separated from the protein of interest.

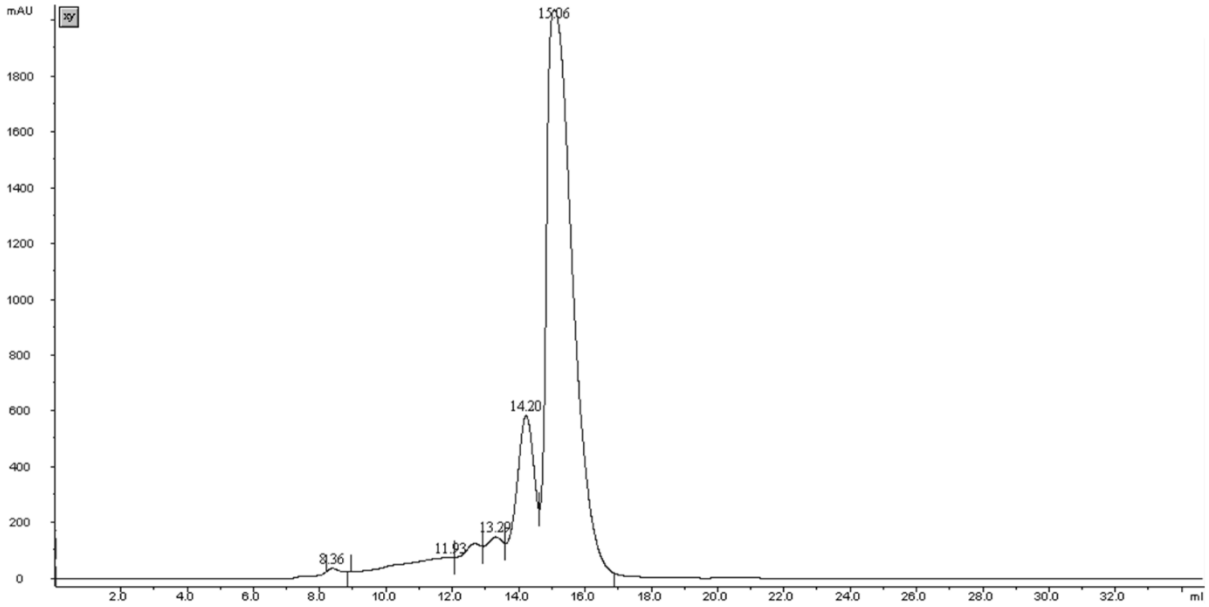


Figure 4.43: FPLC SEC chromatogram of $FrsA_{TE}$ after pTEV digestion. The TE domain eluted at retention time 15.06.

In Figure 4.44, the FPLC chromatogram of the pure TE domain is shown, as well as the SDS-PAGE gel of the TE: 1. before the proteolysis, 2. after the proteolysis and 3. after the FPLC purification. There is a small peak in front of the main $FrsA_{TE}$ signal visible, but no additional band in lane 3. of the SDS-PAGE gel. Extensive trials to try to remove this signal by repeated SEC and analysis of the stored purified protein revealed this peak to be impossible to remove and to increase over time. This led to the conclusion, that this peak probably relates to a different conformation of the TE domain that forms an equilibrium with the main conformation and has a different elution behaviour. We thus stopped trying to remove this "impurity" from the sample, as in the crystallisation trials the protein is investigated over a long period where it has enough time to establish equilibrium. It is not predictable, which conformer is more likely to crystallise.

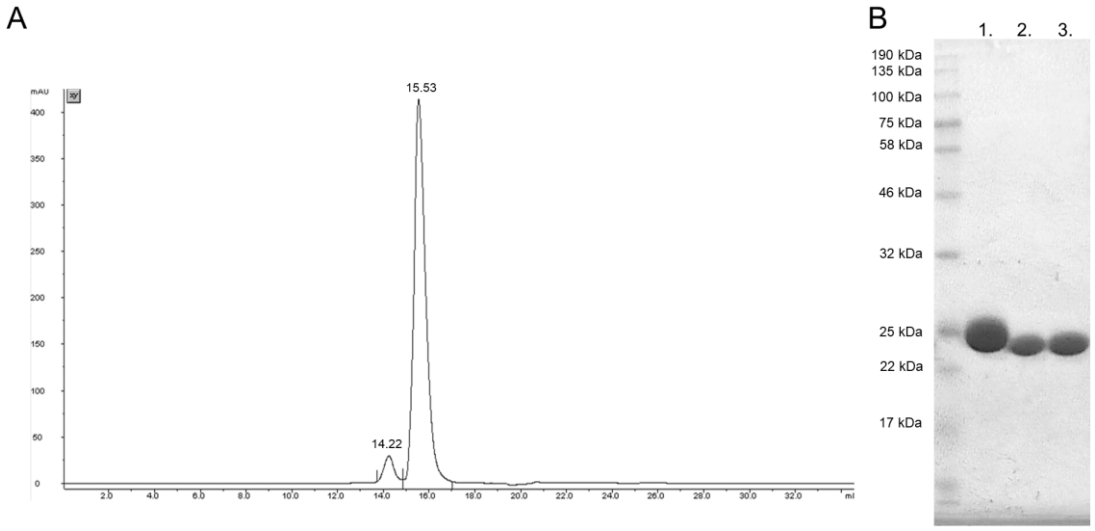


Figure 4.44: Purification of $FrsA_{TE}$. A. FPLC SEC chromatogram of $FrsA_{TE}$ second SEC step. **B.** SDS-PAGE of $FrsA_{TE}$: 1. before the proteolysis (calculated molecular weight 30.42 kDa), 2. after the proteolysis (28.15 kDa) and 3. after the FPLC purification (28.15 kDa).

For crystallisation trials, the purified protein samples were concentrated to 20-25 mg/ml and sent to Switzerland via express delivery. The first protein sample of $FrsA_{TE}$ was in a buffer of 20 mM Tris, 100 mM NaCl, 0.5 mM TCEP, 10% glycerol, flash-frozen with liquid nitrogen and transported on dry ice.

4.8.2 Optimization of storage conditions

The crystallisation trials with the first protein batch did not result in the formation of any crystals after the first weeks. We thus decided to optimize the buffer conditions before sending the next sample because the buffer of the initial solution is known to influence the whole crystallisation process.[187] To test the stability of the protein in different chemical environments, we performed a thermal shift assay (TSA) in cooperation with Tobias Claff from the workgroup of Prof. Christa Müller in the Institute of Pharmaceutical Chemistry of the University of Bonn. This assay uses the fluorescent dye (SYPRO orange) which shows a low fluorescent signal in aqueous, polar environment, but high fluorescence in a nonpolar environment. When the protein unfolds during heating, the hydrophobic core is exposed and the fluorescent signal intensity of the dye increases. During heating, the fluorescence is monitored and the temperature, where half of the protein population is unfolded, is defined as the melting point of the protein (T_M).[188]

We tested several buffers systems with different buffer components, pH values, salt concentrations and additives which are listed in the material section 6.7.3 in Table 6.20. We especially tested the influence of the substrates **22** (buffer 14) and FR-Core (buffer 15) which might have a stabilising effect on protein conformation.[189] Both were measured once freshly mixed and once with a preincubation time of two hours. The collected data was normalised and fitted with a nonlinear sigmoidal fit to determine the inflexion point as melting temperature, see Figure 4.45.

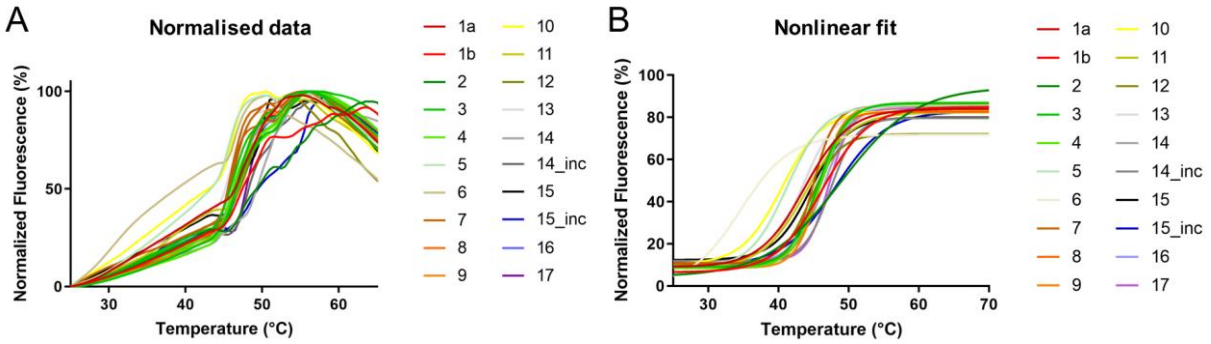


Figure 4.45: Graphs of the thermal shift assay. A. Normalised data of the buffers 1a to 17 (see Table 6.20) **B.** Nonlinear fit data of the buffers 1a to 17.

Unfortunately, no strong influence of the buffer composition on the melting point was identified (see Figure 4.46 A). The pH value of 7.5 appeared to be the optimum and none of the additives increased the T_M significantly. But as buffer 2 with a minimal amount of 20 mM HEPES and 50 mM NaCl gave a slightly higher melting point than Tris buffer (buffer 1a), we subsequently performed a batch of

experiments with buffer systems based on buffer 2 with different precipitants and saccharides added, that are commonly used for crystallisation.[189] The T_M values are shown in Figure 4.46 B. It was observed that only the addition of high amounts of trehalose led to a significant increase of the melting point.

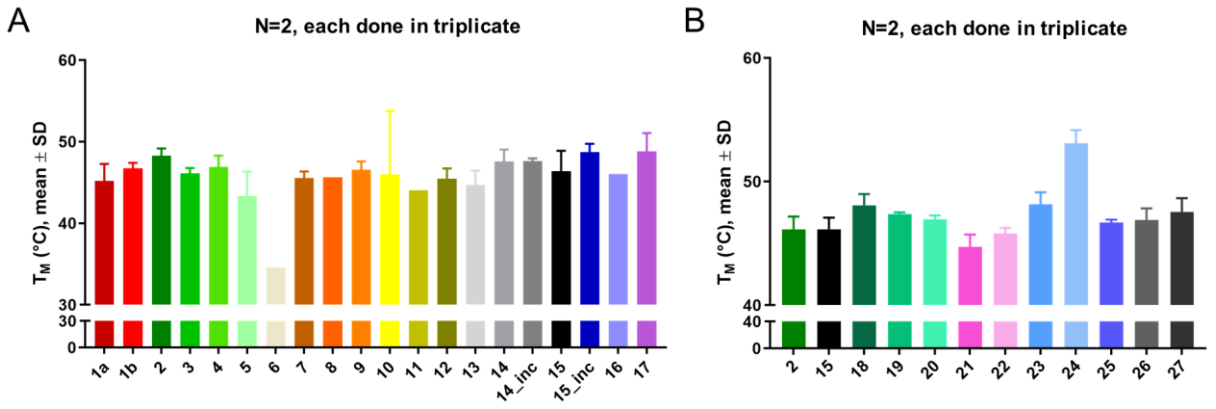


Figure 4.46: Calculated melting temperatures. A. T_M for buffers 1a to 17. **B.** T_M for buffers 2, 15 and 18 to 27 (see Table 6.20).

Despite the effect of the added trehalose, we decided to stick with a minimal buffer and chose to use buffer 2 for all following crystallisation trials. Additionally, we decided to send the freshly purified protein on ice to avoid unnecessary freeze/thaw cycles, as the first trials revealed long stability of the protein in solution at 4 °C.

4.8.3 First crystallisation trials

The crystallisation trials were performed by Jonas Mühle using the sitting drop vapour diffusion method. Here, the protein solution is mixed 1:1 with a precipitant solution in a drop above a second reservoir with the pure precipitant solution. The chamber is sealed, and the water diffuses out of the drop and into the reservoir, or the other way round until the osmolarity of the drop and the reservoir are equal. This leads to a slow change in concentration of both the protein and the precipitant in the drop until equilibrium is reached, which ideally is in the nucleation zone of the phase diagram and leads to crystal growth (see Figure 4.47 and section 6.8).[178]

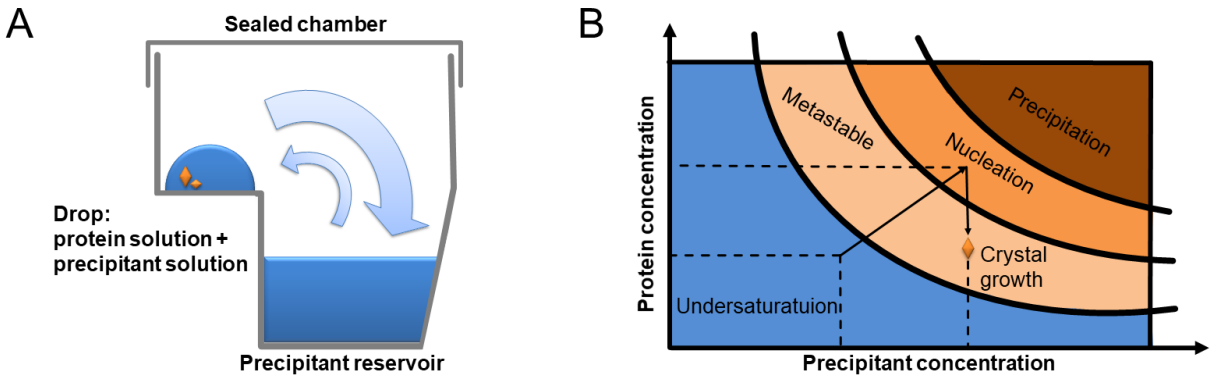


Figure 4.47: The principle of protein crystallisation. A. The sitting drop vapour diffusion. **B.** Phase diagram (Adapted from Dessau *et al.*).[178]

The first crystallisation trial yielded no crystals but could confirm the high long-term stability of the protein at 4 °C and 20 °C. After optimization of the buffer and transport conditions described in 4.8.2, the second trial led to the formation of small crystals. In addition to the batch of *apo* $FrsA_{TE}$, another batch was supplemented with a 2-fold molar excess of FR-Core, to examine if the substrate has (confirmational) stabilizing abilities that might facilitate crystal growth. For each of these batches, the conditions with the most promising results are shown in Figure 4.48. Tryptophan fluorescence confirms that the pictured crystals are protein and not salt crystals. The thin needles were however not big enough for X-ray diffraction analysis, but a starting point for further optimization trials.

After achieving primary hits for crystallisation, the conditions were optimized by different systematic approaches. The "additive screening" approach uses the primary hit conditions and adds commercially available additives, that might improve the crystallisation further. The "grid screening" approach systematically varies two parameters e.g. salt or polyethyleneglycol (PEG) concentration, in a 96-well plate with the primary hit conditions in the centre, to fine-tune the initial conditions.[190]

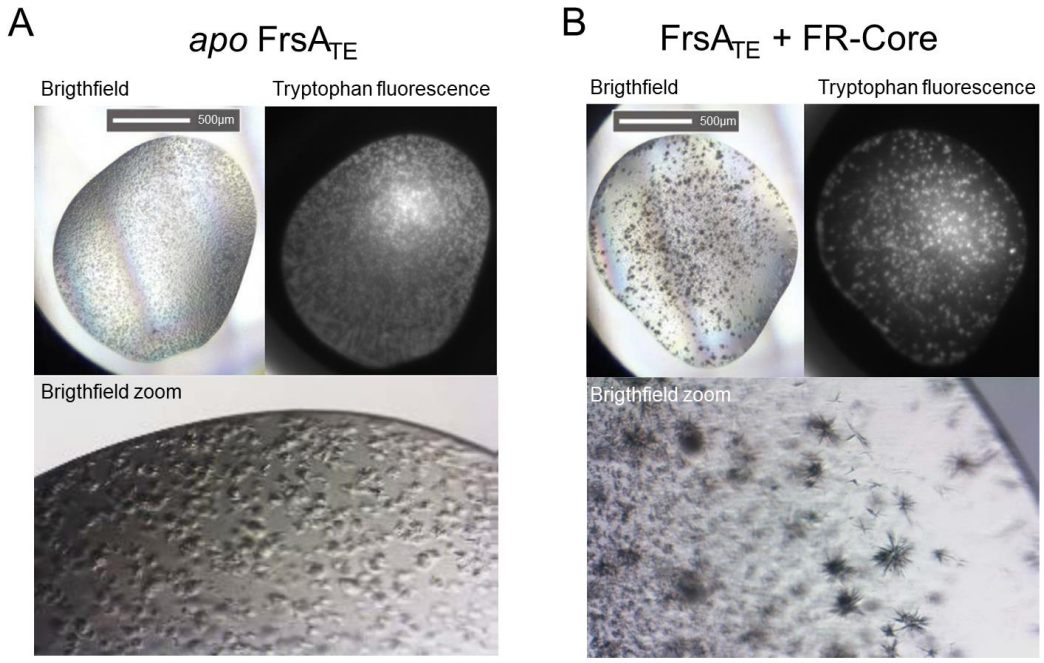


Figure 4.48: Hits of the second crystallisation trial. A. The *apo* $FrsA_{TE}$ protein in precipitant buffer containing 0.1 M Bis-Tris Propane, 0.2 M potassium thiocyanate and 20% PEG 3350 at pH 7.5 and 20 °C. **B.** The $FrsA_{TE}$ protein with 2 molar excess of FR-Core in precipitant buffer containing 0.1 M bis-tris propane, 0.2 M sodium iodide and 20% PEG 3350 at pH 6.5 and 20 °C. Pictures recorded by Jonas Mühle.

The additive screening approach gave the so far best crystals for the *apo* protein: Small rod-shaped $FrsA_{TE}$ *apo* crystals grew in a crystallisation buffer consisting of 0.1 M bis-tris propane pH 7.5, 0.2 M potassium thiocyanate, 4% tert-butanol and 20% PEG 3350 within 3-7 days (see Figure 4.49 A). For the co-crystallisation with FR-Core, the grid screening led to improved primary hit conditions: Thin needles or needle clusters of $FrsA_{TE}$ + FR-Core grew within 28-42 days in a crystallisation mother liquor consisting of 0.1 M bis-tris propane pH 6.5, 0.1-0.2 M sodium iodide and 14% PEG 3350 (see Figure 4.49 B).

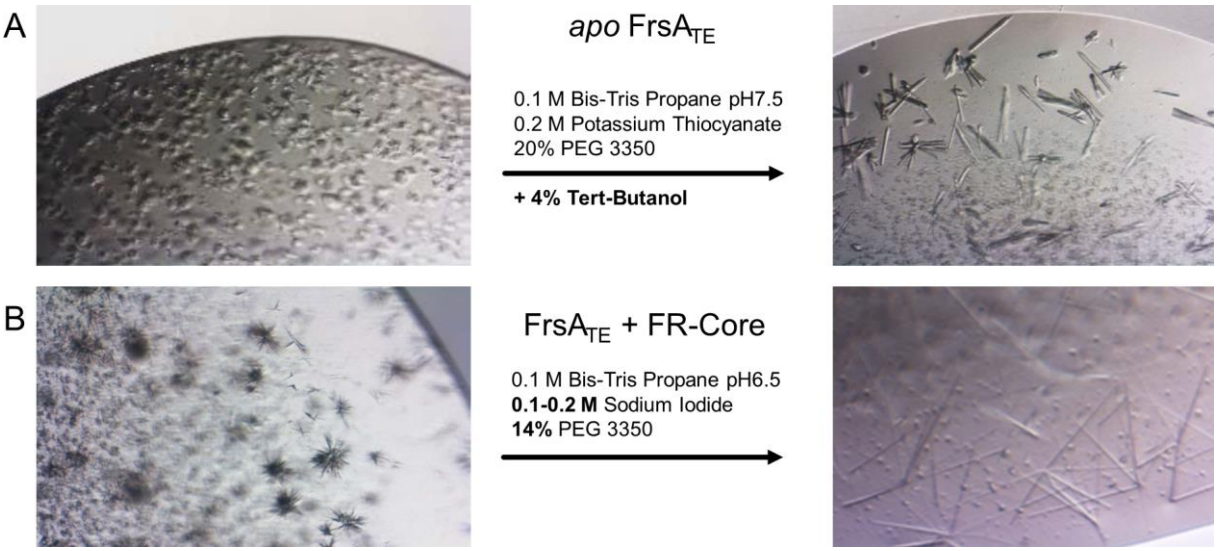


Figure 4.49: Optimizied crystallisation conditions for FrsA$_{TE}$. A. *apo* protein in 0.1 M Bis-tris propane pH 7.5, 0.2 M potassium thiocyanate, 20% PEG 3350 **a**nd 4% tert-butanol as additive. **B.** FrsA$_{TE}$ with FR-Core in 0.1 M Bis-tris propane pH 6.5, 0.1 – 0.2 M sodium iodide and 14% PEG 3350. Pictures recorded by Jonas Mühle.

Although there is a visible improvement in crystallisation, these crystals were still not big enough for X-ray diffraction analysis. Also, needles are not generally ideal for X-ray measurements, as they are normally thin and grow as multi-crystals or inseparable clusters, making it hard to get reasonable 3-dimensional data.[190] For the second optimization trial, the random microseed matrix screening (rMMS) was used, where existing crystals are crushed to small seeds that are added to a new screening.[191–193] This method can also improve the shape of the crystals leading to more suitable samples. The principle of this method is shown in Figure 4.50.

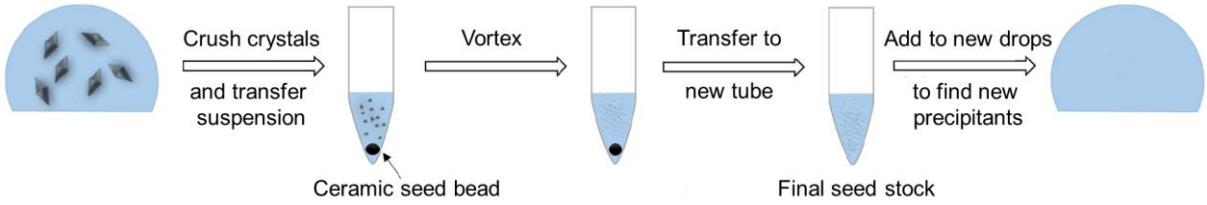


Figure 4.50: Principle of random microseed matrix screening (rMMS). Scheme by Jonas Mühle.

From this batch, the first crystals were prepared under cryo conditions for X-ray diffraction analysis. The crystals were still quite small, but there were some reflexes visible, one at 9.2 Å which is promising for further experiments. The crystal size and the cryo conditions need to be optimized further to obtain a higher resolution. Next, the use of a terbium complex, named crystallophore, that has proven to be a powerful auxiliary for protein crystallography, is attempted.[194] This complex acts as a strong nucleation agent that is expected to lead to bigger derived crystals that are more suitable for X-ray diffraction experiments.

4.8.4 Preparation of FrsA for cryo-EM trials

While X-ray crystallography is still the most established method for generation of 3D structures of proteins, new methods have been developed that cope with some of the drawbacks from classical crystallography. To obtain valuable crystals, high amounts of protein are needed with increasing difficulties for increased protein size. The method of cryo-electron microscopy (EM) for the high-resolution structure determination of biomolecules in solution was awarded the chemistry Nobel Prize in 2017.[195] Cryo-EM has the advantage to preserve the protein in a near-native environment by rapidly freezing it in liquid nitrogen-cooled ethane to get vitreous ice. As the atoms in proteins are mainly light elements that diffract electrons weakly, the contrast is very poor which is coped with multiple copies of the specimen that are computationally averaged.[186] This advance has led to a constantly rising number of near-atomic resolutions of 3D structures reported in the EM data bank (EMDB).[196] In 2014 the first full-length PKS module PikAIII from the bacterium *Streptomyces venezuelae* was investigated by cryo-EM. The 328 kDa module yielded a sub-nanometre-resolution suited for three-dimensional reconstructions, which was a pioneering success.[197] The standalone FrsA TE domain is too small for this method, but the monomodular NRPS FrsA, that contains $FrsA_{TE}$, is in size for cryo-EM experiments. There have only a few structures of whole NRPS modules been solved so far, with only one cryo-EM structure reported with low resolution.[182,198–200] FrsA would be favourably analysed in complex with the essential MLP FrsB, analogously to the structure of EntF with YbdZ or PA2412.[199] A full NRPS module structure could additionally give insight into a lot of further structural aspects.

The purified protein complex FrsA/B was prepared analogously to $FrsA_{TE}$ through expression with a cleavable His8 tag and digestion with pTEV after Ni-NTA purification (see section 4.8.1). Final purification was performed with an SEC column by FPLC. The chromatogram and SDS-PAGE gel of purified FrsA/B are pictured in Figure 4.51. The protein sample was sent to the PSI frozen on dry ice and subsequently, cryo-EM grids were prepared by Jonas Mühle and used for first data collection trials. However, the particles turned out to be rather small; two domains were visible but no secondary structure elements. The particle size thus needs to be optimized to improve three-dimensional structure reconstruction. Therefore, helper chaperons like designed ankyrin repeat proteins (DARPins) could be fused with the protein, that enhances the construct size and adds an asymmetric feature to optimise the shape of the particles.[201]

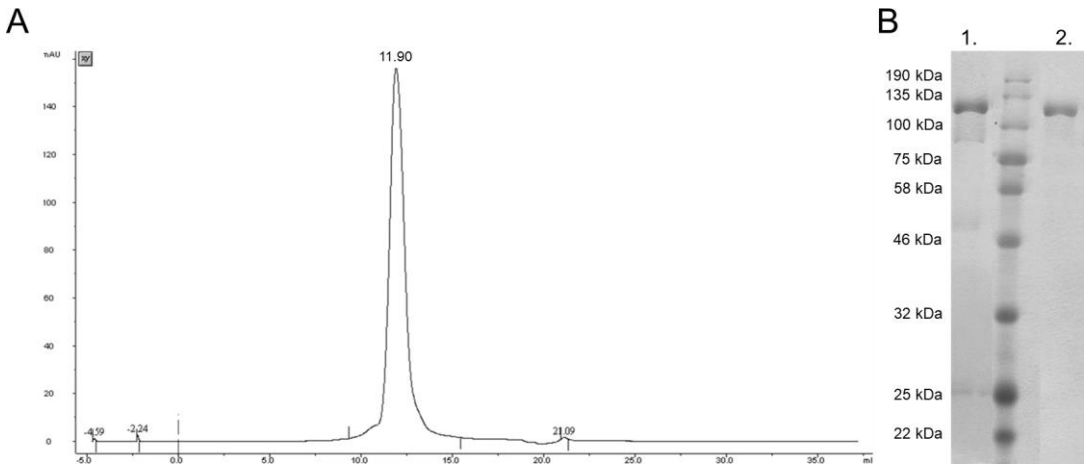


Figure 4.51: Purification of FrsA/B (139 kDa/8 kDa). A. FPLC SEC chromatogram of FrsA/B after SEC purification. **B.** SDS-PAGE of FrsA/B (only FrsA visible): 1. after the proteolysis and 2. after the FPLC purification.

5 Summary

The detailed investigation of secondary metabolites gives insight into many fields of research, like the ecology and interaction of plants, animals and bacteria, but also the effect of these compounds on animals or humans and thus the possibility to put such agents to pharmacological use.

The depsipeptide natural product FR900359 (FR, Figure I) has gained much attention as valuable pharmacological tool and promising therapeutic agent during the last years, because of its unique mechanism of action, the strong and selective inhibition of Gαq proteins. Up to now, two bacterial producers of FR had been described: The endosymbiotic "*Ca*. Burkholderia crenata" living in the leave nodules of the plant *Ardisia crenata* and the recently discovered soil bacterium *Chromobacterium vaccinii*, both harbouring architecturally identical *frs* biosynthetic gene clusters (BGCs). This work focused on the detailed investigation of FR side chain biosynthesis. The side chain assembly, its attachment to the cyclic intermediate FR-Core catalysed by the monomodular non-ribosomal peptide synthetase (NRPS) FrsA and the effect of the side chain on FR bioactivity was examined *in vitro*, *in vivo* and *in silico*. The characterisation of the FrsA thioesterase (TE) domain was the main focus of this investigation as it was shown to catalyse transesterification of the side chain, an activity rarely found in other NRPS systems.

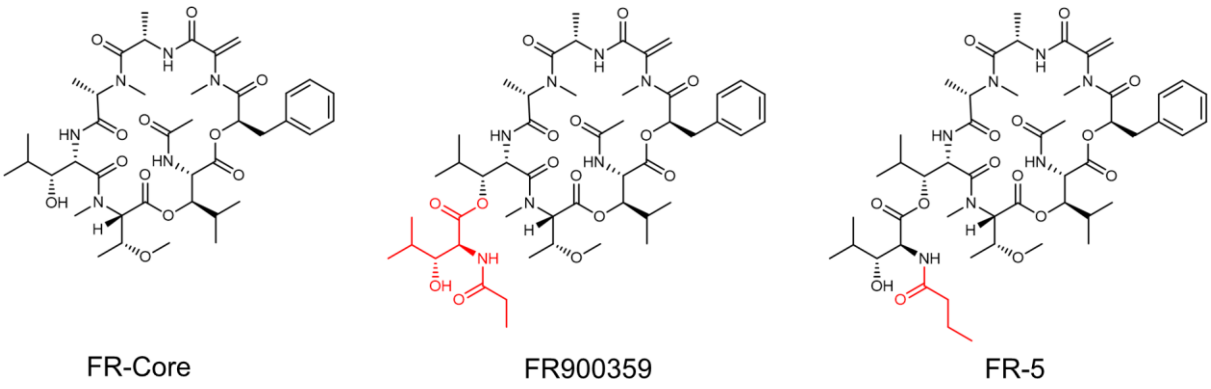


Figure I: Structures of FR900359 (FR), FR-Core and the unnatural derivative FR-5. The side chain of FR and the butyryl residue of FR-5 are depicted in red.

Frist, the two *frs* BGCs were compared and detailed bioinformatic analyses were performed. These revealed a high sequence identity of $FrsA_{CAT}$ to FrsD. Phylogenetic trees of NRPS starter condensation (C) and TE domains were generated to detect closely related domains in the database. The high identity and the close phylogenetic distance between the $FrsA_C$ and $FrsD_C$ domains together with the identical adenylation (A) domains point to one or more duplication events during evolution of *frs*. The noncanonical TE domain of FrsA showed no closer phylogenetic relative than $FrsG_{TE}$, which is only 41.7% identical. Its phylogenetic origin could not be verified with certainty but also suggests an intra-BGC origin.

We then cloned and heterologously expressed multiple *frs* genes and domains in *E. coli* for *in vitro* experiments to obtain information on the catalytic activities and substrate specificities. The A domain

of FrsA was confirmed to activate the expected substrate L-leucine but also activated D-leucine and L-isoleucine. The MbtH-like protein FrsB was crucial for the successful expression and activity of the A domain. Using the tridomain $FrsA_{CAT}$ coexpressed with FrsB and the non-heme diiron monooxygenase FrsH, the *in vitro* assembly of the side chain *N*-propionylhydroxyleucine (marked red in Figure I) was achieved. This proved the β-hydroxylation of the T domain-bound L-leucine by FrsH and the transfer of propionyl-CoA by $FrsA_C$ onto the leucinyl amino group.

Next, the transfer of the side chain onto the intermediate FR-Core was investigated *in vitro*. *N*-propionylhydroxyleucine-SNAC was chemically synthesized as substrate mimic for the TE domain. The other substrate FR-Core needed to be isolated. Therefore, the deletion mutant *C. vaccinii ΔfrsA/vioA* was generated. Without FrsA to generate and attach the side chain, the FR biosynthesis was interrupted at the stage of the cyclic intermediate FR-Core, which was isolated and characterized in this work (Figure I).

In the next step, we were able to prove the transesterifying function of $FrsA_{TE}$ with *N*-propionylhydroxyleucine-SNAC and also with the *in vitro* assembled side chain as substrate (Figure II).

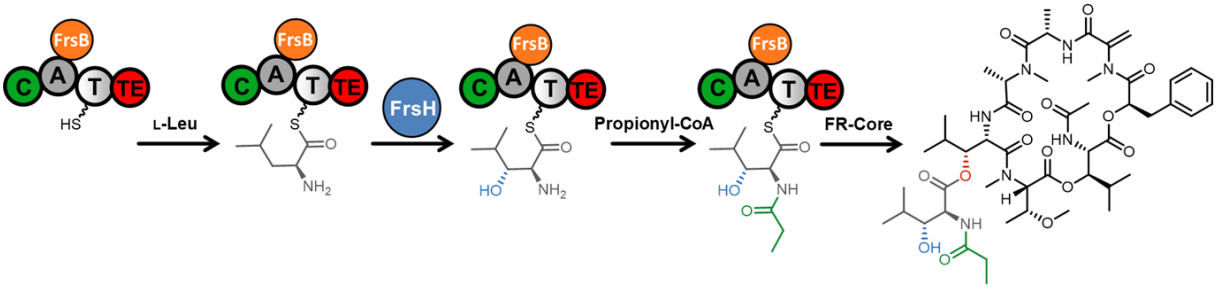


Figure II: *In vitro* assays with FrsA/B and FrsH. Reaction scheme of *N*-Pp-Hle formation and intermolecular transesterification onto FR-Core to yield FR.

The TE domain catalysed the transfer of the side chain onto the free hydroxyl group of the peptide core, yielding the complete FR molecule. $FrsA_A$ and $FrsA_C$ showed in previous assays some promiscuity towards their substrates. We added different amino acids and CoA-bound acyl residues as substrates for the transesterification assay. The modified side chains were accepted as substrate by the TE domain and yielded new FR derivatives. The assays with the altered acyl-CoAs resulted in the formation of the natural derivative FR-2 and the new derivative FR-5 with a butyryl residue in the side chain. The TE domain was able to transfer these variants of the side chain onto FR-Core, so its substrate recognition showed some flexibility as well. The different amino acids revealed the stereochemistry to be relevant for the substrate to be transferred by the TE domain. When D-leucine was supplemented for the *in vitro* side chain assembly, no FR was produced. In contrast, L-isoleucine was accepted, yielding an FR stereoisomer. While we did not investigate this new isomer further, its analysis would be interesting for the selectivity of the FrsH activity.

To explore the influence of the acceptor molecule in the transesterification assay, FR-core was replaced with different substrates. L-hydroxyleucine was used as a minimal substrate and the cyclic depsipeptide lysobactin with an L-hydroxyleucine moiety in its core structure as a different cyclic substrate. Unfortunately, both assays did not show the formation of the calculated results during LC-MS analysis, indicating that hydroxyleucine alone is not sufficient for substrate recognition of the TE domain. Furthermore, we investigated if $FrsG_{TE}$ would be able to catalyse the same reaction as $FrsA_{TE}$, but despite their proximity in the phylogenetic tree, their function proved to be different and $FrsG_{TE}$ did not catalyse the production of FR in the transesterifying assay.

After the successful *in vitro* production of an FR derivative with a butyryl group in the side chain (FR-5, Figure I) we attempted *in vivo* biosynthesis by feeding butyric acid to *C. vaccinii* in a minimal medium. This resulted in sufficient yields of FR-5 for isolation, structure elucidation, and measurement of the biological activity. In-depth NMR measurements confirmed the proposed structure of FR-5,

We investigated the bioactivity of the intermediate FR-Core. FR-Core showed a 16-fold reduced activity against Gαq and a 207-fold reduced binding affinity (Figure III A and B). This demonstrates that the side chain strongly contributes to the interaction between FR and the Gαq protein, which was visualized in *in silico* docking experiments, revealing additional hydrogen bonds and also intramolecular interactions within FR caused by the side chain (Figure III D).

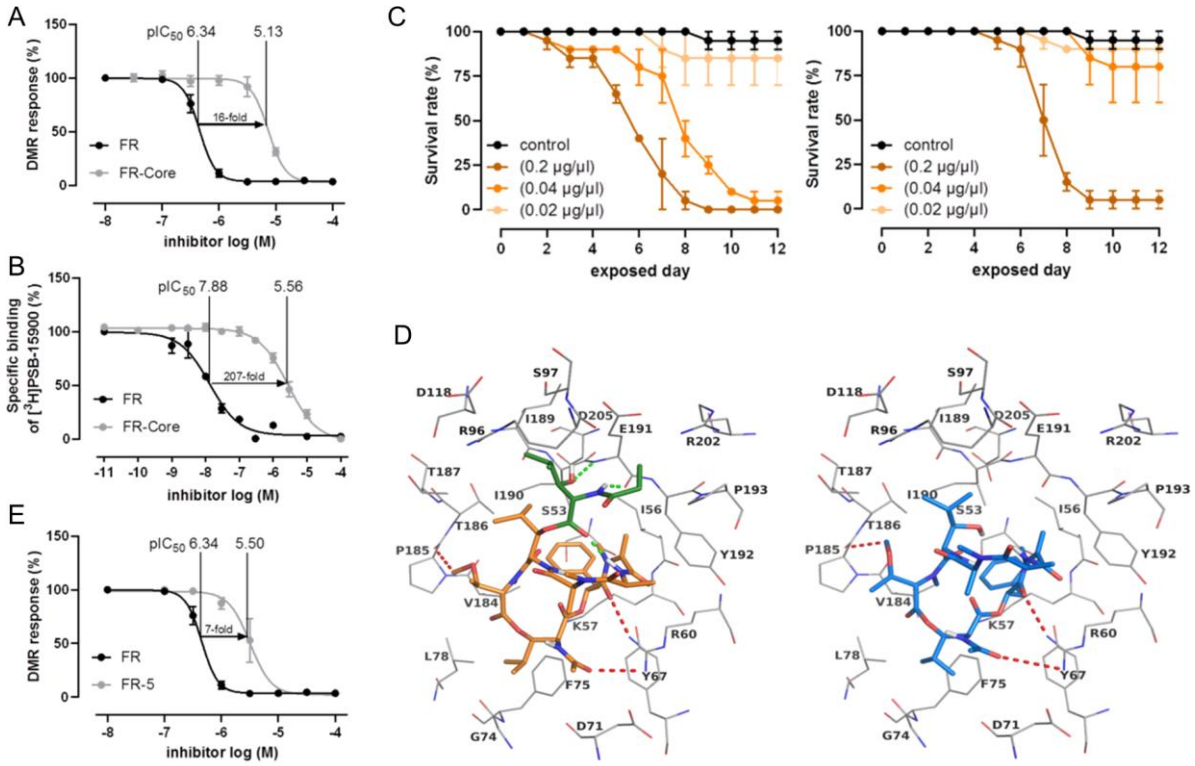


Figure III: Biological activity of FR-Core and FR-5, compared to FR. Taken from Hermes *et al.*[33] A. Concentration-dependent inhibition of activated Gαq proteins by FR and FR-Core as determined by label-free whole cell DMR biosensing. DMR recordings are representative (mean + s.e.m.) of at least four independent biological replicates conducted in triplicate. **B.** Competition binding experiments of FR and FR-Core versus the FR-derived radiotracer [³H]PSB-15900 at human platelet membrane preparation (50 µg protein per vial), incubated at 37 °C for 1 h. **C.** Exposure of nymphs of a stink bug (*Riptortus*

pedestris) to different concentrations of FR (left) and FR-Core (right), survival rate was measured. **D.** Docked poses of FR (left, represented in sticks and coloured in orange, the *N*-Pp-Hle group present only in FR is coloured in green) and FR-Core (represented in sticks and coloured in blue) in the binding pocket of the Gαq protein shown as line representation. Some of the interactions common for FR and FR-Core are indicated by red dotted lines, and the interactions specific for FR are shown as green dotted lines. Oxygen atoms are coloured in red, nitrogen atoms in blue and polar hydrogen atoms in white. **E.** Concentration-dependent inhibition of activated Gαq proteins by FR and FR-5 as determined by label-free whole cell DMR biosensing (see A).

Additionally, FR and FR-Core were fed to the nymphs of a stink bug to investigate its *in vitro* toxicity. FR killed the nymphs in concentrations down to 0.04 µg/ml while FR-Core was only effective in higher concentrations (Figure III C). This supports the assumption that FR could indeed be a defence agent against insect predators and the presence of the side chain improves this activity as well.

Comparison of Gαq inhibiting activities of the investigated FR derivatives revealed, that FR and FR-2 both exhibit nearly the same activity while FR-5 with a slightly enlarged acyl residue in the side chain was significantly less potent (Figure III E). This shows that, in terms of optimal Gαq inhibition, the variability of the acyl residue of the side chain is restricted to smaller residues.

The next part of this work dealt with the comparison of FrsA with the highly similar FrsD. While $FrsD_A$ showed lower substrate promiscuity than $FrsA_A$, $FrsD_C$ accepted acetyl- and propionyl-CoA for acylation in comparable yields to $FrsA_C$. FrsD also assembled the FR side chain *in vitro* which enabled us to test the transesterifying ability of $FrsA_{TE}$ in combination with FrsD. The *in vitro* combination of the two proteins generated FR and also FR-2 when supplied with the respective substrates. Based on the high similarity in sequence and activity of FrsA and FrsD, we discussed scenarios for the stepwise evolution of the FR side chain. We hypothesised FR-Core to be the less active ancestor molecule of FR. Its BGC *frsB-H* might have been extended by gene duplication of *frsD* and duplication of $frsG_{TE}$ followed by extensive adaption for intermolecular transesterification. This theory was supported by the fact, that no close phylogenetic relatives of the *frs* BGC or its single domains could be found in global BGC searches. Still, this hypothesis remains highly speculative and only the discovery and analysis of further related BGCs will give clearer insights into *frs* evolution.

The last part of this work dealt with the structural investigation of FrsA and $FrsA_{TE}$. Both proteins were heterologously expressed and digested to obtain native protein for purification. Size exclusion chromatography led to sufficient amounts of highly pure protein. To gain a three-dimensional structure of FrsA, cryo-EM measurements were attempted, as the crystallisation of such a big protein is challenging. For the smaller TE domain, we started crystallisation trials with promising intermediate results for X-ray diffraction analysis. The addition of FR-Core to the crystallisation attempts revealed a stabilizing effect of this substrate. A co-crystallisation could give further insights in this unusual TE domain and would shed light on the composition of the active site and the interaction with the substrates.

In essence, we investigated the biosynthesis of the FR side chain in detail, achieved its *in vitro* assembly and proved side chain intermolecular transesterification catalysed by $FrsA_{TE}$. Using the promiscuity of multiple domains, we generated and characterised the new analogue FR-5 and discussed the influence

of the side chain and different residues thereof for Gαq inhibitory activity. This research provides the basis for further biosynthetic engineering to generate new FR analogues and to fully understand the purpose of FR in its natural environment.

6 Material and Methods

6.1 Chemicals and reagents

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

6.2 Vectors and organisms

6.2.1 Vectors and plasmids

The vectors used in this work are listed in Table 6.1. The plasmid map of the vectors pET28a, pHis8-TEV and pCDF-Duet-1 are depicted in Supplementary Figure 9.5 to Figure 9.7.

Table 6.1: Used Vectors in this work.

Vector	Resistance	Reference
pET28a(+)	Kanamycin	Merck KGaA (Darmstadt, Germany)
pCDF-Duet-1	Apramycin	Prof. Craig Townsend, modified in AG König
pHis8-TEV	Kanamycin	Dr. René Richarz, AG König
pET-TEV-protease	Kanamycin	Dr. René Richarz, AG König
pCDF-FrsB	Apramycin	Daniel Wirtz, AG König

6.2.2 Organisms

In Table 6.2 all bacterial strains used in this work are summed up. The *frs* sequences of *C. vaccinii* have been deposited at GenBank, accession number MT876545.

Table 6.2: Bacterial strains used in this work.

Strain	Genotype	Reference
Escherichia coli (*E. coli*) BL21 DE3	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*	New England Biolabs (Frankfurt am Main, Germany).
E. coli α-Silver-Select (*α*-SS)	*fhuA2 D(argF-lacZ)U169 phoA glnV44 f80D(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	New England Biolabs (Frankfurt am Main, Germany).

E. coli BAP1 DE3[141]	BL21(DE3) Δ*prp*RBCD::T7prom-sfp,T7prom.prpE	Provided by group of Prof. Bradley Moore
Chromobacterium vaccinii (C. vaccinii) MWU205 DSM25250	Wild typ	DSMZ
C. vaccinii MWU205 Δ*frsA*	Δ*frsA*	Dr. René Richarz, AG König
C. vaccinii MWU205 Δ*vioA*	Δ*vioA*	Dr. René Richarz, AG König

6.3 Media and buffers

6.3.1 Media

All media used for cultivation are listed in Table 6.3. They were prepared with DI water and sterilised via autoclavation, all non-autoclavable ingredients were sterile filtered. The antibiotics used for cultivation are listed in Table 6.4.

Table 6.3: Media composition.

Medium	Ingredients	
Lurina bertani (LB) medium	10 g/L	tryptone
	5 g/L	yeast extract
	10 g/L	NaCl
		Adjust pH to 7.5 using NaOH
LB agar	10 g/L	tryptone
	5 g/L	yeast extract
	5 g/L	NaCl
	15 g/L	agarose
		Adjust pH to 7.5 using NaOH
Terrific broth (TB) medium	12 g/L	tryptone
	24 g/l	yeast extract
	5 g/l	glycerol
	900 ml	DI water
		Ad 100 ml TB salts (10x) after autoclaving

TB salts (10x)	0.17 M	KH_2PO_4,
	0.72 M	K_2HPO_4
		pH = 7,2
M9 minimal medium	33.7 mM	Na_2HPO_4
	22 mM	KH_2PO_4
	8.55 mM	NaCl
	9.35 mM	NH_4Cl
	0.4% (w/v)	Glucose
	1 mM	$MgSO_4$
	0.3 mM	$CaCl_2$
	0.134 mM	EDTA
	0.013 mM	$FeCl_3$
	6.2 µM	$ZnCl_2$
	0.76 µM	$CoCl_2$
	0.42 µM	CoCl
	1.62 µM	H_3BO_3
	0.081 µM	$MgCl_2$
SOC medium	0.5% (w/v)	yeast extract
	2% (w/v)	tryptone
	10 mM	NaCl
	2.5 mM	KCl
		Added steril filtered after autoclavation:
	10 mM	$MgCl_2$
	10 mM	$MgSO_4$
	20 mM	glucose

Table 6.4: Stock solutions of the used antibiotics (x1000).

Stock	Ingredients
Carbenicillin	50 mg/ml carbenicillin dissolved in ethanol
Chloramphenicol	25 mg/ml chloramphenicol dissolved in ethanol
Kanamycin	50 mg/ml kanamycin dissolved in water

Apramycin	50 mg/ml apramycin dissolved in water

6.3.2 Buffers

The buffer compositions for all standard buffers are summarized in Table 6.5, the buffers for protein purification and SDS-Page are listed separately in Table 6.6 and Table 6.7 respectively.

Table 6.5: Buffer composition.

Buffer	**Ingredients**	
TAE buffer 1x	1 mM	EDTA
	0.12% (v/v)	glacial acetic acid
	0.48% (w/v)	Tris
		pH = 8.3
C domain assay buffer	50 mM	Tris
	25 mM	NaCl
	10 mM	$MgCl_2$
		pH = 7.5
FPLC buffer I	20 mM	Tris
	300 mM	NaCl
	125 mM	L-arginine
		pH = 7.5
FPLC buffer II	20 mM	HEPES
	50 mM	NaCl
		pH = 7.5
HPLC-MS eluent stock solution	1250 ml	water
	9.75 g	ammonium acetate
	1 ml	glacial acetic acid
	1250 ml	acetonitrile
Eluent A	223,58 g	HPLC-MS eluent stock solution
	2250 g	Water
Eluent B	223,58 g	HPLC-MS eluent stock solution

	1755 g	acetonitrile

Table 6.6: Protein buffer composition.[202]

Buffer	Ingredients	
Denaturating buffer	100 mM	NaH_2PO_4
	10 mM	Tris
	8 M	urea
		Adjust pH to 8.0 using NaOH
Lysis buffer	50 mM	NaH_2PO_4
	300 mM	NaCl
	10 mM	imidazole
		Adjust pH to 8.0 using NaOH
Wash buffer I	20 mM	NaH_2PO_4
	300 mM	NaCl
	20 mM	imidazole
		Adjust pH to 8.0 using NaOH
Wash buffer II	20 mM	NaH_2PO_4
	300 mM	NaCl
	35 mM	imidazole
		Adjust pH to 8.0 using NaOH
Elution Buffer	50 mM	NaH_2PO_4
	300 mM	NaCl
	300 mM	imidazole
		Adjust pH to 8.0 using NaOH

Table 6.7: Composition of the buffers and solution used for the SDS-PAGE.

Solution	Composition
Staining solution	1 g Coomassie Blue, 100 ml glacial acetic acid, 400 ml methanol, 500 ml water
De-staining solution	200 ml methanol, 100 ml glacial acetic acid, 700 ml water
Separating buffer	1 M Tris-HCl, pH = 8.8
Stacking buffer	0.375 M Tris-HCl, pH = 6.8
10x glycine SDS electrophoresis buffer	250 mM Tris-HCl, 2 M glycine, 1% SDS, pH 9.0
Cathode buffer	100 mM Tris-HCl (pH 8.25), 100 mM Tricin, 0.1% SDS
Anode buffer	200 mM Tris-HCl, pH = 8.9

6.3.3 Kits

The kits and standards used in this work are listed in Table 6.8.

Table 6.8: Used Kits and standards and their manufacturers.

Kit	Manufacturer
Gene Ruler DNA Ladder Mix	Thermo Scientific
PageRuler Unstained Protein Ladder	Thermo Scientific
Blue Prestained Protein Standard	New England Biolabs
Fast Gene Gel/PCR Extraction Kit	NIPPON Genetics Co., Ltd
Fast Gene Plasmid Mini Kit	NIPPON Genetics Co., Ltd

6.4 Microbiological techniques

6.4.1 Cultivation of bacteria

The used bacteria were cultivated under the following conditions on liquid or solid media (Table 6.9). Additional parameters are given with particular experiments.

Table 6.9: Cultivation conditions of bacterial strains.

Strain	Medium (chapter 1.3.1)	Temperature [°C]	Rotation [min^{-1}]
E. coli (all strains)	LB, TB	37	220
C. vaccinii	LB, M9	25	200
C. vaccinii Δ*vioA*	LB	30	200
C. vaccinii Δ*frsA*/*vioA*	LB	30	200

6.4.2 Strain maintenance in cryogenic cultures

For long-time storage, bacterial strains need to be conserved at –80 °C. Therefore, an overnight culture was mixed 1:1 with sterile 50% (v/v) glycerol solution, filled into a Fisherbrand® cryogenic storage vial and stored in a -80 °C freezer. To inoculate medium from a cryogenic culture, the storage vial was kept frozen in a -20 °C vial-rack and immediately stored back to -80 °C after use.

6.4.3 Concentration determination of bacterial cultures

The growth of bacterial culture was determined by the increase of the optical density of the culture at 600 nm (OD_{600}). It was measured with an Eppendorf BioPhotometer kinetic against the particular sterile medium as a reference.

6.4.4 Transformation

Introduction of DNA into an organism is called transformation. In this work chemical (heat-shock) and electric transformation (electroporation) were used. Before the transformation can take place, the cells of the used bacterial stain need to be made competent.

6.4.4.1 Production of chemocompetent cells

700 µl of an overnight culture were transferred in 700 ml fresh LB medium and cultivated at 37 °C and 220 min^{-1}. At an OD_{600} of 0.4 the cells were centrifuged (4 °C, 7,441 g, 10 min) and the supernatant discarded. From now on the cells were kept on ice. The pellet was resuspended in 10 ml cold 70 mM $CaCl_2$, 20 mM $MgSO_4$. The cells were incubated on ice for 30 min, then centrifuged again (4 °C, 7,441 g, 10 min) and resuspended in 3,5 ml of 70 mM $CaCl_2$, 20 mM $MgSO_4$. After another 30 min on ice 875 µl glycerol was added. The cell suspension was aliquoted to 100 µl, shock frosted in liquid nitrogen and then stored at -80 °C.

6.4.4.2 Chemical transformation

One aliquot of chemocompetent cells was thawed on ice and 10 µl of the ligation or up to 100 ng of plasmid were added and incubated on ice for 20 min. Afterwards, a heat-shock at 42 °C was performed for 90 s. After 2 min on ice, 1 ml SOC medium was added. The cells were incubated for 1 h at 37 °C and 220 rpm. Then they were plated on selective agar and further incubated at 37 °C overnight.

6.4.4.3 Production of electrocompetent cells

An overnight preculture was used to inoculate 200 ml of LB medium. The culture was incubated at 37 °C and 220 rpm. At an OD_{600} of 0.4 the cells were centrifuged (4 °C, 2,907 g, 5 min) and the supernatant discarded. From now on the cells were kept on ice. The cells were successively washed with 200 ml, 100 ml and 50 ml of sterile, cold 10% (v/v) glycerol by centrifugation (4 °C, 2,907 g, 5 min) and the supernatant discarded. In the final step, the cells were resuspended in 1 ml 10% (v/v) glycerol and aliquoted to 70 µl. Cells were always prepared freshly before use.

6.4.4.4 Electroporation

The freshly prepared electrocompetent cells were mixed with up to 100 ng of the DNA to be introduced and loaded into a pre-chilled electroporation cuvette (1 mm). 25 kV/cm was applied in the Biorad MicroPulser™ and 1 mL SOC medium was immediately added to the cells. The cells were incubated for 1 h at 37 °C and 220 rpm. Then the cells were plated on selective agar and further incubated at 37 °C overnight.

6.5 Molecular biological methods

6.5.1 Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) is a molecular biological method to amplify DNA sequences. The initial solution and PCR protocol depend on the used polymerase. In this work, Q5 High-fidelity DNA

polymerase (New England Biolabs) and GoTaq polymerases (Promega) were used. The *Taq* polymerase was used for colony screening PCR reactions, the Q5 polymerase for the amplification of DNA for plasmid construction. As a template, the genomic DNA of *Chromobacterium vaccinii* MWU205 was used.

6.5.1.1 Q5 polymerase PCR

The composition of a Q5 polymerase reaction is shown in Table 6.10. The reaction was conducted in a thermocycler using the protocol in Table 6.11.

Table 6.10: Composition of a Q5 polymerase PCR reaction.

Substance	Amount
5x Q5 buffer (NEB)	5 µl
5x Q5 GC Enhancer (NEB)	5 µl
dNTPs (10 mM)	0,5 µl
Forward primer (10 µM)	0,625 µl
Reverse primer (10 µM)	0,625 µl
DNA	1 µl, < 1 µg
Q5 High-fidelity DNA polymerase (NEB)	0,125 µl
H_2O	ad 25 µl

Table 6.11: Protocol of a Q5 polymerase PCR reaction.

Step	Temperature	Time	
1. Initial denaturation	98 °C	30 s	
2. Denaturation	98 °C	10 s	30 Cycles
3. Annealing	Variable	20 s	
4. Polymerisation	72 °C	20-30 s per kb	
5. Final polymerization	72 °C	120 s	

6.5.1.2 *Taq* polymerase colony PCR

The composition of a typical *Taq* polymerase PCR reaction is shown in Table 6.12; the protocol is shown in Table 6.13. To prove specific recombinant DNA, the PCR protocol for the *Taq* polymerase can be used for a colony screening PCR.

Therefore, several different colonies were picked from an agar plate with a pipette tip. One part of the colony was transferred onto a designated area of another agar plate. The rest of each colony was

suspended in 10 µl sterile water. This suspension was then used as template DNA for the *Taq* PCR. For positive control, genomic DNA of *C. vaccinii* was used and sterile water as a negative control.

Table 6.12: Initial composition of a *Taq* polymerase PCR reaction.

Substance	Amount
GoTaq Green buffer (5x)	4 µl
$MgCl_2$ (25 mM)	1 µl
DMSO	1 µl
dNTPs (10 mM)	0,33 µl
Forward primer (10 µM)	0,33 µl
Reverse primer (10 µM)	0,33 µl
GoTaq polymerase	0,1 µl
DNA	1 µl, < 1 µg
H_2O	20 µl

Table 6.13: Protocol of a *Taq* polymerase PCR reaction.

Step	Temperature	Time	
1. Cell lysis	95 °C	300 s	
2. Denaturation	95 °C	45 s	30 Cycles
3. Annealing	Variable	45 s	
4. Polymerisation	72 °C	30 s per kb	
5. Final polymerisation	72 °C	300 s	

6.5.1.3 Primer design

Primers were designed with CloneManager 9.2 software and purchased from Eurofins MWG Operon (Ebersberg, Germany). For primer design, an annealing sequence of around 18 base pairs was searched in the region of interest. For constructs with N-terminal His tags, a stop codon was added outside the annealing sequence. Then, a restriction site outside the protein-coding sequence was added for *in-frame* cloning, so later on the amplified domain/gene could be ligated into an expression vector. The reading frame of the protein sequence has to be taken into account. Outside of the restriction site, three to six base pairs overhang were added to ensure correct restriction. Primers were diluted in sterile water to a final concentration of 100 pmol/µl and stored at -20 °C.

6.5.1.4 Used oligonucleotides (Primer)

In Table 6.14 all primers used for cloning of enzymes and enzymatic domains from *frs* are listed. In Table 6.15 the primers used for the construction of *C. vaccinii* MWU205 deletion mutants are listed.

Table 6.14: Primers used for cloning. Restriction sites are bold and stop codons underlined. All overhangs are given in lowercase, while target specific sequences are given in uppercase.

Primer name	Sequence (5'→3')
Cv_frsA_His6-N_for_*Bam*HI	gat **ggatcc** ATGAAAAACAGTGAATCGC
Cv_frsA_His6-N_rev_*Hind*III	tat **aagctt** TTATTGCTTGACAGCGGTGAC
Cv-frsA-T_His6-N_rev_*Hind*III	tat **aagctt** tcaGCTGTCGCCGCCTTCGGC
Cv-frsA-T_His6-N_for_*Bam*HI	gat **ggatcc** GGCTCGCATTATCAG
Cv-frsA-A_His6-N_for_*Bam*HI	tat **ggatcc** CCGTCGCAGCCGGTGTCC
Cv-frsA-A_His6-N_rev_*Hind*III	gtc **aagctt** ttaCCGCTGATAATGCGAGCC
Cv-frsA-TE_His6-N_for_*Bam*HI	tat **ggatcc** GCCGAAGGCGGCGACAGC
Cv_frsB_for_*Nde*I	gcg **catATG**AGCAATCCCTTTGATGAT
Cv_frsB_rev_*Pac*I_pCDF	gcg **ttaattaa** ttaTTTATCATCGCACTCCAT
Cv_frsD_His6-N_for_*Hind*III	cac **aagctt** tgATGGAAATATGGCTGGCG
Cv_frsD_His6-N_rev_*Xho*I	tat **ctcgag** TCAACTCCTGACAGCGTG
Cv_frsH_His6-N_for_*Bam*HI	gat **ggatcc** ATGACCGTATCCGATAAC
Cv_frsH-His6-N_rev_*Xho*I	gata **ctcgag** tTACAGCAGCATGGTTTG
Cv_frsG-TE_His6-C_for _*Nhe*I	tat **gctagc** ATGGACGGCGAGATCGATGAC
Cv_frsG-TE_His6-N_rev _*Hind*III	tat **aagctt** TCAAGAATTACGGCGGGTGGACTG

Table 6.15: Primers used for the construction of *C. vaccinii* MWU205 deletion mutants. Restriction sites are bold, Gibson homology arms are underlined. All overhangs are given in lowercase, while target specific sequences are given in uppercase. All primers designed by Dr. René Richarz.

Primer name	Sequence (5'→3')	Description
*Bam*HI-FRT_for	tga **ggatcc** AGCTTCAAAAGCGCTCTGA	Sequential cloning of FRT into pUC19
*Sal*I-FRT_rev	tgt **gtcgac** GGGGATCTTGAAGTTCCT	Sequential cloning of FRT into pUC19

*Sph*I-frsA-up_for	agt **gcatgc** GGAAAGTACGTCTGGTCTTG	Sequential cloning of the *frsA*-up region into pUC19::FRT
*Sal*I-frsA-up_rev	tct **gtcgac** TACATCCAGCTGTGCTGAAG	Sequential cloning of the *frsA*-up region into pUC19::FRT
*Bam*HI-frsA-dn_for	ttc **ggatcc** ATTGGTCCTGTTCTCGAGTC	Sequential cloning of the *frsA*-dn region into pUC19::*frsA*-up-FRT
*Sac*I-frsA-dn_rev	tga **gagctc** AGTCCCGCATATGATCGATG	Sequential cloning of the *frsA*-dn region into pUC19::*frsA*-up-FRT
FRT_for	CGAATTAGCTTCAAAAGCGCTCTGA	One step cloning of FRT into pEX18Tc
FRT_rev	CGAATTGGGGATCTTGAAGTTCCT	One step cloning of FRT into pEX18Tc
Gib-vioA-up_for	gcatgcctgcaggtcgactctaga**ggatcc** TGACCCTTGGAACAGGATG	One step cloning of *vioA*-up into pEX18Tc
FRT-vioA-up_rev	aggaacttcaagatccccaattcg CTGCTGCATGTCGAAAATG	One step cloning of *vioA*-up into pEX18Tc
FRT-vioA-dn_for	tcagagcgcttttgaagctaattcg CGTCCATGTGCACAAGTAC	One step cloning of *vioA*-dn into pEX18Tc
Gib-vioA-dn_rev	tacgaattcgagctcggtacccggg GCTCGCCATTGATCGAAAC	One step cloning of *vioA*-dn into pEX18Tc
PCR-frsA-KO_for	GTAATGTCAAAGGCTTGG	Mutant verification of *C. vaccinii* Δ*frsA* by PCR / Sequencing
PCR-frsA-KO_rev	ATTGAATTGCTGACACCG	Mutant verification of *C. vaccinii* Δ*frsA* by PCR / Sequencing
PCR-vioA-KO_for	AGCTCTACCTGTGGCAG	Mutant verification of *C. vaccinii* Δ*vioA* by PCR / Sequencing
PCR-vioA-KO_rev	TCCCAGGAGAAATGGTTG	Mutant verification of *C. vaccinii* Δ*vioA* by PCR / Sequencing

6.5.1.5 Purification of PCR products

The PCR products were purified before further use. The “FastGene Gel/PCR Extraction Kit” (NIPPON Genetics Co.) was used according to the manufacturer’s instructions. For the elution step, autoclaved DI water was used instead of the provided elution buffer.

6.5.1.6 Concentration determination of DNA

The concentration of the purified DNA, PCR fragments, gDNA or plasmids was quantified using photometric absorption measurements at 260 nm with an Eppendorf µCuvette in an Eppendorf BioPhotometer.

6.5.2 DNA isolation

6.5.2.1 Isolation of plasmid DNA

To isolate plasmid DNA from bacteria, the "Plasmid mini kit" (NIPPON Genetics Co.) was used, following the manufacture's instructions. Therefore 4 ml of an overnight culture were used. Plasmids were eluted in 20-40 µl autoclaved DI water. The concentration of the purified plasmid was quantified like described in 6.5.1.6.

6.5.2.2 Isolation of genomic DNA from *C. vaccinii*

Genomic DNA was isolated from stationary liquid cultures with the "Wizard® Genomic DNA Purification Kit" (Promega). The kit was used, following the producer's gDNA extraction protocol for gram-negative bacteria.

6.5.2.3 Agarose gel electrophoresis

The agarose gel electrophoresis is used to separate DNA fragments by size to analyse or purify them. The standard gel contains 1% (w/v) agarose in 1x TAE buffer, solved through heating in the microwave. The solution was cooled down to 60 °C and then poured into an electrophoresis chamber. A comb with a certain amount of sample wells is inserted and the gel cooled down. DNA samples were mixed with loading dye (6x) in a 5:1 ratio (v/v) and filed into the sample wells. The "GeneRuler DNA Ladder Mix" (Thermo Fisher Scientific) was used as a size/reference marker. The gel was run in 1x TAE buffer at 120 V for 20 to 30 min. Afterwards, the gel was stained in ethidium bromide solution for 5 min and destained in water for another 5 min. The DNA fragments were visualized in UV light.

6.5.2.4 DNA isolation from agarose gels

For preparative gel electrophoresis, 0.7% agarose gels were used. The gel was run analogously to 6.5.2.3 and the selected band cut out with a clean scalpel. The "FastGene Gel/PCR Extraction Kit" (NIPPON Genetics Co.) was used for purification, according to the manufacturer's instructions. PCR products were eluted in 20-40 µl autoclaved DI water.

6.5.3 DNA sequencing

To check the correct sequence of plasmid DNA, the constructs were sent to Eurofins Genomics Germany GmbH (Ebersberg, Germany) for Sanger sequencing. Results were supplied as FASTA in .txt-format, or as chromatogram in .pd4-format and analysed with SnapGene 5 or Chromas. The primers used for sequencing are listed in Table 6.16.

Table 6.16: Primers used for Sanger sequencing.

Primer	Sequence	Reference
T7	TAATACGACTCACTATAGGG	Eurofins genomics standard primer
pET-RP	CTAGTTATTGCTCAGCGG	Eurofins genomics standard primer
DuetUP2	TTGTACACGGCCGCATAATC	Addgene vector database
Cv_FR-Gap1_for	TCGAGATGATGAAGGCTG	Dr. René Richarz
Cv_FR-Gap1_rev1	GGCATCAACACTTGATAAG	Dr. René Richarz
Cv_FR-Gap1_rev2	CATCCAGCACGTACAGC	Dr. René Richarz
Cv_FR-Gap1_rev3	AGGCTTTGCAGATGGCG	Dr. René Richarz
Cv_FR-Gap2_for	TCCACCTCGATTTGTACG	Dr. René Richarz
Cv_FR-Gap2_for2	CCTATGTGATCTACACCTC	Dr. René Richarz

6.5.4 Restriction enzyme digestion of DNA

For restriction of DNA, restriction endonucleases type II supplied by New England Biolabs were used according to the manufacturer's instructions. The composition of typical restriction digestion is listed in Table 6.17. The analytical digestions of vectors were incubated at 37 °C for 1 h, for preparative digestions, the incubation time was chosen according to the manufacturer's recommendations. The enzymes were heat-inactivated and digested DNA was analysed by gel electrophoresis. For preparative digestion, before the analysis, digested vector molecules were dephosphorylated (see 6.5.5). The selected DNA fragments were purified from the gel (see 6.5.2.4.) or with the PCR purification kit (see 6.5.1.5).

Table 6.17: Composition of restriction digestion.

Volume analytical digestion	Volume preparative digestion	Compound
8 µl	**25 µl**	DNA
1 µl	**3 µl**	10x CutSmart buffer (NEB)
0.5 µl	**1 µl**	Restriction enzyme 1
0.5 µl	**1 µl**	Restriction enzyme 2

6.5.5 Dephosphorylation of vector molecules

Before the ligation, the 5'-end of the linearized plasmid was dephosphorylated to avoid re-ligation. Therefore, alkaline phosphatase FastAP (Fermentas) was added to the restriction digest 30 min before ending the incubation. This enzyme was heat-inactivated together with the restriction enzymes.

6.5.6 Ligation

The ligation of complementary DNA ends was performed with T4 ligase (Promega). Therefore, 1 µl of ligase, 1 µl reaction buffer and 8 µl of DNA were incubated at 4 °C overnight. The DNA consisted of a 1:1, 1:3 or 1:5 molar ratio of vector to insert. The ligation mixture was then transformed into competent cells, see 6.4.4.

6.5.7 Cloning of plasmids

After ligation of the plasmid and transformation in a cloning strain like αSS, colonie PCRs (see 6.5.1.2) were performed to screen for clones carrying the plasmid with the inserted gene. These colonies were grown overnight, the plasmid isolated (6.5.2.1) digested (6.5.4) and analysed via gel electrophoresis (6.5.2.3). Positively screened plasmids were sent for Sanger sequencing (6.5.3). Cryogenic stocks of clones with the verified plasmid sequence were stored and used for further cultivation and plasmid isolation. All used constructs are listed in Table 6.18.

Table 6.18: Summary of all used expression constructs.

Construct	Vector	Insert	Gene flanking restriction sites	Size insert
pCH01	pET28a	*frsA*	*Bam*HI/*Hind*III	3829 bp
pCH02	pET28a	$frsA_{CAT}$	*Bam*HI/*Hind*III	3091 bp
pCH03	pET28a	$frsA_{AT}$	*Bam*HI/*Hind*III	1816 bp
pCH04	pET28a	$frsA_{A}$	*Bam*HI/*Hind*III	1528 bp
pCH05	pET28a	$frsA_{TTE}$	*Bam*HI/*Hind*III	1000 bp
pCH06	pET28a	$frsA_{TE}$	*Bam*HI/*Hind*III	769 bp
pCH07	pET28a	*frsD*	*Hind*III/*Xho*I	3093 bp
pCH08	pET28a	$frsG_{TE}$	*Nhe*I/*Hind*III	769 bp
pCH09	pHis8-TEV	*frsA*	*Bam*HI/*Hind*III	3829 bp
pCH10	pHis8-TEV	$frsA_{TE}$	*Bam*HI/*Hind*III	769 bp
pDW01	pCDF-duet	*frsB*	*Nde*I/*Pac*I	225 bp

6.5.7.1 Cloning of *in vitro* expression constructs in *E. coli*

To investigate the *frs* gene cluster, the genes or DNA stretches encoding the biosynthetic modules or single domains were cloned into expression plasmids. These plasmids have coding sequences for N-terminal His tags, attaching 6-8 histidines to the proteins during translation, qualifying them for Ni-affinity purification. This general workflow was used for cloning: Via PCR and specific primers, the selected genes were amplified (6.5.1.1), purified (6.5.1.5) and double digested with enzymes also used for digestion of the selected plasmid (6.5.4). The linearized plasmid was additionally dephosphorylated (6.5.5) and then the insert was ligated into the plasmid (6.5.6) and the generated plasmid was sequenced

(6.5.3). The verified plasmids were introduced into expression strains *E. coli* BL21 or BAP1 via electrical or chemical transformation (6.4.4), and the transformation success checked by colony PCR (6.5.1.2). Successful clones were maintained in cryogenic cultures (6.4.2).

6.5.7.2 Construction of coexpression strains in *E. coli*

For coexpression of different constructs, a combination of pET28a and pCDFDuet-1 vectors was used. For coexpression, both vectors were transformed simultaneously into competent *E. coli* cells (see 6.4.4) and selection was performed on selective agar containing the respective antibiotics for both vectors. Alternatively, competent cells of an existing clone were prepared for electroporation and the second vector was transformed into these cells.

6.5.8 Construction of the *frsA* and *vioA* knock-out vectors

The knock-out vectors were constructed by Dr. René Richarz as described in Hermes *et al*.[33] At first, a classical multiple cloning procedure (sequential cloning) was used to generate the *C. vaccinii* MWU205 Δ*frsA* mutant. Later, a more elaborate method, based on Gibson assembly (one-step cloning) was employed to delete *vioA* (see Figure 4.15). For both approaches, the upstream (up) and downstream (dn) DNA sequences of the respective genes (ca. 1.1 kb for *frsA* and ca. 0.4 kb for *vioA*) were amplified by Q5 PCR with suitable primers (Table 6.15). In addition, the 1.8 kb FRT cassette (Gem^R Gfp^+) (FRT) from pPS858[153] was amplified by Q5 PCR. All PCR reactions were prepared in a 25 µl scale as described above with either 2–22 ng *C. vaccinii* MWU205 gDNA (*frsA* and *vioA* up- and downstream) or 0.5–4.5 ng pPS858 (FRT) as a template. The fragments for sequential cloning were then cloned into pUC19[203] in a directional manner in the following order: FRT > *frsA*-up > *frsA*-dn. For this, standard restriction/ligation based cloning techniques were used. The insert of the resulting plasmid pUC19::Δ*frsA*, consisting of all three fragments, was then subcloned into pEX18Tc by using the flanking restriction sites for *Sac*I and *Sph*I. This yielded the final knock-out vector pEX18TC::Δ*frsA*. In contrast, the one-step cloning approach followed the protocol of Gibson *et al*.[204] For this, a reaction mixture (20 µl scale) containing 4.5 nM of the PCR amplified inserts (FRT, *vioA*-up, *vioA*-dn), 1.375 nM *Bam*HI linearized pEX18Tc, 0.0054 U/µl T5 exonuclease (NEB), 0.034 U/µl Phusion polymerase (NEB), 5.4 U/µl *Taq* DNA ligase (NEB) and 1x ISO buffer (10 mM $MgCl_2$, 200 µM deoxynucleotide triphosphates, 10 mM DTT, 5% (w/v) PEG-8000, 1 mM NAD, 100 mM Tris-HCl pH 7.5) was set up. This mixture was then incubated at 50 °C for 1 h. 5 µl of this reaction mixture were then transformed into chemically competent *E. coli* NEB Turbo cells (NEB). Positive clones were initially identified by colony PCR. For this, screening reactions were performed in a 25 µl scale with 1x GoTaq reaction buffer with 1.5 mM $MgCl_2$, 4% DMSO, 200 µM dNTP, 100 nM of forward and reverse primer, DNA template (bacterial suspension in water) and GoTaq G2 DNA polymerase (0.025 U/µl, Promega). Plasmids from positive clones were then verified by restriction digest analysis as well as terminal-end Sanger sequencing. This yielded the final knock-out vector pEX18Tc::Δ*vioA*.

6.5.9 Preparation of knockout mutants

The knock-put strains *C. vaccinii* MWU205 Δ*frsA* and *C. vaccinii* MWU205 Δ*vioA* were prepared by Dr. René Richarz by the protocol below published in Hermes *et al*.[33] Transfer of pEX18Tc::Δ*frsA* or pEX18Tc::Δ*vioA* into *C. vaccinii* MWU205 occurred by triparental conjugation employing *E. coli* NEBTurbo harbouring one of the knockout vectors and the conjugational helper strain *E. coli* ET12567 pUB307. All three strains were grown in 15 ml LB at 30 °C, 220 rpm to an OD600 of 0.4-0.6. Cells were then washed two times with 10 ml LB to remove any antibiotics and were finally resuspended in an appropriate amount of LB. The resulting suspensions were then mixed in a 3:1 ratio of donor strains and acceptor strain (600 µl *E. coli* NEBTurbo with pEX18Tc::Δ*frsA* or pEX18Tc::Δ*vioA*, 600 µl *E. coli* ET12567 pUB307, 200 µl *C. vaccinii* MWU205), centrifuged (2 min, 9418 g, RT) and the pellet resuspended in a small part of the supernatant (ca. 100 µl). This mixture was applied to an LB agar plate without antibiotics in the form of a 'puddle'. After drying, the plate was incubated for 24 h at 30 °C. On the next day, the resulting cell layer was scraped off and resuspended in 1 ml LB without NaCl (NS-LB). The suspension was then plated on NS-LB agar containing Amp200, Gem30 and 15% sucrose (w/v) and the plates were incubated for 60–72 h at 25 °C. Resulting clones were screened for a successful double homologous recombination event by colony PCR as described above. Three positive clones were then further tested for an integration of the FRT cassette at the right genomic locus. For this, the respective regions were amplified by Q5 PCR with primers binding outside the sequences used for construction of the knockout vectors (Supplementary Table 6, Supplementary Fig. 7a). All PCR reactions were prepared in a 25 µl scale with 75-200 ng of *C. vaccinii* MWU205 mutant gDNA as template as described above. Resulting PCR products which showed the expected size difference (minus 1,897 bp for Δ*frsA*::FRT, plus 879 bp for Δ*vioA*::FRT) compared to the wild type (6,144 bp for *frsA*, 2,043 bp for *vioA*), were purified and applied to terminal-end Sanger sequencing along with the primers used for amplification. As pUB307 is a self-transferable helper plasmid, all correct clones were routinely tested for the loss of pUB307 by plating them on kanamycin and tetracycline. Only clones, who failed to grow on either antibiotic, were considered for further usage as this indicates a loss of the plasmid. Alternatively, cells were cured from pUB307 by growing them in LB supplemented with 5% (w/v) SDS at 30 °C for 24–48 h and replating them on selective agar. To remove the FRT cassette from the genome of the *C. vaccinii* Δ*frsA*::FRT or Δ*vioA*::FRT deletion mutants, the genes *flp* and *sacB* from pFlp2[153] were introduced into the broad host vector pBMTL-2[205]as described by Wang *et al*.[206] The resulting vector pBMTL-2::*flp-sacB* was then transferred to the deletion mutants by electroporation following the protocol established for *Chromobacterium violaceum*.[207] After the transformation clones were grown overnight at 30 °C, which has been shown to be sufficient for FRT removal by Flp mediated site-specific recombination.[153] Positive clones that lost the FRT cassette were identified by colony PCR and plated on NS-LB with Amp200 and 15% sucrose (w/v) to remove the pBMTL-2::*flp* plasmid. Loss of the plasmid was then confirmed by colony PCR as well as testing the clones for kanamycin susceptibility. To verify modification of the correct genomic locus, three clones were further investigated by Q5 PCR

as described above. In this case, removal of the FRT cassette was indicated by the formation of a smaller PCR product (loss of 1,725 bp) compared to the mutants with FRT cassette. The mutants verified in this manner were termed *C. vaccinii* MWU205 Δ*frsA* and *C. vaccinii* MWU205 Δ*vioA*, respectively. The double mutant *C. vaccinii* MWU205 Δ*frsA*/Δ*vioA* was constructed by me as described above by using *C. vaccinii* MWU205 Δ*frsA* as starting strain.[33] Mutant strains were grown and extracted analogous to the wild type strain, see 6.10.1.

6.6 Protein expression and purification

Here, the standard protocol for protein expression and purification is given. All differing procedures for special experiments are stated with the particular experiments or in the results part. The used buffers are listed in Table 6.6.

6.6.1 Protein overexpression

An LB preculture inoculated with 5 µl of cryogenic culture was cultivated overnight at 37 °C and 220 rpm. A variable amount of TB medium with appropriate antibiotics in baffled flasks was inoculated with 1% of preculture at 37 °C and 220 rpm until it reached an OD_{600} of 0.9-1.1. Then, it was cooled on ice to 16 °C, 0.4 mM isopropyl-β-D-thiogalactopyranoside (IPTG) was added and incubation proceeded at 16 °C, 200 rpm for a further 16 h. Finally, the cells were harvested via centrifugation (4,000 *g*, 4 °C, 15 min).

Expression of constructs including T domains was performed in *E. coli* BAP1, to ensure *in vivo* phosphopantetheinylation of the T domains. All other constructs were expressed in *E. coli* BL21 (DE3). For the expression of FrsH, TB media was supplemented with 25 µM Fe(III)-citrate.

6.6.2 Cell lysis

For *in vitro* assays with heterologously overexpressed proteins, after expression the harvested cells (6.6.1) were resuspended in 2.5 ml lysis buffer per g pellet and lysed on ice with a sonicator. The sonicator was used at 50 W with 40 Hz for 10 s, then the suspension was cooled on ice for 10 s and the sonication repeated nine times. Afterwards, the lysate was centrifuged (10.000 *g*, 4 °C, 10 min). The pellet of cell debris was resuspended in denaturating buffer (Table 6.6) for an SDS-PAGE sample of the insoluble proteins.

6.6.3 Ni-NTA affinity chromatography

All recombinant proteins were expressed with a polyhistidine tag, predominantly a hexa histidine tag (His6-tag) at the N-terminus of the protein. These histidines can coordinate polyvalent cations like Nickel and offer the possibility to use affinity chromatography to separate tagged proteins from the cell lysate. After the cell lysis (6.6.2), the supernatant of the lysate was incubated on ice with 0.5-1 ml Ni-NTA-agarose (Qiagen), which was washed with lysis buffer beforehand, and the suspension was kept in light movement. After 1 h, the suspension was filtered with a propylene column (Qiagen 1 ml). The

Ni-NTA agarose was washed with 4 ml of wash buffer I and 4 ml of wash buffer II and eluted with 2.5 ml of elution buffer. Of all protein fractions 15 µl were analysed by SDS-PAGE, see 6.7.

6.6.4 Rebuffering and concentrating of proteins

The elution fraction of the Ni-NTA chromatography was rebuffered with a PD10 column (GE) following the manufacturer's gravity protocol. The eluate was concentrated with Vivaspin 500 columns (Sartorius) with a molecular weight cut off (MWCO) 30 kDa or 10 kDa according to the manufacturer's instructions. The selected MWCO was a maximum of half the size of the corresponding protein. High amounts (>>10 mg) of protein were also desalted using spin filters by concentrating and diluting the solution at least 3 times with the desired buffer.

6.6.5 Protease digestion for His Tag removal

To generate a pure protein without any tags for crystallisation experiments, the particular genes were ligated into a pHis8-TEV vector. This special vector has a recognition site for the pTEV protease which cleaves the peptide bond specifically at one site. Therefore the protein with His8-tag and the pTEV were purified via Ni-NTA affinity chromatography and rebuffered as described in 6.6.3 and 6.6.4. Than protein and protease were mixed in an mg/ml ratio 1:5 and incubated overnight at 4 °C. Afterwards 0.5 ml Ni-NTA was added to remove the protease and all undigested proteins. The suspension was shaken on ice for 1 h, filtered with a propylene column and washed with 6 ml of wash buffer I. The flow through and the wash fraction were run through a second column with 0.5 ml Ni-NTA, pooled, concentrated and passed on to FPLC purification (6.6.6). Both Ni-NTA columns were eluted with 1 ml of elution buffer for SDS-PAGE analysis.

6.6.6 Fast protein liquid chromatography (FPLC)

For further purification of expressed proteins after the affinity chromatography, a size exclusion chromatography (SEC) step was performed. Therefore an Äkta-FPLC (Cytiva) with a 280 nm UV detector was used, connected to a Superdex 200 Increase 10/300 GL column, a 500 µl sample loop and the FPLC buffers I or II for isocratic elution (Buffer composition Table 6.5). All solutions were filtered and degassed before use. The flow was set as 1 ml/min for wash and equilibration and 0.4 ml/min for elution. After 1 column volume (CV) equilibration 100- 500 µl protein solution, filtrated or centrifuged (10 min, 11,627g, 4 °C), were injected and 200 µl fractions collected between 0.35 and 1.2 CV runtime. All collected fractions were analysed via SDS-PAGE, see 6.7, and the fractions with the desired protein were pooled and concentrated with a Vivaspin column. Chromatograms were analysed with UniCorn 4.1 Software. The superdex column was washed with 2 CV DI water and then with 2 CV 20% EtOH for storage.

6.6.7 Protein storage

To store proteins over longer periods and ensure their activity, they were shock frosted in liquid nitrogen. Therefore, the protein buffer had to contain 5-10% glycerol and the protein was frozen in small amounts, 50-200 µl, in PCR tubes and directly transferred into a -80 °C freezer.

6.7 Protein analysis

6.7.1 Sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE)

The buffers for the SDS-PAGE are listed in Table 6.7. The separating gel was prepared first, filled in SDS-PAGE-gel chamber Novex NC2015 1.5 mm (Life technologies) and covered with isopropanol to get a smooth surface. After the gel had polymerised, the isopropanol was removed, the stacking gel was applied and a comb was inserted to create defined gel pockets. The composition of the gels is listed in Table 6.19.

Table 6.19: Composition of the SDS-PAGE gels.

Gel	Composition
Separating gel	4 ml acrylamide (30% acrylamide, bisacrylamide ratio 37.5:1), 3.3 ml water, 2.5 mL separating buffer, 100 µl 10% SDS, 100 µl 10% APS, 4 µl TEMED
Stacking gel	0.51 ml acrylamide (30% acrylamide, bisacrylamide ratio 37.5:1), 2.04 ml water, 375 µl stacking buffer, 30 µl 10% SDS, 30 µl 10% APS, 3 µl TEMED

15 µl of a protein sample, collected during protein purification (6.6.3 and 6.6.6), were mixed with 5 µl NuPAGE LDS sample buffer (4x) (Thermo Scientific) and 2 µl NuPAGE Reducing Agent (10x) (Thermo Scientific). The mixed solution was incubated at 70 °C for 10 min. After heating, 5-10 µl of protein solution were loaded in a pocket of the SDS-PAGE gel. Additionally, 3 µl of PageRuler Unstained Protein Ladder (Thermo Scientific) were applied in a free gel pocket. The glycine SDS electrophoresis buffer was used for standard measurements and the anode and cathode buffer system for small proteins like FrsB or lone standing T domains. The proteins were first stacked by using 90 V for 30 min and then separated by 120 V for 90 min.

The gels were washed in DI water for 5 min, stained for 20 min in staining solution and then de-stained in the de-staining solution for 16 h. Incubation took place on a shaking plate and for de-staining paper was added to the solution.

6.7.2 Protein concentration determination

The concentration of purified protein was determined using photometric absorption measurements by the method of Gill and Hippel at 280 nm with an Eppendorf µCuvette in an Eppendorf BioPhotometer kinetic.[208] The molar extinction coefficient at λ=280 nm was estimated for every protein separately by

the number of tryptophans, tyrosines and disulfide bonds in the protein. The photometer directly calculated the protein concentration (mg/ml). The appropriate buffer was used as a blank measurement.

6.7.3 Thermal shift assay

The thermal shift assay (TSA) is used to determine the thermal stability of a protein in different environments. The used method is the Thermofluor assay, in which a compound with a low fluorescence signal in a polar environment (such as in aqueous solution) but with high fluorescence in a nonpolar environment, is added to a protein solution. The fluorescence of the solution is monitored while the solution is heated. When the protein chain unfolds, the hydrophobic core gets exposed to the fluorescent dye and the signal increases until all protein molecules are completely denatured. The temperature, at which half of the protein population is unfolded is defined as T_M and equal to the melting point of the protein.[188]. In this work, the fluorescent dye SYPRO orange (Merck KGaA, Darmstadt, Germany) was used.

For the assay, SEC-purified protein with removed His tag (6.6.5) was used in a final concentration of 10-20 µM with 2-20-times SYPRO orange concentration. All buffers tested are listed in Table 6.20. The triplicate samples were measured with the Quiagen Rotor-Gene Q6 plex at a temperature gradient from 25 to 90 °C with 1 °C heating intervals. The excitation wavelength was λ=470 nm and fluorescence was recorded at λ=610 nm continuously during the measurement. The data evaluation was performed with GraphPad Prism7 and the values were normalised and fitted with a Boltzmann sigmoidal fit.

Table 6.20: Buffers used for TSA.

Buffer	Composition
1a	20 mM Tris, 50 mM NaCl, pH = 7.5
1b	20 mM Tris, 50 mM NaCl, 0.5 mM TCEP, pH = 7.5
2	20 mM HEPES, 50 mM NaCl, pH = 7.5
3	50 mM HEPES, 50 mM NaCl, pH = 7.5
4	20 mM HEPES, 50 mM NaCl, pH = 8
5	20 mM HEPES, 50 mM NaCl, pH = 7
6	20 mM HEPES, 50 mM NaCl, pH = 6.5
7	20 mM HEPES, 50 mM NaCl, pH = 7.5
8	20 mM HEPES, 100 mM NaCl, 0.5 mM TCEP, pH = 7.5
9	20 mM HEPES, 50 mM NaCl, 1 mM TCEP, pH = 7.5
10	20 mM HEPES, 50 mM NaCl, 5 mM $MgCl_2$, pH = 7.5
11	20 mM HEPES, 50 mM NaCl, 5% glycerol, pH = 7.5
12	20 mM HEPES, 50 mM NaCl, 10% glycerol, pH = 7.5
13	20 mM HEPES, 50 mM NaCl, 50 mM L-Arginine, pH = 7.5
14	20 mM HEPES, 50 mM NaCl, 5 µM **22**, pH = 7.5
15	20 mM HEPES, 50 mM NaCl, 5 µM FR-Core, pH = 7.5
16	20 mM MOPS, 50 mM NaCl, pH = 7.5
17	20 mM potassium phosphate, 50 mM NaCl, pH = 7.5
18	20 mM HEPES, 50 mM NaCl, 10% PEG4000, pH = 7.5
19	20 mM HEPES, 50 mM NaCl, 20% PEG4000, pH = 7.5
20	20 mM HEPES, 50 mM NaCl, 10% PEG8000, pH = 7.5
21	20 mM HEPES, 50 mM NaCl, 50 mM sodium citra,t pH = 7.5
22	20 mM HEPES, 50 mM NaCl, 50 mM ammonium sulfat, pH = 7.5

23	20 mM HEPES, 50 mM NaCl, 10% trehalose, pH = 7.5
24	20 mM HEPES, 50 mM NaCl, 40% trehalose, pH = 7.5
25	20 mM HEPES, 50 mM NaCl, 10% saccarose, pH = 7.5
26	20 mM HEPES, 50 mM NaCl, 50 µM FR-Core, pH = 7.5
27	20 mM HEPES, 50 mM NaCl, 100 µM FR-Core, pH = 7.5

6.8 Crystallisation trials

Gel-filtrated $FrsA_{TE}$ at 23 mg/ml was crystallized using the sitting-drop vapor diffusion crystallisation method. Prior to crystallisation, the protein was centrifuged at 21,000 g, 4 °C for 20 min. Droplets were dispensed using the Mosquito crystallisation robot (TTP labtech) by mixing 200 nl protein solution with 200 nl crystallisation mother liquor into MRC 2-well plates (SwissCI) at room temperature. The crystallisation plates were centrifuged for 1 min at 80 g and subsequently monitored over a period of 3 months using the RockImager robot (Formulatrix).

Small rod-shaped FrsATE *apo* crystals grew in a crystallisation buffer consisting of 0.1 M bis-tris propane pH 7.5, 0.2 M potassium thiocyanate, 4% tert-butanol and 20% PEG 3350 within 3-7 d.

In order to crystallize $FrsA_{TE}$ in complex with FR-Core, the protein solution was supplemented with a 2-fold molar excess of FR-Core dissolved to 50 mM in DMSO. This resulted in a solution consisting of 21 mg/ml $FrsA_{TE}$, 1.5 mM FR-Core, 3% DMSO and 90% of all gel-filtration buffer components.

Thin needles or needle clusters of $FrsA_{TE}$ + FR-Core grew within 28-42 days in a crystallisation mother liquor consisting of 0.1 M bis-tris propane pH 6.5, 0.1 – 0.2 M sodium iodide and 14% PEG 3350.

6.9 Chemical synthesis of precursors 21 and 22

The synthesis of **21** and **22** was performed as described in Hermes *et al*. in cooperation with Jim Küppers (AK Gütshow).[33] Thin-layer chromatography was carried out on Merck (Darmstadt, Germany) aluminium sheets, silica gel 60 F254. Detection was performed with UV light at 254 nm. Preparative column chromatography was performed on Merck silica gel (0.063-0.200 mm, 60 Å). Melting points were determined on a Büchi (Essen, Germany) 510 oil bath apparatus. ^{1}H NMR (500 MHz) and ^{13}C NMR (125 MHz) spectra were recorded on a Bruker Avance DRX 500. Chemical shifts δ are given in ppm referring to the signal centre using the solvent peaks for reference: DMSO-d_6 2.49/39.7 ppm. LC-MS analyses were carried out on an API2000 (Applied Biosystems, Darmstadt, Germany) mass spectrometer coupled to an Agilent (Santa Clara, CA, USA) 1100 LC system using an EC50/2 Nucleodur C18 Gravity column (Macherey-Nagel, Düren, Germany; 50 × 2.0 mm, particle size 3 µm). Purity of the compounds was determined using the diode array detector (DAD) of the LC-MS instrument between 200 and 400 nm. HRMS were recorded on a microTOF-Q (Bruker, Köln, Germany) mass spectrometer connected to a Dionex (Thermo Scientific, Braunschweig, Germany) Ultimate 3000 LC via an ESI interface using a Nucleodur C_{18} Gravity column (50 × 2.0 mm I.D., 3 µm, Macherey-Nagel, Düren, Germany).

6.9.1 (2*S*,3*R*)-3-Hydroxy-4-methyl-2-propionamidopentanoic acid (21)

A stirred solution of propanoic acid (370 mg, 5.00 mmol) and *N*-methylmorpholine (506 mg, 5.00 mmol) in THF (8.6 ml) was cooled to -10 °C. Isobutyl chloroformate (683 mg, 5.00 mmol) was added and the reaction was allowed to stir for 0.5 h. The temperature was adjusted to 0 °C, followed by treatment with (2*S*,3*R*)-2-amino-3-hydroxy-4-methylpentanoic acid (3-OH-leucine, **20** 1.10 g, 7.50 mmol, purchased from Iris Biotech GmbH, Marktredwitz, Germany) in 1 M NaOH (5.2 ml). After stirring the reaction for further 24 h at room temperature, the mixture was diluted with H_2O (16 ml) and washed with ethyl acetate (2 × 16 ml). The combined ethyl acetate layer was extracted with sat. aq. $NaHCO_3$ solution (3 × 16 ml). All aqueous layers were combined, adjusted to pH ~2 by adding 1 M HCl and extracted with ethyl acetate (3 × 32 ml). This combined organic layer was dried over Na_2SO_4, filtered and evaporated to dryness. The crude residue was purified by column chromatography on silica gel using a gradient of petroleum ether/ethyl acetate (1:1) + 1% AcOH to 100% ethyl acetate + 1% AcOH to give a white solid (762 mg, 75%); mp 104–106 °C.

^{1}H NMR (500 MHz, DMSO-d_6) δ 0.77 (d, ^3J = 6.7 Hz, 3H) and 0.90 (d, ^3J = 6.6 Hz, 3H, CH(CH_3)$_2$), 0.98 (t, ^3J = 7.6 Hz, 3H, CH_2CH_3), 1.50 – 1.58 (m, 1H, C*H*(CH$_3$)$_2$), 2.12 – 2.20 (m, 2H, CH_2), 3.50 (dd, ^3J = 8.7 Hz, ^3J = 2.8 Hz, 1H, C*H*-OH), 4.40 (dd, ^3J = 9.1 Hz, ^3J = 2.8 Hz, 1H, C*H*-NH), 7.53 (d, ^3J = 9.1 Hz, 1H, NH). Two proton signals (CO_2H, OH) do not appear. ^{13}C NMR (125 MHz, DMSO-d_6) δ 10.08 (CH_2CH_3), 19.12, 19.21 (CH(*C*H$_3$)$_2$), 28.48 (*C*H(CH$_3$)$_2$), 30.92 (CH_2CH_3), 54.54 (CHNH), 76.18 (CHOH), 173.09, 173.30 (CO_2H, CONH). LC-MS (ESI) (90% H_2O to 100% MeOH in 10 min, then 100% MeOH for 10 min, DAD 196–400 nm), *m/z* = 204.0 ([M+H]$^+$).

6.9.2 (2*S*,3*R*)-*S*-2-Acetamidoethyl-3-hydroxy-4-methyl-2-propionamidopentanethioate (22)

(2*S*,3*R*)-3-Hydroxy-4-methyl-2-propionamidopentanoic acid (**5**, 610 mg, 3.00 mmol) was dissolved in anhydrous MeCN (120 ml) under nitrogen atmosphere. A solution of DCC (650 mg, 3.15 mmol) and HOBt × H_2O (482 mg, 3.15 mmol) in MeCN (120 ml) was slowly added, followed by *N*-acetylcysteamine (375 mg, 3.15 mmol). The reaction mixture was stirred at room temperature for 24 h. Subsequently, the urea was filtered off and the filtrate was evaporated to dryness. The crude residue was purified by preparative column chromatography using ethyl acetate/MeOH (9:1) as eluent to obtain clear oil (82 mg, 9%).

^{1}H NMR (500 MHz, DMSO-d_6) δ 0.74 (d, ^3J = 6.7 Hz, 3H) and 0.90 (d, ^3J = 6.6 Hz, 3H, CH(CH_3)$_2$), 1.02 (t, ^3J = 7.6 Hz, 3H, CH_2CH_3), 1.53 – 1.61 (m, 1H, C*H*(CH$_3$)$_2$), 1.78 (s, 3H, $COCH_3$), 2.26 (q, ^3J = 7.6 Hz, 2H, CH_2CH$_3$), 2.79 – 2.89 (m, 2H, CH_2S), 3.08 – 3.19 (m, 2H, NHCH_2), 3.59 (ddd, ^3J = 9.1 Hz, ^3J = 6.9 Hz, ^3J = 2.4 Hz, 1H, C*H*OH), 4.54 (dd, ^3J = 8.9 Hz, ^3J = 2.4 Hz, 1H, C*H*NH), 5.01 (d, ^3J = 6.9 Hz, 1H, OH), 7.98 (t, ^3J = 5.7 Hz, 1H, N*H*CH$_2$), 8.08 (d, ^3J = 8.9 Hz, 1H, CHN*H*). ^{13}C NMR (125 MHz, DMSO-d_6) δ 9.90 (CH_2CH_3), 18.83, 19.17 (CH(*C*H$_3$)$_2$), 22.66 (*C*H$_3$CO), 27.99 (SCH_2), 28.39 (*C*H(CH$_3$)$_2$), 30.99 (*C*H$_2$CH$_3$), 38.24 (NHCH$_2$), 61.53 (CHNH), 76.00 (CHOH), 169.44 (H$_3$C*C*ONH), 174.03 (CH_2*C*ONH), 201.86 (COS). LC-MS (ESI) (90% H_2O to 100% MeOH in 10 min, then 100%

MeOH for 10 min, DAD 220–400 nm), *m/z* = 305.0 ($[M+H]^+$). HRMS, calcd. for $C_{13}H_{24}N_2O_4S$: $[M+H]^+$ *m/z* 305.1530; found: 305.1544.

6.10 Isolation of precursors

6.10.1 Cultivation and extraction of *C. vaccinii*

The cultivation of *Chromobacterium vaccinii* MWU205, *C. vaccinii* Δ*frsA* and *C. vaccinii* Δ*frsA*/Δ*vioA* was performed in two different media, LB and M9 minimal medium. The *C. vaccinii* strains were grown in each medium, supplemented with 50 µg/ml carbenicillin. The cultivation was performed at 25 °C and 200 rpm for 2 days. For the isolation of compounds, 4 L of culture were extracted with 4 L *n*-butanol by shaking for 12 h and subsequent centrifugation (1,860 g, 5 min).

6.10.2 Isolation of FR-Core

For the isolation of FR-Core (**8**), the crude material, butanoic extract from 4 L culture of *C. vaccinii* Δ*frsA* or *C. vaccinii* Δ*frsA*/Δ*vioA* (see 6.10.1), was fractionated on a Grace Reveleris X2 flash chromatography system with integrated evaporative light scattering (ELSD) and UV-Vis detection via a Reveleris C_{18} flash column (220 g, 40 µm). A stepwise gradient solvent system of increasing polarity and a flow rate of 65 ml/min was used starting with 50/50 H_2O/MeOH for 13 min, then changing to 30/70 H_2O/MeOH within 1 min and hold again for 13 min. The gradient was then changed within 1 min to 25/75 H_2O/MeOH and held for 25.0 min, then within 1 min to 20/80 H_2O/MeOH, held for 13 min, then within 1 min to 15/85 H_2O/MeOH and held for 25 min. Finally, the gradient was changed within 1 min to 100% MeOH and held for an additional 10 min. According to the measured ELSD and UV signals, an FR-Core containing fraction was collected at 70 min. Final purification was done by HPLC with a semi-preparative Macherey-Nagel Nucleodur C_{18} column (250 x 8 mm, 5 µm) using isocratic elution with 19/81 H_2O/MeOH (flow 2.0 ml/min). HPLC was carried out using a Waters HPLC system, controlled by Waters Millennium software, 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. Pure FR-Core was isolated as a white powder (t_R: 12 min, 5 mg).

6.10.3 Isolation of FR-5

After the feeding experiments (see 6.11.4), for the isolation of FR-5 (**19**), M9 medium was supplemented with 20 mM butyric acid and *C. vaccinii* was cultivated 2 d as described above. The crude material (butanoic extract from 4.5 L culture of *C. vaccinii*) was fractionated as described in 6.10.2. A **19** containing fraction was collected at around 80 min. Final purification was done as described in 6.10.2. Pure **19** was isolated as a white powder (t_R: 25 min, 10 mg).

6.10.4 Structure Elucidation of FR-Core and FR-5

NMR spectra were recorded either on a Bruker Ascend 600 NMR spectrometer with a Prodigy cryoprobe operating at 600 MHz (^{1}H) and 150 MHz (^{13}C) using acetonitrile-d_3 as solvent (Deutero GmbH, 99.8% D) or on a Bruker Avance 300 DPX NMR spectrometer at 300 MHz (^{1}H) and 75 MHz (^{13}C) using $CDCl_3$ as solvent (Deutero GmbH, 99.8% D). NMR spectra were processed using Bruker TopspinVersion 1.3 and MestReNova 8.0.1 software packages. Spectra were referenced to residual solvent signals with resonances at $\delta_{H/C}$ 1.93/117.7 for acetonitrile-d_3 or $\delta_{H/C}$ 7.26/77.0 for $CDCl_3$.

6.11 *In vitro* enzyme assays

6.11.1 γ-$^{18}O_4$-ATP-Exchange Assay for A domains

The γ-$^{18}O_4$-ATP-Exchange Assay of Phelan *et al*. was developed to test the potential of A domains to activate certain amino acids. [144] The solutions used for this assay are listed in Table 6.21. For each tested amino assay, 2 µl of each solution 1, 2 and 3 was mixed and incubated 1 h at room temperature. Then, 6 µl stop solution were added and the denatured protein centrifuged to get a clear solution for MALDI-TOF-MS analysis. As a negative control, an assay containing solution 2 without an amino acid with only 15 mM pyrophosphate was used.

The sample was analysed with a MALDI-TOF spectrometer Brucker Autoflex III (Brucker Daltonik, Bremen). Therefore, 1 µl of the matrix/assay solution was transferred to the ground steel sample carrier and dried on air. The measurement was conducted in negative mode by averaging the signal of 2000 laser pulses in an *m/z* range of 300-600 *m/z*. The data was processed with FlexAnalysis 3.3 and analysed with FlexControl 3.3 software. Absolute substrate conversion was calculated by dividing the peak area at 506 Da through the combined 506, 508, 510, 512, and 514 Da peak areas. Divided by 83.33 for the molar ratio of the equilibrium between labelled and unlabelled PPi, the % exchange was calculated.

Table 6.21: Used solutions for the γ-$^{18}O_4$-ATP-exchange-assay.

Solution	Composition
Solution 1	3 mM amino acid, 15 mM pyrophosphate, 20 mM Tris-HCl pH 7.5
Solution 2	3 mM γ-$^{18}O_4$-ATP, 15 mM $MgCl_2$; 2 mM Tris-HCl pH 7.5
Solution 3	5 mM enzyme in 20 mM Tris-HCl pH 7.5, 5% glycerol
Stop solution	10 mg/ml 9-aminoacridine in actetone

6.11.2 $C_{starter}$ domain and hydroxylation assay

The $C_{starter}$ domain assay was developed to prove the function of the C domain of FrsA and FrsD, as well as the hydroxylating function of FrsH. Therefore, phosphopantetheinylated FrsA or FrsD was expressed in *E. coli* BAP1 to directly generate the *holo*-form of the T domain,[141] isolated as described above and buffered in C domain assay buffer (see Table 6.5) using PD-10 columns (GE) following the

manufacturer's protocol. The assays were performed in a one-pot-reaction: Purified FrsH was reduced with a 5-fold concentration of $Na_2S_2O_4$ and equimolar concentration of methylviologen by mixing and incubation on ice for 5 min. To activate the A domain, 1 mM ATP and 1 mM of L-leucine were added. As substrate for the C domain, 500 µM propionyl-S-CoA (CoALA Biosciences, Austin, USA) or acetyl-S-CoA (Sigma) was added. The protein solution of FrsA/FrsD (500 µl, 10 µM) was mixed with 500 µl of the activated FrsH, resulting concentration of the proteins were 5 µM FrsA/FrsD and 25 µM FrsH. The assay was incubated at 20 °C for 3 h. The reaction was stopped by adding 10% (w/v) TCA and incubated on ice for 30 min. Precipitated proteins were pelleted and washed twice with assay buffer. Alkaline hydrolysis was performed with 0.1 M KOH at 70 °C for 20 min. The solution was lyophilized, and the pellet solved in a minimal volume of MeOH for LC-MS analysis (6.12.1). For the negative control, the proteins were heat-inactivated at 80 °C for 20 min before the addition of the substrates.

6.11.3 Transesterification assay

To test the activity of the TE domain, two different setups were used. The first using the synthesized precursor **22**, the second assembling the side chain *in vitro*. For the first attempt, the substrate FR-Core (**8**) was dissolved in MeOH to a final concentration of 50 mM, and substrate **6** was dissolved in MeOH to a final concentration of 100 mM. The enzymatic assays were conducted in C domain buffer with 2 mM **22** and 0.5 mM FR-Core in a total volume of 1 ml. Reactions were initiated by the addition of $FrsA_{TE}$ (5 µM). The assays were incubated at 22 °C for 5 h and then extracted three times with 0.75 ml CH_2Cl_2. The organic phase was evaporated, and the pellet resuspended in 50 µL MeOH for LC-MS analysis, see 6.12.1. Alternatively, for the *in vitro* assay, FrsA and FrsH were used with L-leucine and propionyl-S-CoA to form the side-chain under the same conditions as in the $C_{starter}$ assay. Additionally, 0.5 mM FR-Core was added and the reaction was incubated and processed like the TE domain assay described above. In this assay mixture, the substrates were eventually varied to D-leucine and L-isoleucine for the amino acid and to acetyl-S-CoA (Sigma) and butyryl-S-CoA (CoALA Biosciences, Austin, USA) for the C domain substrate. For the negative control, FrsH and/or FrsA/$FrsA_{TE}$ were heat-inactivated at 80 °C for 20 min prior to the addition of the substrates.

6.11.4 Precursor feeding experiments

To investigate if *C. vaccinii* wt can use and incorporate different building blocks in FR biosynthesis, feeding of precursors in minimal medium was tested. Therefore, a preculture of *C. vaccinii* in M9 minimal medium was inoculated with a colony from a fresh LB agar plate and cultivated for 24 h at 25 °C and 200 rpm. Then three flasks with 40 ml M9 medium supplemented with 20 mM sodium butyrate and 50 µg/ml carbenicillin were prepared as biological triplicates and inoculated with the preculture. As blank and a negative control one flask with M9 medium with supplements but without bacteria and one flask M9 without butyrate but with bacteria was prepared each. All flasks were cultivated for 48 h at 25°C and 200 rpm before extraction with 40 ml *n*-butanol per flask overnight at 160 rpm. The phases were separated by centrifugation (1,860 g, 5 min) and the butanol fraction

evaporated to dryness. A 1 mg/ml MeOH solution of every extract was submitted to LC-MS analysis, see 6.12.2.

6.12 HPLC and MS analysis

6.12.1 HPLC-HR-MS/MS Analysis

Mass spectra were recorded on a micrOTOF-QII mass spectrometer (Bruker) with ESI-source coupled with an HPLC Dionex Ultimate 3000 (Thermo Scientific) using an EC10/2 Nucleoshell C_{18} 2.7 µm column (Macherey-Nagel). The column temperature was 25 °C. MS data were acquired over a range from 100-3,000 *m/z* in positive mode. Auto MS/MS fragmentation was achieved with rising collision energy (35-50 keV over a gradient from 500-2000 *m/z*) with a frequency of 4 Hz for all ions over a threshold of 100. HPLC begins with 90% H_2O containing 0.1% acetic acid. The gradient starts after 1 min to 100% acetonitrile (0.1% acetic acid) in 20 min. 5 µl of a 1 mg/ml sample solution (MeOH) was injected to a flow of 0.3 ml/min.

6.12.2 HPLC-MS analysis

For HPLC-MS analysis without fragmentation a Waters 2695 Separation Module was used, coupled to a Waters 996 Photodiode Array detector, Waters QDa detector and ESI. A Waters X Bridge Shield RP column with 2.1 x 100 mm and particle size 3.5 µm at 25 °C was used. The composition of the eluents is listed in Table 6.5, the gradient in Table 6.22, the flow was 0.3 ml/min and 5 µl sample were injected. MS data were collected in positive and negative mode between 140-1250 *m/z* and some additional mass traces were measured.

Table 6.22: Gradient for HPLC-MS analysis.

Time [min]	Flow [ml/min]	Eluent A [%]	Eluent B [%]
0.01	0.30	80.0	20.0
20.00	0.30	0.0	100.0
30.00	0.30	0.0	100.0
31.00	0.30	80.0	20.0
45.00	0.30	80.0	20.0

6.13 Bioinformatics

6.13.1 Alignments

The sequence alignments of *frsA, frsD,* $FrsA_{TE}$ and $FrsG_{TE}$ and all alignments of "*Ca*. B. crenata" and *C.vaccinii frs* genes and proteins were performed with the online platform EMBOSS needle.[131]

6.13.2 Structural models

The 3D structural models of the TE domains were calculated by the I-TASSER server[209,210] and aligned with the Swiss-PDB-viewer.[211]

6.13.3 Phylogenetic tree of $C_{starter}$ and TE domains

For the phylogenetic trees of C and TE domains, protein sequences were retrieved from public databases (for source organisms and accession numbers see Table 6.23 and Table 6.24) and aligned using the MUSCLE algorithm[212] in MEGA 10.[213] In case of multi-domain proteins, only the C, resp. TE domain was analysed. With this program maximum likelihood trees were calculated as follows: The evolutionary history was inferred by using the Maximum Likelihood method and Le_Gascuel_2008 model.[214] The tree with the highest log likelihood (C: -58208.75, TE: -8887.94) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter C= 1.6801, TE = 1.9188)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], C: 0.6%, TE: 1.14% sites). The trees are drawn to scale, with branch lengths measured in the number of substitutions per site. C domain analysis involved 139; TE analysis 39 amino acid sequences. For C domain analysis, all positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). For TE analysis, all positions containing gaps and missing data were eliminated (complete deletion option). There were a total of 409 (C) and 132 (TE) positions in the final datasets.

Table 6.23: Databank accession numbers and source organisms of protein sequences used for phylogenetic analyses of $C_{starter}$ domains. Related to Figure 4.3 and Supplementary Figure 9.23.

Acceccion	Organism	Acceccion	Organism
MT876545	*Chromobacterium vaccinii* MWU205	WP_084480120.1	*Methylosarcina lacus*
MT876545	*Chromobacterium vaccinii* MWU205	WP_078996481.1	*Lysobacter enzymogenes*
KNE75171.1	*Candidatus* Burkholderia crenata	WP_096414584.1	*Lysobacter capsici*
KNE75168.1	*Candidatus* Burkholderia crenata	WP_119629200.1	*Methylocaldum marinum*
WP_162850377.1	*Rhodococcus jostii*	WP_174932260.1	*Burkholderia lata*
WP_007299123.1	*Rhodococcus imtechensis*	WP_120026908.1	*Amycolatopsis panacis*
RZK84214.1	*Rhodococcus sp.*	WP_054261810.1	*Propionispora sp*. 2/2-37
WP_169695704.1	*Rhodococcus opacus*	WP_091794482.1	*Lysobacter sp.* yr284
ELB93997.1	*Rhodococcus wratislaviensis* IFP 2016	WP_040376786.1	*Peribacillus psychrosaccharolyticus*
TMK21117.1	*Alphaproteobacteria bacterium*	WP_138885197.1	*Lysobacter enzymogenes*
TMJ40383.1	*Alphaproteobacteria bacterium*	WP_052756236.1	*Lysobacter capsici*
WP_036714060.1	*Paenibacillus ehimensis*	KJS54231.1	*Streptomyces rubellomurinus subsp. indigoferus*
WP_176165338.1	*Streptomyces sp.* NA00687	WP_166530797.1	*Agaribacter marinus*
WP_106046991.1	*Bacillus atrophaeus*	TQF01515.1	*Kitasatospora sp.* MMS16-CNU292

WP_183596885.1	*Paenibacillus phyllosphaerae*	WP_057921675.1	*Lysobacter capsici*
NQD50526.1	*Bacillus altitudinis*	WP_049786543.1	*Mycetohabitans rhizoxinica*
WP_156746001.1	*Mycobacterium sp.* 1423905.2	CBW76866.1	*Mycetohabitans rhizoxinica* HKI 454
WP_186229549.1	*Burkholderia gladioli*	WP_146064026.1	*Mycetohabitans endofungorum*
OKH99041.1	*Streptomyces sp.* CB02923	WP_174993913.1	*Burkholderia arboris*
WP_157638970.1	*Burkholderia ubonensis*	OPD64588.1	*Lysobacter enzymogenes*
WP_161220653.1	*Streptomyces sp.* SID6137	WP_156123907.1	*Paraburkholderia mimosarum*
WP_155634472.1	*Burkholderia stagnalis*	WP_081052109.1	*Burkholderia cepacia*
REK66978.1	*Brevibacillus sp.*	WP_026869105.1	*Inquilinus limosus*
APD71879.1	*Streptomyces sp.*	WP_143200560.1	*Kitasatospora sp.* CB01950
TDL85614.1	*Vibrio vulnificus*	SFN60854.1	*Actinomadura madurae*
WP_171565203.1	*Brevibacillus sp.* MCWH	WP_051213701.1	*Glomeribacter sp.* 1016415
WP_090670197.1	*Paenibacillus tianmuensis*	WP_067343588.1	*Marinomonas spartinae*
WP_139641914.1	*Streptomyces sedi*	WP_169750257.1	*Streptosporangium amethystogenes*
PBC50623.1	*Rhodococcus sp.* ACS1	WP_092071365.1	*Dendrosporobacter quercicolus*
WP_161218580.1	*Streptomyces sp.* SID6139	WP_045303061.1	*Saccharothrix sp.* ST-888
WP_135342059.1	*Streptomyces palmae*	WP_181556777.1	*Anoxybacillus caldiproteolyticus*
WP_155754571.1	*Burkholderia stagnalis*	WP_034840793.1	*Inquilinus limosus*
WP_161783329.1	*Burkholderia sp.* A1	WP_157441225.1	*Actinoplanes awajinensis*
WP_187438659.1	*Streptomyces sp.* sk2.1	WP_172236540.1	*Bradyrhizobium sp*. LMG 8443
AEA62641.1	*Burkholderia gladioli* BSR3	WP_143956941.1	*Mycobacterium sp*. KBS0706
SEC10944.1	*Rhodococcus koreensis*	WP_165781571.1	*Streptosporangium minutum*
WP_158315302.1	*Bacillus megaterium*	WP_176072366.1	*Paraburkholderia mimosarum*
WP_072274058.1	*Peribacillus simplex*	WP_056114189.1	*Lysobacter sp.* Root690
PAE69549.1	*Bacillus subtilis*	WP_165940603.1	*Burkholderia sp.* SRS-46
WP_120709248.1	*Rhizobium jaguaris*	WP_083780791.1	*Bradyrhizobium sp.* BTAi1
WP_094237092.1	*Tumebacillus algifaecis*	WP_067015022.1	*Marinomonas spartinae*
QAR15116.1	*Streptomyces costaricanus*	PMS18374.1	*Burkholderia dabaoshanensis*
WP_104825816.1	*Rhizobium sp.* NXC24	WP_162791553.1	*Dyella sp.* L4-6
BBA33607.1	*Methylocaldum marinum*	WP_103564835.1	*Actinomadura sp.* RB29
WP_090514810.1	*Paenibacillus sp. cl6col*	RDD80443.1	*Dyella sp.* L4-6
WP_094237092.1	*Tumebacillus algifaecis*	WP_146014032.1	*Burkholderia dabaoshanensis*
WP_098680385.1	*Bacillus altitudinis*	WP_132240709.1	*Micromonospora sp.* CNZ303
WP_081114287.1	*Bacillus stratosphericus*	WP_091621489.1	*Micromonospora peucetia*
WP_068017976.1	*Rhodoplanes sp.* Z2-YC6860	WP_093292387.1	*Thermoactinomyces sp.* DSM 45892
WP_181799076.1	*Kitasatospora sp.* MMS16-CNU292	WP_184260366.1	*Granulicella mallensis*
WP_061420802.1	*Bacillus pumilus*	WP_067109092.1	*Mycobacterium sp.* 852002-51057_SCH5723018
WP_125058029.1	*Streptomyces rimosus*	WP_110573648.1	*Marinomonas alcarazii*
WP_128788709.1	*Streptomyces sp. endophyte*_N2	WP_033674841.1	*Bacillus gaemokensis*
PYS27789.1	*Acidobacteria bacterium*	WP_172427363.1	*Streptomyces griseofuscus*
WP_048411362.1	*Chromobacterium sp.* LK1	WP_157419571.1	*Actinomadura kijaniata*
WP_099396555.1	*Iodobacter sp.* BJB302	WP_186145144.1	*Burkholderia gladioli*
AXE32757.1	*Chromobacterium sp.* IIBBL 112-1	WP_007952768.1	*Pelosinus fermentans*
WP_082113610.1	*Chromobacterium vaccinii*	WP_136371369.1	*Cohnella sp.* CC-MHH1044
WP_122983744.1	*Chromobacterium subtsugae*	WP_076788272.1	*Burkholderia sp.* b13
KZE85028.1	*Chromobacterium sp.* F49	WP_146012759.1	*Trinickia caryophylli*
SUX53575.1	*Chromobacterium vaccinii*	WP_146064107.1	*Mycetohabitans endofungorum*
OVE47519.1	*Chromobacterium violaceum*	WP_142401196.1	*Mycobacterium marinum*
WP_081545507.1	*Chromobacterium haemolyticum*	WP_067015692.1	*Mycobacterium sp.* 1081908.1
WP_118266927.1	*Chromobacterium rhizoryzae*	WP_052407967.1	*Allokutzneria albata*
WP_106076243.1	*Chromobacterium amazonense*	MPZ86334.1	*Actinophytocola sp.*
WP_107800191.1	*Chromobacterium sp. Panama*	WP_141996174.1	*Amycolatopsis cihanbeyliensis*
WP_156746001.1	*Mycobacterium sp.* 1423905.2	KAK48742.1	*Caballeronia jiangsuensis*
WP_175879783.1	*Mycobacterium sp.* IS-2888	WP_099661199.1	*Sporosarcina sp.* P29

WP_172880913.1	*Bacillus safensis*	WP_066488029.1	*Burkholderia sp.* BDU8
WP_097231083.1	*Streptomyces zhaozhouensis*	WP_073604292.1	*Vibrio aerogenes*
WP_053416733.1	*Viridibacillus arvi*	SAL02479.1	*Caballeronia arationis*
WP_143665067.1	*Streptomyces cacaoi*	NVK71858.1	*Oceanospirillaceae bacterium*
WP_123647243.1	*Lysobacter enzymogenes*	WP_061162849.1	*Caballeronia temeraria*
WP_035299007.1	*Brevibacillus thermoruber*	SLC21331.1	*Mycobacteroides abscessus subsp. massiliense*
WP_076794250.1	*Burkholderia sp.* b14		*Methylosarcina lacus*

Table 6.24: Names, databank accession numbers and source organisms of protein sequences used for phylogenetic analyses of TE domains.

Name	Acceccion	Organism
BcFrsA-TE	KNE75171.1	*Candidatus* Burkholderia crenata
BcFrsG-TE	KNE75165.1	*Candidatus* Burkholderia crenata
cvFrsA-TE	MT876545	*Chromobacterium vaccinii* MWU205
cvFrsG-TE	MT876545	*Chromobacterium vaccinii* MWU205
DhbF - TE	WP_144530663.1	*Bacillus subtilis* ATCC 21332
EntF - TE	AYG20286.1	*Escherichia coli str.* K-12 *substr*. MG1655
ERYA3_TE	Q03133	*Saccharopolyspora erythraea*
FtdB - TE	ADJ54381.1	*Streptomyces sp.* SPB78
HbnA	ALV82446.1	*Streptomyces variabili*
HSAF - TE	ABL86391.1	*Lysobacter enzymogenes*
IkaA - TE	AJD77023.1	*Streptomyces sp.* ZJ306
KirHI	AGS67165.1	*Streptomyces collinus* Tu 365
KSE_70420	BAJ32800	*Kitasatospora setae* KM-6054
LipX2	ABB05100	*Streptomyces aureofaciens*
LybB - TE1	AEH59100.1	*Lysobacter sp.* ATCC 53042
LybB - TE2	AEH59100.1	*Lysobacter sp.* ATCC 53042
MassC - TE1	EIK63041.1	*Pseudomonas fluorescens* SS101
MassC - TE2	EIK63041.1	*Pseudomonas fluorescens* SS101
NocB - TE	AAT09805.1	*Nocardia uniformis subsp. tsuyamanensis*
ObiF - TE	KX134687.1	*Pseudomonas fluorescens* ATCC 39502
Pys-Pent - TE	WP_011533552	*Pseudomonas entomophila*
Pys-Pflu -TE	WP_064118616	*Pseudomonas fluorescens* HKI 0770
RomH	WP_078586793.1	*Streptomyces rimosus* NRRL B-2659
RomI	WP_004571777.1	*Streptomyces rimosus* NRRL B-2659
SGR814 - TE	AGK81502	*Streptomyces fulvissimus* DSM 40593
SlgL	CBA11558	*Streptomyces lydicus*
Sln6 - TE	DAB41476.1	*Streptomyces sp.* CNB-091
Sln9 - TE	DAB41479.1	*Streptomyces sp.* CNB-091
SrfA-C - TE	2VSQ	*Bacillus subtilis* ATCC 21332
SSHG - TE	EFE85271.1	*Streptomyces albus* J1074
SwrW - TE	I7GF64	*Serratia marcescens*
TaaE - TE1	CCJ67640.1	*Pseudomonas costantinii* DSM 16734
TaaE - TE2	CCJ67640.1	*Pseudomonas costantinii*
TrdC	ADY38535.1	*Streptomyces sp.* SCSIO 1666
TycC - TE	AAC45930.1	*Brevibacillus brevis*
ViscC - TE1	CAY48789.1	*Pseudomonas fluorescens* SBW25

ViscC - TE2	CAY48789.1	*Pseudomonas fluorescens* SBW25
WlipC - TE1	AFJ23826.1	*Pseudomonas putida*
WlipC - TE2	AFJ23826.1	*Pseudomonas putida*

7 References

1 R. Schrage, Schmitz A.-L., E. Gaffal, S. Annala, S. Kehraus, D. K. Wenzel, G. M. König and E. Kostenis, *Nature Communications*, 2015, **6**, 10156.

2 A. Nishimura, K. Kitano, J. Takasaki, M. Taniguchi, N. Mizuno, K. Tago, T. Hakoshima and H. Itoh, *Proceedings of the National Academy of Sciences of the United States of America*, 2010, **107**, 13666–13671.

3 J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigó, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S.

Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh and X. Zhu, *Science (New York, N.Y.)*, 2001, **291**, 1304–1351.

4 D. Calebiro and Z. Koszegi, *Molecular and Cellular Endocrinology*, 2019, **483**, 24–30.

5 D. Hilger, M. Masureel and B. K. Kobilka, *Nature Structural & Molecular Biology*, 2018, **25**, 4–12.

6 D. Tietze, D. Kaufmann, A. A. Tietze, A. Voll, R. Reher, G. König and F. Hausch, *Journal of Chemical Information and Modeling*, 2019, **59**, 4361–4373.

7 A. P. Campbell and A. V. Smrcka, *Nature Reviews. Drug Discovery*, 2018, **17**, 789–803.

8 A. S. Hauser, M. M. Attwood, M. Rask-Andersen, H. B. Schiöth and D. E. Gloriam, *Nature Reviews. Drug Discovery*, 2017, **16**, 829–842.

9 E. Kostenis, E. M. Pfeil and S. Annala, *The Journal of Biological Chemistry*, 2020, **295**, 5206–5215.

10 S. Mangmool and H. Kurose, *Toxins*, 2011, **3**, 884–899.

11 R. Reher, T. Kühl, 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 and D. Imhof, *ChemMedChem*, 2018, **13**, 1634–1643.

12 H. Zhang, A. L. Nielsen and K. Strømgaard, *Medicinal Research Reviews*, 2020, **40**, 135–157.

13 M. Fujioka, S. Koda, Y. Morimoto and K. Biemann, *The Journal of Organic Chemistry*, 1988, **53**, 2820–2825.

14 J. Takasaki, T. Saito, M. Taniguchi, T. Kawasaki, Y. Moritani, K. Hayashi and M. Kobori, *The Journal of Biological Chemistry*, 2004, **279**, 47438–47445.

15 A. Nesterov, M. Hong, C. Hertel, P. Jiao, L. Brownell and E. Cannon, *Journal of Biomolecular Screening*, 2010, **15**, 518–527.

16 X.-F. Xiong, H. Zhang, C. R. Underwood, K. Harpsøe, T. J. Gardella, M. F. Wöldike, M. Mannstadt, D. E. Gloriam, H. Bräuner-Osborne and K. Strømgaard, *Nature Chemistry*, 2016, **8**, 1035–1041.

17 A. Miyamae, M. Fujioka, S. Koda and Y. Morimoto, *Journal of the Chemical Society, Perkin Transactions 1*, 1989, 873.

18 K. Zaima, J. Deguchi, Y. Matsuno, T. Kaneda, Y. Hirasawa and H. Morita, *Journal of Natural Medicines*, 2013, **67**, 196–201.

19 K. Machida, D. Arai, R. Katsumata, S. Otsuka, J. K. Yamashita, T. Ye, S. Tang, N. Fusetani and Y. Nakao, *Bioorganic & Medicinal Chemistry*, 2018, **26**, 3852–3857.

20 Y. Hayakawa, M. Akimoto, A. Ishikawa, M. Izawa and K. Shin-ya, *The Journal of Antibiotics*, 2016, **69**, 187–188.

21 M. Taniguchi, K.-I. Suzumura, K. Nagai, T. Kawasaki, J. Takasaki, M. Sekiguchi, Y. Moritani, T. Saito, K. Hayashi, S. Fujita, S.-I. Tsukamoto and K.-I. Suzuki, *Bioorganic & Medicinal Chemistry*, 2004, **12**, 3125–3133.

22 K. Kitajima, A. M. Fox, T. Sato and D. Nagamatsu, *Biological Invasions*, 2006, **8**, 1471–1482.

23 L. Ma, W. Li, H. Wang, X. Kuang, Q. Li, Y. Wang, P. Xie and K. Koike, *Journal of Pharmaceutical and Biomedical Analysis*, 2015, **102**, 400–408.

24 D. Lapadula, E. Farias, C. E. Randolph, T. J. Purwin, D. McGrath, T. H. Charpentier, L. Zhang, S. Wu, M. Terai, T. Sato, G. G. Tall, N. Zhou, P. B. Wedegaertner, A. E. Aplin, J. Aguirre-Ghiso and J. L. Benovic, *Molecular Cancer Research*, 2019, **17**, 963–973.

25 Z. Jia, K. Koike, T. Ohmoto and M. Ni, *Phytochemistry*, 1994, **37**, 1389–1396.

26 M. Taniguchi, K.-I. Suzumura, K. Nagai, T. Kawasaki, T. Saito, J. Takasaki, K.-I. Suzuki, S. Fujita and S.-I. Tsukamoto, *Tetrahedron*, 2003, **59**, 4533–4538.

27 M. Taniguchi, K. Nagai, N. Arao, T. Kawasaki, T. Saito, Y. Moritani, J. Takasaki, K. Hayashi, S. Fujita, K.-I. Suzuki and S.-I. Tsukamoto, *The Journal of Antibiotics*, 2003, **56**, 358–363.

28 I. M. Miller, in *Advances in Botanical Research*, ed. J. A. Callow, Academic Press, 1990, vol. 17, pp. 163–234.

29 B. Lemaire, E. Smets and S. Dessein, *Research in microbiology*, 2011, **162**, 528–534.

30 A. Carlier, L. Fehr, M. Pinto-Carbó, T. Schäberle, R. Reher, S. Dessein, G. König and L. Eberl, *Environmental Microbiology*, 2016, **18**, 2507–2522.

31 M. Pinto-Carbó, K. Gademann, L. Eberl and A. Carlier, *Current Opinion in Plant Biology*, 2018, **44**, 23–31.

32 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 and G. M. König, *Angewandte Chemie (International ed. in English)*, 2018, **57**, 836–840.

33 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. Asuka Inoue, P. Mergaert, M. Gütschow, E. Kostenis, G. M. König and M. Crüsemann, *Nature Communications*, 2020, accepted.

34 S. D. Soby, S. R. Gadagkar, C. Contreras and F. L. Caruso, *International Journal of Systematic and Evolutionary Microbiology*, 2013, **63**, 1840–1846.

35 K. Vöing, A. Harrison and S. D. Soby, *Genome Announcements*, 2015, **3**.

36 R. Reher, M. Kuschak, N. Heycke, S. Annala, S. Kehraus, H.-F. Dai, C. E. Müller, E. Kostenis, G. M. König and M. Crüsemann, *Journal of Natural Products*, 2018, **81**, 1628–1635.

37 M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W.-T. Liu, M. Crüsemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C.-C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrewe, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C.-C. Liaw, Y.-L. Yang, H.-U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio,

A. Klitgaard, C. B. Larson, C. A. B. P, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Dai, R. Neupane, J. Gurr, A. M. C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P.-M. Allard, P. Phapale, L.-F. Nothias, T. Alexandrov, M. Litaudon, J.-L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D.-T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. O. Palsson, K. Pogliano, R. G. Linington, M. Gutiérrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein and N. Bandeira, *Nature Biotechnology*, 2016, **34**, 828–837.

38 R. D. Süssmuth and A. Mainz, *Angewandte Chemie (International ed. in English)*, 2017, **56**, 3770–3821.

39 A. Stanišić and H. Kries, *ChemBioChem: a European Journal of Chemical Biology*, 2019, **20**, 1347–1356.

40 M. E. Horsman, T. P. A. Hari and C. N. Boddy, *Natural Product Reports*, 2016, **33**, 183–202.

41 T. M. Makris, M. Chakrabarti, E. Munck and J. D. Lipscomb, *Proceedings of the National Academy of Sciences of the United States of America*, 2010, **107**, 15391–15396.

42 D. A. Herbst, B. Boll, G. Zocher, T. Stehle and L. Heide, *The Journal of Biological Chemistry*, 2013, **288**, 1991–2003.

43 L. Ray, K. Yamanaka and B. S. Moore, *Angewandte Chemie (International ed. in English)*, 2016, **55**, 364–367.

44 C. Rausch, I. Hoof, T. Weber, W. Wohlleben and D. H. Huson, *BMC Evolutionary Biology*, 2007, **7**, 78.

45 T. Stachelhaus, H. D. Mootz and M. A. Marahiel, *Chemistry & Biology*, 1999, **6**, 493–505.

46 N. M. Gaudelli, D. H. Long and C. A. Townsend, *Nature*, 2015, **520**, 383–387.

47 T. A. Lundy, S. Mori and S. Garneau-Tsodikova, *RSC Chemical Biology*, 2020, **1**, 233–250.

48 D. T. Rensing, S. Uppal, K. J. Blumer and K. D. Moeller, *Organic Letters*, 2015, **17**, 2270–2273.

49 H. Kaur, P. W. R. Harris, P. J. Little and M. A. Brimble, *Organic Letters*, 2015, **17**, 492–495.

50 Y. Trusov and J. R. Botella, *Frontiers in Chemistry*, 2016, **4**, 24.

51 S. A. Kautsar, K. Blin, S. Shaw, T. Weber and M. H. Medema, *Nucleic Acids Research*, 2020.

52 J. C. Navarro-Muñoz, N. Selem-Mojica, M. W. Mullowney, S. A. Kautsar, J. H. Tryon, E. I. Parkinson, E. L. C. de Los Santos, M. Yeong, P. Cruz-Morales, S. Abubucker, A. Roeters, W. Lokhorst, A. Fernandez-Guerra, L. T. D. Cappelini, A. W. Goering, R. J. Thomson, W. W. Metcalf, N. L. Kelleher, F. Barona-Gomez and M. H. Medema, *Nature Chemical Biology*, 2020, **16**, 60–68.

53 M. G. Chevrette, K. Gutiérrez-García, N. Selem-Mojica, C. Aguilar-Martínez, A. Yañez-Olvera, H. E. Ramos-Aboites, P. A. Hoskisson and F. Barona-Gómez, *Natural Product Reports*, 2020, **37**, 566–599.

54 D. P. Fewer and M. Metsä-Ketelä, *The FEBS Journal*, 2020, **287**, 1429–1449.

55 T. Kawasaki, M. Taniguchi, Y. Moritani, K. Hayashi, T. Saito, J. Takasaki, K. Nagai, O. Inagaki and H. Shikama, *Thrombosis and Haemostasis*, 2003, **90**, 406–413.

56 M. Kuschak, V. Namasivayam, M. Rafehi, J. H. Voss, J. Garg, J. G. Schlegel, A. Abdelrahman, S. Kehraus, R. Reher, J. Küppers, K. Sylvester, S. Hinz, M. Matthey, D. Wenzel, B. K. Fleischmann, A. Pfeifer, A. Inoue, M. Gütschow, G. M. König and C. E. Müller, *British Journal of Pharmacology*, 2019.

57 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 and K. Strømgaard, *European Journal of Medicinal Chemistry*, 2018, **156**, 847–860.

58 A. Miyamae, M. Fujioka, S. Koda and Y. Morimoto, *J. Chem. Soc., Perkin Trans. 1*, 1989, 873.

59 T. Kawasaki, M. Taniguchi, Y. Moritani, T. Uemura, T. Shigenaga, H. Takamatsu, K. Hayashi, J. Takasaki, T. Saito and K. Nagai, *Thrombosis and Haemostasis*, 2005, **94**, 184–192.

60 T. Uemura, H. Takamatsu, T. Kawasaki, M. Taniguchi, E. Yamamoto, Y. Tomura, W. Uchida and K. Miyata, *European Journal of Pharmacology*, 2006, **536**, 154–161.

61 T. Uemura, T. Kawasaki, M. Taniguchi, Y. Moritani, K. Hayashi, T. Saito, J. Takasaki, W. Uchida and K. Miyata, *British Journal of Pharmacology*, 2006, **148**, 61–69.

62 V. Inamdar, A. Patel, B. K. Manne, C. Dangelmaier and S. P. Kunapuli, *Platelets*, 2015, **26**, 771–778.

63 M. M. Meleka, A. J. Edwards, J. Xia, S. A. Dahlen, I. Mohanty, M. Medcalf, S. Aggarwal, K. D. Moeller, O. V. Mortensen and P. Osei-Owusu, *Pharmacological Research*, 2019, **141**, 264–275.

64 R. Carr, C. Koziol-White, J. Zhang, H. Lam, S. S. An, G. G. Tall, R. A. Panettieri and J. L. Benovic, *Molecular Pharmacology*, 2016, **89**, 94–104.

65 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 and D. Wenzel, *Science Translational Medicine*, 2017, **9**.

66 K. Klepac, A. Kilic, T. Gnad, L. M. Brown, B. Herrmann, A. Wilderman, A. Balkow, A. Glode, K. Simon, M. E. Lidell, M. J. Betz, S. Enerback, J. Wess, M. Freichel, M. Bluher, G. Konig, E. Kostenis, P. A. Insel and A. Pfeifer, *Nature Communications*, 2016, **7**, 10895.

67 C. D. van Raamsdonk, K. G. Griewank, M. B. Crosby, M. C. Garrido, S. Vemula, T. Wiesner, A. C. Obenauf, 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 and B. C. Bastian, *The New England Journal of Medicine*, 2010, **363**, 2191–2199.

68 V. Chua, D. Lapadula, C. Randolph, J. L. Benovic, P. B. Wedegaertner and A. E. Aplin, *Molecular Cancer Research*, 2017, **15**, 501–506.

69 M. D. Onken, C. M. Makepeace, K. M. Kaltenbronn, S. M. Kanai, T. D. Todd, S. Wang, T. J. Broekelmann, P. K. Rao, J. A. Cooper and K. J. Blumer, *Science Signaling*, 2018, **11**.

70 S. Annala, X. Feng, N. Shridhar, F. Eryilmaz, J. Patt, J. 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 and E. Kostenis, *Science Signaling*, 2019, **12**.

71 M. D. Onken and J. A. Cooper, *Uveal melanoma cells use ameboid and mesenchymal mechanisms of cell motility crossing the endothelium*, 2020.

72 M. S. Grace, T. Lieu, B. Darby, F. C. Abogadie, N. Veldhuis, N. W. Bunnett and P. McIntyre, *British Journal of Pharmacology*, 2014, **171**, 3881–3894.

73 D. Karpinsky-Semper, C.-H. Volmar, S. P. Brothers and V. Z. Slepak, *Molecular Pharmacology*, 2014, **85**, 758–768.

74 D. Kamato, L. Thach, R. Bernard, V. Chan, W. Zheng, H. Kaur, M. Brimble, N. Osman and P. J. Little, *Frontiers in Cardiovascular Medicine*, 2015, **2**, 14.

75 E. M. Wauson, M. L. Guerra, J. Dyachok, K. McGlynn, J. Giles, E. M. Ross and M. H. Cobb, *Molecular Endocrinology (Baltimore, Md.)*, 2015, **29**, 1114–1122.

76 D. Bolognini, C. E. Moss, K. Nilsson, A. U. Petersson, I. Donnelly, E. Sergeev, G. M. König, E. Kostenis, M. Kurowska-Stolarska, A. Miller, N. Dekker, A. B. Tobin and G. Milligan, *The Journal of Biological Chemistry*, 2016, **291**, 18915–18931.

77 S. J. Bradley, C. H. Wiegman, M. M. Iglesias, K. C. Kong, A. J. Butcher, B. Plouffe, E. Goupil, J.-M. Bourgognon, T. Macedo-Hatch, C. LeGouill, K. Russell, S. A. Laporte, G. M. Konig, E. Kostenis, M. Bouvier, K. F. Chung, Y. Amrani and A. B. Tobin, *Proceedings of the National Academy of Sciences of the United States of America*, 2016, **113**, 4524–4529.

78 G.-W. Chang, C.-C. Hsiao, Y.-M. Peng, F. A. Vieira Braga, N. A. M. Kragten, E. B. M. Remmerswaal, M. D. B. van de Garde, R. Straussberg, G. M. Konig, E. Kostenis, V. Knauper, L. Meyaard, R. A. W. van Lier, K. P. J. M. van Gisbergen, H.-H. Lin and J. Hamann, *Cell Reports*, 2016, **15**, 1757–1770.

79 X. Jiang, J. Yang, Z. Shen, Y. Chen, L. Shi and N. Zhou, *Insect Biochemistry and Molecular Biology*, 2016, **75**, 78–88.

80 S. H. Kim, D. A. MacIntyre, A. C. Hanyaloglu, A. M. Blanks, S. Thornton, P. R. Bennett and V. Terzidou, *Molecular and Cellular Endocrinology*, 2016, **420**, 11–23.

81 J. P. Kukkonen, *Biochemical and Biophysical Research Communications*, 2016, **476**, 379–385.

82 Y. Liao, B. Lu, Q. Ma, G. Wu, X. Lai, J. Zang, Y. Shi, D. Liu, F. Han and N. Zhou, *The Journal of Biological Chemistry*, 2016, **291**, 7505–7516.

83 S. Lieb, T. Littmann, N. Plank, J. Felixberger, M. Tanaka, T. Schäfer, S. Krief, S. Elz, K. Friedland, G. Bernhardt, J. Wegener, T. Ozawa and A. Buschauer, *Pharmacological Research*, 2016, **114**, 13–26.

84 R. Badolia, V. Inamdar, B. K. Manne, C. Dangelmaier, J. A. Eble and S. P. Kunapuli, *The Journal of Biological Chemistry*, 2017, **292**, 14516–14531.

85 A. Brüser, A. Zimmermann, B. C. Crews, G. Sliwoski, J. Meiler, G. M. König, E. Kostenis, V. Lede, L. J. Marnett and T. Schöneberg, *Scientific Reports*, 2017, **7**, 2380.

86 D. Devost, R. Sleno, D. Pétrin, A. Zhang, Y. Shinjo, R. Okde, J. Aoki, A. Inoue and T. E. Hébert, *The Journal of Biological Chemistry*, 2017, **292**, 5443–5456.

87 A. Holdfeldt, A. Dahlstrand Rudin, M. Gabl, Z. Rajabkhani, G. M. König, E. Kostenis, C. Dahlgren and H. Forsman, *Journal of Leukocyte Biology*, 2017, **102**, 871–880.

88 C. S. N. Klose, T. Mahlakõiv, J. B. Moeller, L. C. Rankin, A.-L. Flamar, H. Kabata, L. A. Monticelli, S. Moriyama, G. G. Putzel, N. Rakhilin, X. Shen, E. Kostenis, G. M. König, T. Senda, D. Carpenter, D. L. Farber and D. Artis, *Nature*, 2017, **549**, 282–286.

89 Z.-G. Gao, A. Inoue and K. A. Jacobson, *Biochemical Pharmacology*, 2018, **151**, 201–213.

90 M. Grundmann, N. Merten, D. Malfacini, A. Inoue, P. Preis, K. Simon, N. Rüttiger, N. Ziegler, T. Benkel, N. K. Schmitt, S. Ishida, I. Müller, R. Reher, K. Kawakami, A. Inoue, U. Rick, T. Kühl, D. Imhof, J. Aoki, G. M. König, C. Hoffmann, J. Gomeza, J. Wess and E. Kostenis, *Nature Communications*, 2018, **9**, 341.

91 X. Lian, S. Beer-Hammer, G. M. König, E. Kostenis, B. Nürnberg and M. Gollasch, *Frontiers in Physiology*, 2018, **9**, 1234.

92 T. Littmann, T. Ozawa, C. Hoffmann, A. Buschauer and G. Bernhardt, *Scientific Reports*, 2018, **8**, 17179.

93 S. C. Nimmagadda, S. Frey, B. Edelmann, C. Hellmich, L. Zaitseva, G. M. König, E. Kostenis, K. M. Bowles and T. Fischer, *Leukemia*, 2018, **32**, 846–849.

94 N. Patelis, K. Kakavia, K. Maltezos, C. Damaskos, E. Spartalis, S. Matheiken and S. Georgopoulos, *Current Pharmaceutical Design*, 2018, **24**, 4558–4563.

95 C. Ruzza, F. Ferrari, R. Guerrini, E. Marzola, D. Preti, R. K. Reinscheid and G. Calo, *Pharmacology Research & Perspectives*, 2018, **6**, e00445.

96 Z. Shen, X. Yang, Y. Chen and L. Shi, *Insect Biochemistry and Molecular Biology*, 2018, **98**, 1–15.

97 S. C. Wright, M. C. A. Cañizal, T. Benkel, K. Simon, C. Le Gouill, P. Matricon, Y. Namkung, V. Lukasheva, G. M. König, S. A. Laporte, J. Carlsson, E. Kostenis, M. Bouvier, G. Schulte and C. Hoffmann, *Science Signaling*, 2018, **11**.

98 R. D. Cervantes-Villagrana, S. R. Adame-García, I. García-Jiménez, V. M. Color-Aparicio, Y. M. Beltrán-Navarro, G. M. König, E. Kostenis, G. Reyes-Cruz, J. S. Gutkind and J. Vázquez-Prado, *The Journal of Biological Chemistry*, 2019, **294**, 531–546.

99 C. Coombs, A. Georgantzoglou, H. A. Walker, J. Patt, N. Merten, H. Poplimont, E. M. Busch-Nentwich, S. Williams, C. Kotsi, E. Kostenis and M. Sarris, *Nature Communications*, 2019, **10**, 5166.

100 J. K. Ebner, G. M. König, E. Kostenis, P. Siegert, K. Aktories and J. H. C. Orth, *Bone*, 2019, **127**, 592–601.

101 X. Feng, N. Arang, D. C. Rigiracciolo, J. S. Lee, H. Yeerna, Z. Wang, S. Lubrano, A. Kishore, J. A. Pachter, G. M. König, M. Maggiolini, E. Kostenis, D. D. Schlaepfer, P. Tamayo, Q. Chen, E. Ruppin and J. S. Gutkind, *Cancer Cell*, 2019, **35**, 457-472.e5.

102 C. V. Jørgensen, H. Zhou, M. J. Seibel and H. Bräuner-Osborne, *Journal of Pharmacological and Toxicological Methods*, 2019, **97**, 59–66.

103 M.-C. Kienitz, A. Niemeyer, G. M. König, E. Kostenis, L. Pott and A. Rinne, *Journal of Molecular and Cellular Cardiology*, 2019, **130**, 107–121.

104 S. Murat, M. Bigot, J. Chapron, G. M. König, E. Kostenis, G. Battaglia, F. Nicoletti, E. Bourinet, J. Bockaert, P. Marin and F. Vandermoere, *Molecular Psychiatry*, 2019, **24**, 1610–1626.

105 C. Avet, A. Mancini, B. Breton, C. Le Gouill, A. S. Hauser, C. Normand, H. Kobayashi, F. Gross, M. Hogue, V. Lukasheva, S. Morissette, E. Fauman, J.-P. Fortin, S. Schann, X. Leroy, D. E. Gloriam and M. Bouvier, *Selectivity Landscape of 100 Therapeutically Relevant GPCR Profiled by an Effector Translocation-Based BRET Platform*, 2020.

106 N. Barki, D. Bolognini, U. Börjesson, L. Jenkins, J. Riddell, D. I. Hughes, T. Ulven, B. D. Hudson, E. R. Ulven, N. Dekker, A. B. Tobin and G. Milligan, *Chemogenetic analysis of how receptors for short chain fatty acids regulate the gut-brain axis*, 2020.

107 N. Caengprasath, N. Gonzalez-Abuin, M. Shchepinova, Y. Ma, A. Inoue, E. W. Tate, G. Frost and A. C. Hanyaloglu, *Endosomal free fatty acid receptor 2 signaling is essential for propionate-induced anorectic gut hormone release*, 2020.

108 Z. Cao, L. Yan, Z. Shen, Y. Chen, Y. Shi, X. He and N. Zhou, *Biochimica et Biophysica Acta. Molecular Cell Research*, 2020, 118718.

109 J. Gröper, G. M. König, E. Kostenis, V. Gerke, C. A. Raabe and U. Rescher, *Cells*, 2020, **9**.

110 A. Holdfeldt, S. Lind, C. Hesse, C. Dahlgren and H. Forsman, *Biochemical Pharmacology*, 2020, **180**, 114143.

111 M. Kurz, A.-L. Krett and M. Bünemann, *Molecular Pharmacology*, 2020, **97**, 267–277.

112 E. M. Pfeil, M. Vescovo, T. Vögtle, J. Brands, U. Rick, N. Merten, I.-M. Albrecht, K. Kawakami, Y. Ono, F. M. Ngako Kadji, J. Aoki, F. Häberlein, M. Matthey, J. Garg, S. Hennen, M.-L. Jobin, K. Seier, D. Calebiro, A. Pfeifer, A. Heinemann, D. Wenzel, G. König, B. Nieswandt, B. K. Fleischmann, A. Inoue, K. Simon and E. Kostenis, *SSRN Electronic Journal*, 2020.

113 A. Pietraszewska-Bogiel, L. Joosen, A. O. Chertkova and J. Goedhart, *ACS Omega*, 2020, **5**, 2648–2659.

114 J. Röthe, R. Kraft, T. Schöneberg and D. Thor, *Biological Procedures Online*, 2020, **22**, 4.

115J. Wenzel, C. E. Hansen, C. Bettoni, M. A. Vogt, B. Lembrich, R. Natsagdorj, G. Huber, J. Brands, K. Schmidt, J. C. Assmann, I. Stölting, K. Saar, J. Sedlacik, J. Fiehler, P. Ludewig, M. Wegmann, N. Feller, M. Richter, H. Müller-Fielitz, T. Walther, G. M. König, E. Kostenis, W. Raasch, N. Hübner, P. Gass, S. Offermanns, C. de Wit, C. A. Wagner and M. Schwaninger, *Proceedings of the National Academy of Sciences of the United States of America*, 2020, **117**, 1753–1761.

116A. D. White, F. G. Jean-Alphonse, F. Fang, K. A. Peña, S. Liu, G. M. König, A. Inoue, D. Aslanoglou, S. H. Gellman, E. Kostenis, K. Xiao and J.-P. Vilardaga, *Proceedings of the National Academy of Sciences of the United States of America*, 2020, **117**, 7455–7460.

117E. Lorza-Gil, G. Kaiser, E. Rexen Ulven, G. M. König, F. Gerst, M. B. Oquendo, A. L. Birkenfeld, H.-U. Häring, E. Kostenis, T. Ulven and S. Ullrich, *Scientific Reports*, 2020, **10**, 16497.

118J. Holze, M. Bermudez, E. M. Pfeil, M. Kauk, T. Bödefeld, M. Irmen, C. Matera, C. Dallanoce, M. de Amici, U. Holzgrabe, G. M. König, C. Tränkle, G. Wolber, R. Schrage, K. Mohr, C. Hoffmann, E. Kostenis and A. Bock, *ACS Pharmacology & Translational Science*, 2020, **3**, 859–867.

119W. Wang, C. Jiang, Y. Xu, Q. Ma, J. Yang, Y. Shi and N. Zhou, *Cellular Signalling*, 2020, **73**, 109677.

120R. Frei, J. Nordlohne, U. Hüser, S. Hild, J. Schmidt, F. Eitner and M. Grundmann, *Journal of Leukocyte Biology*, 2020.

121A. Truong, J. H. Yoo, M. T. Scherzer, J. M. S. Sanchez, K. J. Dale, C. G. Kinsey, J. R. Richards, D. Shin, P. C. Ghazi, M. D. Onken, K. J. Blumer, S. J. Odelberg and M. McMahon, *Clinical Cancer Research: an Official Journal of the American Association for Cancer Research*, 2020.

122T. Wang, Z. Cao, Z. Shen, J. Yang, X. Chen, Z. Yang, K. Xu, X. Xiang, Q. Yu, Y. Song, W. Wang, Y. Tian, L. Sun, L. Zhang, S. Guo and N. Zhou, *eLife*, 2020, **9**.

123K. W. Eyring, J. Liu, G. M. König, S. Hidema, K. Nishimori, E. Kostenis and R. W. Tsien, *Oxytocin signals via Gi and Gq to drive persistent CA2 pyramidal cell firing and strengthen CA3-CA1 neurotransmission*, 2020.

124S. M. Khan, R. D. Martin, S. Gora, C. Bouazza, J. Jones-Tabah, A. Zhang, S. MacKinnon, P. Trieu, P. B. Clarke, J. C. Tanny and T. E. Hébert, *An interaction between Gβγ and RNA polymerase II regulates transcription in cardiac fibroblasts*, 2018.

125D. 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 and E. Kostenis, *The Journal of Biological Chemistry*, 2019, **294**, 5747–5758.

126M. W. Boesgaard, K. Harpsøe, M. Malmberg, C. R. Underwood, A. Inoue, J. M. Mathiesen, G. M. König, E. Kostenis, D. E. Gloriam and H. Bräuner-Osborne, *Journal of Biological Chemistry*, 2020.

127C. J. Hutchings, *Expert Opinion on Biological Therapy*, 2020, 1–11.

128C. H. Chau, P. S. Steeg and W. D. Figg, *The Lancet*, 2019, **394**, 793–804.

129S. Coats, M. Williams, B. Kebble, R. Dixit, L. Tseng, N.-S. Yao, D. A. Tice and J.-C. Soria, *Clinical Cancer Research: an Official Journal of the American Association for Cancer Research*, 2019, **25**, 5441–5448.

130F. He, N. Wen, D. Xiao, J. Yan, H. Xiong, S. Cai, Z. Liu and Y. Liu, *Current Medicinal Chemistry*, 2020, **27**, 2189–2219.

131F. Madeira, Y. M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. N. Tivey, S. C. Potter, R. D. Finn and R. Lopez, *Nucleic Acids Research*, 2019, **47**, W636-W641.

132I. Schamari, PhD thesis, Rheinischen Friedrich-Wilhelms-Universität Bonn, 2018.

133F. I. Kraas, V. Helmetag, M. Wittmann, M. Strieker and M. A. Marahiel, *Chemistry & Biology*, 2010, **17**, 872–880.

134S. P. Niehs, B. Dose, K. Scherlach, M. Roth and C. Hertweck, *ChemBioChem: a European Journal of Chemical Biology*, 2018, **19**, 2167–2172.

135T. P. A. Hari, P. Labana, M. Boileau and C. N. Boddy, *ChemBioChem: a European Journal of Chemical Biology*, 2014, **15**, 2656–2661.

136M. Klapper, D. Braga, G. Lackner, R. Herbst and P. Stallforth, *Cell Chemical Biology*, 2018, **25**, 659-665.e9.

137K. Graupner, K. Scherlach, T. Bretschneider, G. Lackner, M. Roth, H. Gross and C. Hertweck, *Angewandte Chemie (International ed. in English)*, 2012, **51**, 13173–13177.

138C. A. Galea, K. D. Roberts, Y. Zhu, P. E. Thompson, J. Li and T. Velkov, *Biochemistry*, 2017, **56**, 657–668.

139K. D. Patel, F. B. d'Andrea, N. M. Gaudelli, A. R. Buller, C. A. Townsend and A. M. Gulick, *Nature Communications*, 2019, **10**, 3868.

140F. Studier and B. A. Moffatt, *Journal of Molecular Biology*, 1986, **189**, 113–130.

141B. A. Pfeifer, S. J. Admiraal, H. Gramajo, D. E. Cane and C. Khosla, *Science (New York, N.Y.)*, 2001, **291**, 1790–1792.

142A. Shirai, A. Matsuyama, Y. Yashiroda, A. Hashimoto, Y. Kawamura, R. Arai, Y. Komatsu, S. Horinouchi and M. Yoshida, *The Journal of Biological Chemistry*, 2008, **283**, 10745–10752.

143A. Rath, M. Glibowicka, V. G. Nadeau, G. Chen and C. M. Deber, *Proceedings of the National Academy of Sciences of the United States of America*, 2009, **106**, 1760–1765.

144V. V. Phelan, Y. Du, J. A. McLean and B. O. Bachmann, *Chemistry & Biology*, 2009, **16**, 473–478.

145M. Zhu, L. Wang and J. He, *ACS Chemical Biology*, 2019, **14**, 256–265.

146S. Vobruba, S. Kadlcik, R. Gazak and J. Janata, *PloS ONE*, 2017, **12**, e0189684.

147J. Li and S. E. Jensen, *Chemistry & Biology*, 2008, **15**, 118–127.

148H. D. Mootz, R. Finking and M. A. Marahiel, *The Journal of Biological Chemistry*, 2001, **276**, 37289–37298.

149M. Kaniusaite, J. Tailhades, E. A. Marschall, R. J. A. Goode, R. B. Schittenhelm and M. J. Cryle, *Chemical Science*, 2019, **10**, 9466–9482.

150S. A. Sieber, C. T. Walsh and M. A. Marahiel, *Journal of the American Chemical Society*, 2003, **125**, 10862–10866.

151E. Gross, J. Meinhofer and S. Udenfriend, *The peptides. Analysis - synthesis - biology*, Elsevier, Orlando, 1979-1987.

152Y.-Q. Cheng, M. Yang and A. M. Matter, *Applied and Environmental Microbiology*, 2007, **73**, 3460–3469.

153T. T. Hoang, R. R. Karkhoff-Schweizer, A. J. Kutchma and H. P. Schweizer, *Gene*, 1998, **212**, 77–86.

154C. Sánchez, A. F. Braña, C. Méndez and J. A. Salas, *ChemBioChem: a European Journal of Chemical Biology*, 2006, **7**, 1231–1240.

155L. C. M. Ngoka and M. L. Gross, *Journal of the American Society for Mass Spectrometry*, 1999, **10**, 360–363.

156C. C. Ladner and G. J. Williams, *Journal of Industrial Microbiology & Biotechnology*, 2016, **43**, 371–387.

157L. Shi and B. P. Tu, *Current Opinion in Cell Biology*, 2015, **33**, 125–131.

158N. M. Gaudelli and C. A. Townsend, *Nature Chemical Biology*, 2014, **10**, 251–258.

159M. Peschke, C. Brieke, M. Heimes and M. J. Cryle, *ACS Chemical Biology*, 2018, **13**, 110–120.

160M. Kotowska and K. Pawlik, *Applied Microbiology and Biotechnology*, 2014, **98**, 7735–7746.

161S. P. Niehs, B. Dose, S. Richter, S. J. Pidot, H.-M. Dahse, T. P. Stinear and C. Hertweck, *Angewandte Chemie (International ed. in English)*, 2020, **59**, 7766–7771.

162S. Grüschow, E. J. Rackham, B. Elkins, P. L. A. Newill, L. M. Hill and R. J. M. Goss, *ChemBioChem: a European Journal of Chemical Biology*, 2009, **10**, 355–360.

163C. J. Dutton, S. P. Gibson, A. C. Goudie, K. S. Holdom, M. S. Pacey, J. C. Ruddock, J. D. Bu'Lock and M. K. Richards, *The Journal of Antibiotics*, 1991, **44**, 357–365.

164P. Wongkittichote, N. Ah Mew and K. A. Chapman, *Molecular Genetics and Metabolism*, 2017, **122**, 145–152.

165K. H. McClean, M. K. Winson, L. Fish, A. Taylor, S. R. Chhabra, M. Camara, M. Daykin, J. H. Lamb, S. Swift, B. W. Bycroft, G. S. Stewart and P. Williams, *Microbiology*, 1997, **143**, 3703–3711.

166Y. Wang, A. Ikawa, S. Okaue, S. Taniguchi, I. Osaka, A. Yoshimoto, Y. Kishida, R. Arakawa and K. Enomoto, *Bioscience, Biotechnology, and Biochemistry*, 2008, **72**, 1958–1961.

167H. Wang, P. Jiang, Y. Lu, Z. Ruan, R. Jiang, X.-H. Xing, K. Lou and D. Wei, *Biochemical Engineering Journal*, 2009, **44**, 119–124.

168R. Schröder, N. Janssen, J. Schmidt, A. Kebig, N. Merten, S. Hennen, A. Müller, S. Blättermann, M. Mohr-Andrä, S. Zahn, J. Wenzel, N. J. Smith, J. Gomeza, C. Drewke, G. Milligan, K. Mohr and E. Kostenis, *Nature Biotechnology*, 2010, **28**, 943–949.

169R. Schröder, J. Schmidt, S. Blättermann, L. Peters, N. Janssen, M. Grundmann, W. Seemann, D. Kaufel, N. Merten, C. Drewke, J. Gomeza, G. Milligan, K. Mohr and E. Kostenis, *Nature Protocols*, 2011, **6**, 1748–1760.

170T. Kittilä, M. Schoppet and M. J. Cryle, *ChemBioChem: a European Journal of Chemical Biology*, 2016, **17**, 576–584.

171M. H. Medema, P. Cimermancic, A. Sali, E. Takano and M. A. Fischbach, *PLoS Computational Biology*, 2014, **10**, e1004016.

172M. A. Fischbach, C. T. Walsh and J. Clardy, *Proceedings of the National Academy of Sciences of the United States of America*, 2008, **105**, 4601–4608.

173N. Waglechner, A. G. McArthur and G. D. Wright, *Nature Microbiology*, 2019, **4**, 1862–1871.

174M. E. Hillenmeyer, G. A. Vandova, E. E. Berlew and L. K. Charkoudian, *Proceedings of the National Academy of Sciences of the United States of America*, 2015, **112**, 13952–13957.

175A. Vingadassalon, F. Lorieux, M. Juguet, G. Le Goff, C. Gerbaud, J.-L. Pernodet and S. Lautru, *ACS Chemical Biology*, 2015, **10**, 601–610.

176R. Ueoka, A. R. Uria, S. Reiter, T. Mori, P. Karbaum, E. E. Peters, E. J. N. Helfrich, B. I. Morinaka, M. Gugger, H. Takeyama, S. Matsunaga and J. Piel, *Nature Chemical Biology*, 2015, **11**, 705–712.

177S. Götze, J. Arp, G. Lackner, S. Zhang, H. Kries, M. Klapper, M. García-Altares, K. Willing, M. Günther and P. Stallforth, *Chemical Science*, 2019, **10**, 10979–10990.

178M. A. Dessau and Y. Modis, *Journal of Visualized Experiments*, 2011.

179T. P. Korman, J. M. Crawford, J. W. Labonte, A. G. Newman, J. Wong, C. A. Townsend and S.-C. Tsai, *Proceedings of the National Academy of Sciences of the United States of America*, 2010, **107**, 6246–6251.

180S. C. Tsai, L. J. Miercke, J. Krucinski, R. Gokhale, J. C. Chen, P. G. Foster, D. E. Cane, C. Khosla and R. M. Stroud, *Proceedings of the National Academy of Sciences of the United States of America*, 2001, **98**, 14808–14813.

181S. A. Samel, B. Wagner, M. A. Marahiel and L.-O. Essen, *Journal of Molecular Biology*, 2006, **359**, 876–889.

182A. Tanovic, S. A. Samel, L.-O. Essen and M. A. Marahiel, *Science (New York, N.Y.)*, 2008, **321**, 659–663.

183V. Kallnik, C. Schulz, C. Schultz, P. Schweiger and U. Deppenmeier, *Applied Microbiology and Biotechnology*, 2011, **90**, 1285–1293.

184C. M. Halliwell, G. Morgan, C. P. Ou and A. E. Cass, *Analytical Biochemistry*, 2001, **295**, 257–261.

185A. P. Araújo, G. Oliva, F. Henrique-Silva, R. C. Garratt, O. Cáceres and L. M. Beltramini, *Biochemical and Biophysical Research Communications*, 2000, **272**, 480–484.

186S. Jonic and C. Vénien-Bryan, *Current Opinion in Pharmacology*, 2009, **9**, 636–642.

187E. A. Bhat, M. Abdalla and I. A. Rather, *Global Journal of Biotechnology and Biomaterial Science*, 2018, **4**, 1–7.

188M. K. Grøftehauge, N. R. Hajizadeh, M. J. Swann and E. Pohl, *Acta Crystallographica. Section D, Biological Crystallography*, 2015, **71**, 36–44.

189A. McPherson, *Methods (San Diego, Calif.)*, 2004, **34**, 254–265.

190A. McPherson and B. Cudney, *Acta Crystallographica. Section F, Structural Biology Communications*, 2014, **70**, 1445–1467.

191M. Till, A. Robson, M. J. Byrne, A. V. Nair, S. A. Kolek, P. D. Shaw Stewart and P. R. Race, *Journal of Visualized Experiments*, 2013.

192A. D'Arcy, T. Bergfors, S. W. Cowan-Jacob and M. Marsh, *Acta Crystallographica. Section F, Structural Biology Communications*, 2014, **70**, 1117–1126.

193A. D'Arcy, F. Villard and M. Marsh, *Acta Crystallographica. Section D, Biological Crystallography*, 2007, **63**, 550–554.

194S. Engilberge, F. Riobé, S. Di Pietro, L. Lassalle, N. Coquelle, C.-A. Arnaud, D. Pitrat, J.-C. Mulatier, D. Madern, C. Breyton, O. Maury and E. Girard, *Chemical Science*, 2017, **8**, 5909–5917.

195J.-P. Renaud, A. Chari, C. Ciferri, W.-T. Liu, H.-W. Rémigy, H. Stark and C. Wiesmann, *Nature Reviews. Drug Discovery*, 2018, **17**, 471–492.

196L. A. Earl, V. Falconieri, J. L. Milne and S. Subramaniam, *Current Opinion in Structural Biology*, 2017, **46**, 71–78.

197S. Dutta, J. R. Whicher, D. A. Hansen, W. A. Hale, J. A. Chemler, G. R. Congdon, A. R. H. Narayan, K. Håkansson, D. H. Sherman, J. L. Smith and G. Skiniotis, *Nature*, 2014, **510**, 512–517.

198J. M. Reimer, M. Eivaskhani, I. Harb, A. Guarné, M. Weigt and T. M. Schmeing, *Science (New York, N.Y.)*, 2019, **366**, eaaw4388.

199B. R. Miller, E. J. Drake, C. Shi, C. C. Aldrich and A. M. Gulick, *Journal of Biological Chemistry*, 2016, **291**, 22559–22571.

200M. J. Tarry, A. S. Haque, K. H. Bui and T. M. Schmeing, *Structure*, 2017, **25**, 783-793.e4.

201P. R. Mittl, P. Ernst and A. Plückthun, *Current Opinion in Structural Biology*, 2020, **60**, 93–100.

202Qiagen, *The QIAexpressionist™. A Handbook for High-Level Expression and Purification of 6xHis-Tagged Proteins*, 5th edn., 2003.

203J. Norrander, T. Kempe and J. Messing, *Gene*, 1983, **26**, 101–106.

204D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison and H. O. Smith, *Nature Methods*, 2009, **6**, 343–345.

205M. D. Lynch and R. T. Gill, *Biotechnology and Bioengineering*, 2006, **94**, 151–158.

206C. Wang, S. R. Wesener, H. Zhang and Y.-Q. Cheng, *Chemistry & Biology*, 2009, **16**, 585–593.

207L. Broetto, R. Cecagno, F. H. Sant'anna, S. Weber and I. S. Schrank, *Applied Microbiology and Biotechnology*, 2006, **71**, 450–454.

208S. C. Gill and P. H. von Hippel, *Analytical Biochemistry*, 1989, **182**, 319–326.

209J. Yang and Y. Zhang, *Nucleic Acids Research*, 2015, **43**, W174-81.

210C. Zhang, P. L. Freddolino and Y. Zhang, *Nucleic Acids Research*, 2017, **45**, W291-W299.

211N. Guex and M. C. Peitsch, *Electrophoresis*, 1997, **18**, 2714–2723.

212R. C. Edgar, *Nucleic Acids Research*, 2004, **32**, 1792–1797.

213S. Kumar, G. Stecher, M. Li, C. Knyaz and K. Tamura, *Molecular Biology and Evolution*, 2018, **35**, 1547–1549.

214S. Q. Le and O. Gascuel, *Molecular Biology and Evolution*, 2008, **25**, 1307–1320.

8 List of abbreviations

A domain	adenylation domain
Ac	acetyl
AMP	adenosine monophosphate
ATP	adenosine triphosphate
BGC	biosynthetic gene cluster
BLAST	basic local alignment search tool
bp	base pairs
"Ca. B. crenata"	*"Candidatus* Burkholderia crenata"
C. vaccinii	*Chromobacterium vaccinii* MWU205 DSM25250
C domain	condensation domain
CoA	coenzyme A
CV	column volume
Dha	dehydroalanine
DI water	deionised water
DNA	deoxyribonucleic acid
E domain	epimerisation domain
EDTA	ethylenediaminetetraacetic acid
EIC	extracted ion chromatogram
ESI	electrospray ionization
EtOH	ethanol
E. coli	*Escherichia coli*
FR	FR900359
FPLC	fast protein liquid chromatography
GDP	guanosine diphosphate
GPCRs	G protein-coupled receptors
GTP	guanosine triphosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	histidine
Hle	3-hydroxyleucine
H_2O	water
HPLC-HR-MS/MS	high-performance liquid chromatography coupled to high-resolution tandem mass spectrometry
Ile	isoleucine
IPTG	isopropyl-β-D-thiogalactopyranoside
kDa	kilodalton

KOH	potassium hydroxide
LC-MS	liquid chromatography coupled to mass spectrometry
LB	Luria Bertani medium
Leu	leucine
MALDI-TOF	matrix-assisted laser desorption/ionization coupled to time of flight detector
MeOH	methanol
MLP	MbtH-like protein
MT domain	methyltransferase domain
MWCO	molecular weight cut-off
N-Ac-Hle	*N*-acetyl-3-hydroxyleucine
N-Pp-Hle	*N*-propionyl-3-hydroxyleucine
Ni-NTA	nickel nitrilotriacetic acid
NMR	nuclear magnetic resonance spectroscopy
NRP	non-ribosomal peptide
NRPS	non-ribosomal peptide synthetase
OD_{600}	optical density at 600 nm
PCR	polymerase chain reaction
PKS	polyketid synthetase
Pp	propionyl
Ppant	phosphopantetheine arm
PPi	inorganic pyrophosphate
PTX	pertussistoxin
rMMS	random microseed matrix screening
rpm	rounds per minute
SDS-PAGE	sodium dodecyl sulphate–polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
Ser	serine
TAE	Tris-acetate-EDTA buffer
TB	Terrific broth medium
T domain	thiolation domain
TE domain	thioesterase domain
TSA	thermal shift assay
UV	ultraviolet
wt	wild type
YM	YM-254890

9 Appendix

9.1 Sequence alignments

```
#=======================================
#
# Aligned_sequences: 2
# 1: FrsA_A
# 2: FrsD_A
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1633
# Identity:    1630/1633 (99.8%)
# Similarity:  1630/1633 (99.8%)
# Gaps:           0/1633 ( 0.0%)
# Score: 8138.0
#
#
#=======================================

FrsA_A             1 cggcggagccgtcgcagccggtgtccgacatcgagctgctggacgaggcc     50
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A             1 cggcggagccgtcgcagccggtgtccgacatcgagctgctggacgaggcc     50

FrsA_A            51 gagcgccggcaactgctggtcgactggaaccgcaccggaccggaccacgg    100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A            51 gagcgccggcaactgctggtcgactggaaccgcaccggaccggaccacgg    100

FrsA_A           101 ccaggccaccttcccgcaactgttcgaaacccaggcggccctcaccccgc    150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A           101 ccaggccaccttcccgcaactgttcgaaacccaggcggccctcaccccgc    150

FrsA_A           151 acgccgtcgcgctggaaagcccggacgcccggctcagctatgccgaactg    200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A           151 acgccgtcgcgctggaaagcccggacgcccggctcagctatgccgaactg    200

FrsA_A           201 gacgcccgcgccaaccggctggcgcgccatctgcaaagcctgggcgtcgg    250
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A           201 gacgcccgcgccaaccggctggcgcgccatctgcaaagcctgggcgtcgg    250

FrsA_A           251 cgccgacgtgctggtcggcatctgcctggagcgctcgatcgacatggtgg    300
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A           251 cgccgacgtgctggtcggcatctgcctggagcgctcgatcgacatggtgg    300

FrsA_A           301 tcgcggtgctgggcgcgctgaagtccggcgccgcctatctgccgctgtcg    350
                     |||||||.|||||||.||||||||||||||||||||||||||||||||||
FrsD_A           301 tcgcggtactgggcgtgctgaagtccggcgccgcctatctgccgctgtcg    350

FrsA_A           351 ccggagtacccgacggaacggctggcctacatgctgggcgactcgatggc    400
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A           351 ccggagtacccgacggaacggctggcctacatgctgggcgactcgatggc    400

FrsA_A           401 ccccgtgctgctgaccgactcggcacaagtcgagcggctgccgtcgtatt    450
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A           401 ccccgtgctgctgaccgactcggcacaagtcgagcggctgccgtcgtatt    450

FrsA_A           451 ggggccgggtagtcgaactggaccggctcgacctggacgctctgccggac    500
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A           451 ggggccgggtagtcgaactggaccggctcgacctggacgctctgccggac    500

FrsA_A           501 agcgcgccggaacgggcgctgcgcgccgagcacctggcctatgtgatcta    550
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A           501 agcgcgccggaacgggcgctgcgcgccgagcacctggcctatgtgatcta    550

FrsA_A           551 cacctccggctccaccggccaaccgaagggcgtggcggtcagccacgccg    600
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
FrsD_A          551 cacctccggctccaccggccaaccgaagggcgtggcggtcagccacgccg    600

FrsA_A          601 gcctggccggcctggccggcagccagacagagcggttcgcgctgcaaggc    650
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A          601 gcctggccggcctggccggcagccagacagagcggttcgcgctgcaaggc    650

FrsA_A          651 ccgacgcgggtgctgcaattcgcctcgctgagtttcgacgcggcggtgat    700
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A          651 ccgacgcgggtgctgcaattcgcctcgctgagtttcgacgcggcggtgat    700

FrsA_A          701 ggaaatgctgatggccttctgcagcggcggccggctggtgctgccggcgg    750
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A          701 ggaaatgctgatggccttctgcagcggcggccggctggtgctgccggcgg    750

FrsA_A          751 cggggccgctgctgggcgaacagctgctggacacgctgaaccgccatgaa    800
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A          751 cggggccgctgctgggcgaacagctgctggacacgctgaaccgccatgaa    800

FrsA_A          801 attagccacgcgctgatctcgccgtcggcgctgagcaccgcggacgcggc    850
                    ||.||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A          801 atcagccacgcgctgatctcgccgtcggcgctgagcaccgcggacgcggc    850

FrsA_A          851 gttggcgccggtcctgcggacgctggtggtgggcggggaagcctgcccgg    900
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A          851 gttggcgccggtcctgcggacgctggtggtgggcggggaagcctgcccgg    900

FrsA_A          901 gcgcgacggtggcggcctggtcggcgggacggcggatggtgaacgcctac    950
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A          901 gcgcgacggtggcggcctggtcggcgggacggcggatggtgaacgcctac    950

FrsA_A          951 ggtccgaccgaggcgacggcctgcgtgacgatgagcgagccgctgtccgg   1000
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A          951 ggtccgaccgaggcgacggcctgcgtgacgatgagcgagccgctgtccgg   1000

FrsA_A         1001 cgacggcgcgccgaagctgggccgtccgacgcacaacgcgcggctgtacg   1050
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A         1001 cgacggcgcgccgaagctgggccgtccgacgcacaacgcgcggctgtacg   1050

FrsA_A         1051 tgctggatggcgcgctgcaactggcgccggtgggggtggcgggcgagctg   1100
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A         1051 tgctggatggcgcgctgcaactggcgccggtgggggtggcgggcgagctg   1100

FrsA_A         1101 tacatcgcgggggccgggctggcgcgcggctatctgaaccggccggggct   1150
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A         1101 tacatcgcgggggccgggctggcgcgcggctatctgaaccggccggggct   1150

FrsA_A         1151 gacggcggagcgcttcgtggcgaatccgtacggagagggtgagcggctgt   1200
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A         1151 gacggcggagcgcttcgtggcgaatccgtacggagagggtgagcggctgt   1200

FrsA_A         1201 accgcagcggcgacctggcgcggtggacggaagaaggcgagctggaatac   1250
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A         1201 accgcagcggcgacctggcgcggtggacggaagaaggcgagctggaatac   1250

FrsA_A         1251 ctggggcgcagcgaccagcaggtgaaggtgcggggtttccgtatcgagcc   1300
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A         1251 ctggggcgcagcgaccagcaggtgaaggtgcggggtttccgtatcgagcc   1300

FrsA_A         1301 gggcgagatcgaagcggtgctgaaccggcatccgcaagtgagccagtcgg   1350
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A         1301 gggcgagatcgaagcggtgctgaaccggcatccgcaagtgagccagtcgg   1350

FrsA_A         1351 tggtggtggcgcggcagagccagggcggcgacagccagttggtggcgtac   1400
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A         1351 tggtggtggcgcggcagagccagggcggcgacagccagttggtggcgtac   1400

FrsA_A         1401 gtggcggccgtcggcggggtggaggggtcggagctgcggcgcctggcggc   1450
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A         1401 gtggcggccgtcggcggggtggaggggtcggagctgcggcgcctggcggc   1450
```

```
FrsA_A          1451 ggggcagctgccggagcacatggtgccggcggcggtggtggtgctggaat   1500
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A          1451 ggggcagctgccggagcacatggtgccggcggcggtggtggtgctggaat   1500

FrsA_A          1501 cgctgccgcagttgccgaacgggaagctggaccgcaagtcgctgccggcg   1550
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A          1501 cgctgccgcagttgccgaacgggaagctggaccgcaagtcgctgccggcg   1550

FrsA_A          1551 ccggagtttggcggctcgcattatcagcggccgcgcaacgcgcaggagga   1600
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_A          1551 ccggagtttggcggctcgcattatcagcggccgcgcaacgcgcaggagga   1600

FrsA_A          1601 aatgctgtgcgggctgttcgcggaagtgctgga   1633
                     |||||||||||||||||||||||||||||||||
FrsD_A          1601 aatgctgtgcgggctgttcgcggaagtgctgga   1633


#---------------------------------------
#---------------------------------------
```

Figure 9.1: Emboss Needle Alignment of the A domain of frsA and frsD in *C. vaccinii*. Non-identical base pairs are highlighted in grey

```
# Aligned_sequences: 2
# 1: FrsA_C
# 2: FrsD_C
# Matrix: EDNAFULL
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1323
# Identity:    1218/1323 (92.1%)
# Similarity:  1218/1323 (92.1%)
# Gaps:          54/1323 ( 4.1%)
# Score: 5707.5
#
#
#=======================================

FrsA_C             1 cat-------ttcaggcatcttcagc-acagctggatgtatggatttct     42
                       ||        |..|.|||        || |||||| ||||
FrsD_C             1 -atggaaatatggctggc-------gcaacagct-gatg-----------     30

FrsA_C            43 caggaagtttcaccgaatctgcccaacaatattgccgagtatctgaatct     92
                     |.|||    |||.||||||          ||||||||||||||||.||||
FrsD_C            31 ccgga---ttcgccgaat---------aatattgccgagtatctgcatct     68

FrsA_C            93 cgccggctcgttggatgctggattgtttctgcaggctttaagccaggtcg    142
                     ..||||..||||||||.|.|..|||||.|..|..|.||..|.|||||||
FrsD_C            69 ttccggtccgttggatccagatttgttttcaaaaccttgcggcaggtcg    118

FrsA_C           143 ccagtgagagcgcggagc--tgcaatacaacttt---ccgtcacgatggtc    187
                     |.||.|||| |||.| ||  ||||...|||.||    |..|.|.|||||.|
FrsD_C           119 cgagcgaga-cgccg-gcattgcaggtcaatttttccattgaggatggac    166

FrsA_C           188 -tccagttgaccaagtttcgtcgagatgatgaaggctgggag-ccggact    235
                      .||  |||.||  |.||.|.||.|...||||.|.|| |||| ||.||.|
FrsD_C           167 ggcc--ttgccc--ggttagccgcgtccatgaggact-ggagtcctgatt    211

FrsA_C           236 tcatcgatgtatcgacgcacggcgagccggaacacgcagccctgcgcgcc    285
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_C           212 tcatcgatgtatcgacgcacggcgagccggaacacgcagccctgcgcgcc    261
```

```
FrsA_C          286 atgcgggagcgggtggagaaacccttcgatctggcgcgggacgcgttgtt    335
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_C          262 atgcgggagcgggtggagaaacccttcgatctggcgcgggacgcgttgtt    311

FrsA_C          336 tcgctggaccttgatccgcctggccgacgagcgccacatcttctgccatg    385
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_C          312 tcgctggaccttgatccgcctggccgacgagcgccacatcttctgccatg    361

FrsA_C          386 tgtatcaccacatcgcgatggatggggccggctatgtgatgctgctgcag    435
                    ||||||||||||||||||||||||..||||||||||||||||||||||||||
FrsD_C          362 tgtatcaccacatcgcgatggatgtccggctatgtgatgctgctgcag    411

FrsA_C          436 cgcatagccgaggtttacggcgcgctgcgggaaggccagccggcaccggc    485
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_C          412 cgcatagccgaggtttacggcgcgctgcgggaaggccagccggcaccggc    461

FrsA_C          486 ctgcggtttcgccgatgcggatgccatcgtccgcgaggaagagcgctacc    535
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_C          462 ctgcggtttcgccgatgcggatgccatcgtccgcgaggaagagcgctacc    511

FrsA_C          536 gccagtcggagcagttcgcggtcgaccgggcattctggcaagcgcgctcg    585
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_C          512 gccagtcggagcagttcgcggtcgaccgggcattctggcaagcgcgctcg    561

FrsA_C          586 gccgagctggcgacggcggagccgccgctgccggcggccgatggcccgtt    635
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_C          562 gccgagctggcgacggcggagccgccgctgccggcggccgatggcccgtt    611

FrsA_C          636 cctggcgttcgcccagacggcggtgattccggaagacgcctgcggcgggc    685
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_C          612 cctggcgttcgcccagacggcggtgattccggaagacgcctgcggcgggc    661

FrsA_C          686 tgcggatgacggccgagcggctgggcgtctcccagtcccgtttgctgaca    735
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_C          662 tgcggatgacggccgagcggctgggcgtctcccagtcccgtttgctgaca    711

FrsA_C          736 gcagccatcgtcgcttatttccatcgctggggcggccagcaagagatctt    785
                    ||.||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_C          712 gcggccatcgtcgcttatttccatcgctggggcggccagcaagagatctt    761

FrsA_C          786 gttccggctggcggtatcggcgcgcagcgatgcgacgcgacacgcgcccg    835
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_C          762 gttccggctggcggtatcggcgcgcagcgatgcgacgcgacacgcgcccg    811

FrsA_C          836 gccacctggcgcatgcgttgccgctgctggccagcctgccgccgcgcgcc    885
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_C          812 gccacctggcgcatgcgttgccgctgctggccagcctgccgccgcgcgcc    861

FrsA_C          886 agtctggccgacatcgcgcgacagctggacggcgaggtggagcggatgcg    935
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_C          862 agtctggccgacatcgcgcgacagctggacggcgaggtggagcggatgcg    911

FrsA_C          936 tccgcatacccgctatcgggctgaggacatcgtgcgcgaccaggccggtg    985
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_C          912 tccgcatacccgctatcgggctgaggacatcgtgcgcgaccaggccggtg    961

FrsA_C          986 ccggtttggggcgcggggcgcaggggcctgtgatcaacctcatgcctttt   1035
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_C          962 ccggtttggggcgcggggcgcaggggcctgtgatcaacctcatgcctttt   1011

FrsA_C         1036 gcttaccgcttcgagtttggcgcctgtcgcgtggagtccgcccatcagct   1085
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_C         1012 gcttaccgcttcgagtttggcgcctgtcgcgtggagtccgcccatcagct   1061

FrsA_C         1086 gaccgtcggcgtgctggacacgctggaagtggcggtgcacgaccgcaaga   1135
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_C         1062 gaccgtcggcgtgctggacacgctggaagtggcggtgcacgaccgcaaga   1111

FrsA_C         1136 acggtgacggcctccacctcgatttgtacgcatccgagcgcggctgcccg   1185
```

```
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_C          1112 acggtgacggcctccacctcgatttgtacgcatccgagcgcggctgcccg   1161

FrsA_C          1186 cccgaaccgctgcggcggcatgccctgcggctggcccggttcatcgtcga   1235
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_C          1162 cccgaaccgctgcggcggcatgccctgcggctggcccggttcatcgtcga   1211

FrsA_C          1236 ggcggcggcggagccgtcgcagccggtgtccgacatcgagctgctggacg   1285
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
FrsD_C          1212 ggcggcggcggagccgtcgcagccggtgtccgacatcgagctgctggacg   1261

FrsA_C          1286 aggccgagcgccggcaactgctg   1308
                     |||||||||||||||||||||||
FrsD_C          1262 aggccgagcgccggcaactgctg   1284


#---------------------------------------
#---------------------------------------
```

Figure 9.2: Emboss Needle Alignment of the C domain of frsA and frsD in *C. vaccinii*. Non-identical base pairs are highlighted in grey.

```
        TE_FrsA            1 HPLFCIHPEGGLGWSYIGLALHLDHEQPIYTLQARGLDGMSELAPSIPDM     50
                             .|||||||.|.|.|:|:.|..:||.|||||.|||||:||.||.|.||..|
        TE_FrsG            1 PPLFCIHPGGCLSWTYVSLVRYLDAEQPIYGLQARGIDGQSEPASSIEAM     50

        TE_FrsA           51 AADYIEQIRSIQPNGPYHLLGWSLGGVIAQEVAVQLERVGEKTALLAILD    100
                             ||||:.|||.|||:|||:||||||||.:||.:|.|||.:.::..||.:||
        TE_FrsG           51 AADYVAQIRGIQPHGPYYLLGWSLGGNLAQAMASQLESMDQEVGLLFLLD    100

        TE_FrsA          101 TFPIEILHEAMFGKQACAYDLFARVVQEMYLMPIEEARLKSMYLIGLNHM    150
                             :.| ..:|:   ..:...|.||.:..:..:...:.|.:::::::.:...|:
        TE_FrsG          101 SGP-SPMHK---DDEMIEYPLFTKEFKNTFKFHVSETKMQAIFEVTKRHV    146

        TE_FrsA          151 KITAAFSSSHYGGDLLLFRSLIPYAEDALMPEADTWQPYLSGQLEVHDIE    200
                             ::....:....|..||||:.:||.|...:.....|..|:.|.:|||:.
        TE_FrsG          147 ELIRQSTTPVSQGPALLFRATVPYDESTPLLPPHAWNEYVKGDIEVHEVH    196

        TE_FrsA          201 CTHMDMMQRDVLKIIGPVLESKL    223
                             |.|..|.:.:.::.:|||:|.||
        TE_FrsG          197 CQHAQMNRIEFMEQMGPVIERKL    219

                                       Length: 223
                             # Identity:      93/223 (41.7%)
                             # Similarity:   139/223 (62.3%)
                                  # Gaps:           4/223 (1.8%)
                                  # Score: 482.0
```

Figure 9.3: Emboss Needle Alignment of $FrsA_{TE}$ and $FrsG_{TE}$ from *C. vaccinii*. Residues in the active site which are responsible for hydrolysis are marked in grey.

```
#=======================================
#
# Aligned_sequences: 2
# 1: cvFrsA_CAT
# 2: cvFrsD_CAT
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 1021
# Identity:     943/1021 (92.4%)
# Similarity:   962/1021 (94.2%)
# Gaps:          18/1021 ( 1.8%)
# Score: 4810.0
#
#
```

```
#=======================================

cvFrsA_CAT         1 AMKNSESPIHHFQASSAQLDVWISQEVSPNLPNNIAEYLNLAGSLDAGLF     50
                                       :::|::|::.|:.||||||||:|:|.||..||
cvFrsD_CAT         1 ------------------MEIWLAQQLMPDSPNNIAEYLHLSGPLDPDLF     32

cvFrsA_CAT        51 LQALSQVASESAELQYNFRHDGLQLTKFRRDDEGWEPDFIDVSTHGEPEH    100
                     .:.|.|||||:..||.||..:..:....|..|.|.|||||||||||||||
cvFrsD_CAT        33 FKTLRQVASETPALQVNFSIEDGRPCPVSRVHEDWSPDFIDVSTHGEPEH     82

cvFrsA_CAT       101 AALRAMRERVEKPFDLARDALFRWTLIRLADERHIFCHVYHHIAMDGAGY    150
                     ||||||||||||||||||||||||||||||||||||||||||||||.|||
cvFrsD_CAT        83 AALRAMRERVEKPFDLARDALFRWTLIRLADERHIFCHVYHHIAMDVAGY    132

cvFrsA_CAT       151 VMLLQRIAEVYGALREGQPAPACGFADADAIVREEERYRQSEQFAVDRAF    200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
cvFrsD_CAT       133 VMLLQRIAEVYGALREGQPAPACGFADADAIVREEERYRQSEQFAVDRAF    182

cvFrsA_CAT       201 WQARSAELATAEPPLPAADGPFLAFAQTAVIPEDACGGLRMTAERLGVSQ    250
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
cvFrsD_CAT       183 WQARSAELATAEPPLPAADGPFLAFAQTAVIPEDACGGLRMTAERLGVSQ    232

cvFrsA_CAT       251 SRLLTAAIVAYFHRWGGQQEILFRLAVSARSDATRHAPGHLAHALPLLAS    300
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
cvFrsD_CAT       233 SRLLTAAIVAYFHRWGGQQEILFRLAVSARSDATRHAPGHLAHALPLLAS    282

cvFrsA_CAT       301 LPPRASLADIARQLDGEVERMRPHTRYRAEDIVRDQAGAGLGRGAQGPVI    350
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
cvFrsD_CAT       283 LPPRASLADIARQLDGEVERMRPHTRYRAEDIVRDQAGAGLGRGAQGPVI    332

cvFrsA_CAT       351 NLMPFAYRFEFGACRVESAHQLTVGVLDTLEVAVHDRKNGDGLHLDLYAS    400
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
cvFrsD_CAT       333 NLMPFAYRFEFGACRVESAHQLTVGVLDTLEVAVHDRKNGDGLHLDLYAS    382

cvFrsA_CAT       401 ERGCPPEPLRRHALRLARFIVEAAAEPSQPVSDIELLDEAERRQLLVDWN    450
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
cvFrsD_CAT       383 ERGCPPEPLRRHALRLARFIVEAAAEPSQPVSDIELLDEAERRQLLVDWN    432

cvFrsA_CAT       451 RTGPDHGQATFPQLFETQAALTPHAVALESPDARLSYAELDARANRLARH    500
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
cvFrsD_CAT       433 RTGPDHGQATFPQLFETQAALTPHAVALESPDARLSYAELDARANRLARH    482

cvFrsA_CAT       501 LQSLGVGADVLVGICLERSIDMVVAVLGALKSGAAYLPLSPEYPTERLAY    550
                     ||||||||||||||||||||||||||||.||||||||||||||||||||||
cvFrsD_CAT       483 LQSLGVGADVLVGICLERSIDMVVAVLGVLKSGAAYLPLSPEYPTERLAY    532

cvFrsA_CAT       551 MLGDSMAPVLLTDSAQVERLPSYWGRVVELDRLDLDALPDSAPERALRAE    600
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
cvFrsD_CAT       533 MLGDSMAPVLLTDSAQVERLPSYWGRVVELDRLDLDALPDSAPERALRAE    582

cvFrsA_CAT       601 HLAYVIYTSGSTGQPKGVAVSHAGLAGLAGSQTERFALQGPTRVLQFASL    650
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
cvFrsD_CAT       583 HLAYVIYTSGSTGQPKGVAVSHAGLAGLAGSQTERFALQGPTRVLQFASL    632

cvFrsA_CAT       651 SFDAAVMEMLMAFCSGGRLVLPAAGPLLGEQLLDTLNRHEISHALISPSA    700
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
cvFrsD_CAT       633 SFDAAVMEMLMAFCSGGRLVLPAAGPLLGEQLLDTLNRHEISHALISPSA    682

cvFrsA_CAT       701 LSTADAALAPVLRTLVVGGEACPGATVAAWSAGRRMVNAYGPTEATACVT    750
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
cvFrsD_CAT       683 LSTADAALAPVLRTLVVGGEACPGATVAAWSAGRRMVNAYGPTEATACVT    732

cvFrsA_CAT       751 MSEPLSGDGAPKLGRPTHNARLYVLDGALQLAPVGVAGELYIAGAGLARG    800
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
cvFrsD_CAT       733 MSEPLSGDGAPKLGRPTHNARLYVLDGALQLAPVGVAGELYIAGAGLARG    782

cvFrsA_CAT       801 YLNRPGLTAERFVANPYGEGERLYRSGDLARWTEEGELEYLGRSDQQVKV    850
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
cvFrsD_CAT       783 YLNRPGLTAERFVANPYGEGERLYRSGDLARWTEEGELEYLGRSDQQVKV    832
```

```
cvFrsA_CAT       851 RGFRIEPGEIEAVLNRHPQVSQSVVVARQSQGGDSQLVAYVAAVGGVEGS    900
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
cvFrsD_CAT       833 RGFRIEPGEIEAVLNRHPQVSQSVVVARQSQGGDSQLVAYVAAVGGVEGS    882

cvFrsA_CAT       901 ELRRLAAGQLPEHMVPAAVVVLESLPQLPNGKLDRKSLPAPEFGGSHYQR    950
                     |||||||||||||||||||||||||||||||||||||||||||||||||||
cvFrsD_CAT       883 ELRRLAAGQLPEHMVPAAVVVLESLPQLPNGKLDRKSLPAPEFGGSHYQR    932

cvFrsA_CAT       951 PRNAQEEMLCGLFAEVLDMEKVGRGDSFFDLGGHSLLATRLIRRIRETLD   1000
                     ||||||||||||||::..||..||||||||||||||.|||..|:
cvFrsD_CAT       933 PRNAQEEMLCGLFAEVLEVGSVGIDDSFFDLGGHSLLATRLISRIRAALN    982

cvFrsA_CAT      1001 VELSIRDLFEAPCVTELSRHI   1021
                     |||.||.||:.|.|.|.||...:
cvFrsD_CAT       983 VELPIRQLFDLPSVAELLEVL   1003


#---------------------------------------
#---------------------------------------
```

Figure 9.4: Emboss Needle Alignment of $FrsA_{CAT}$ and $FrsD_{CAT}$ from *C. vaccinii*.

9.2 Plasmid maps

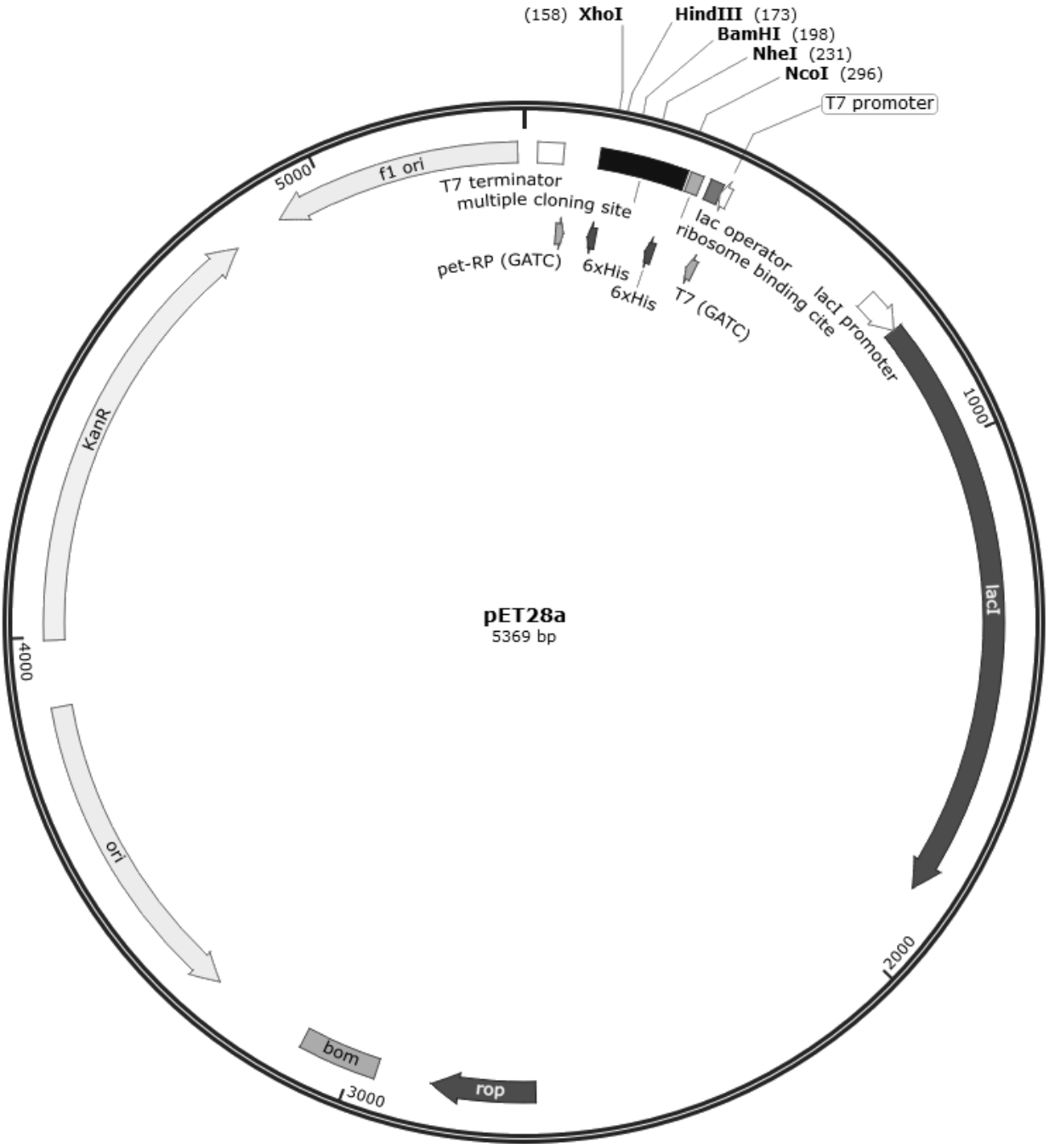


Figure 9.5: Plasmid map of pET28a.

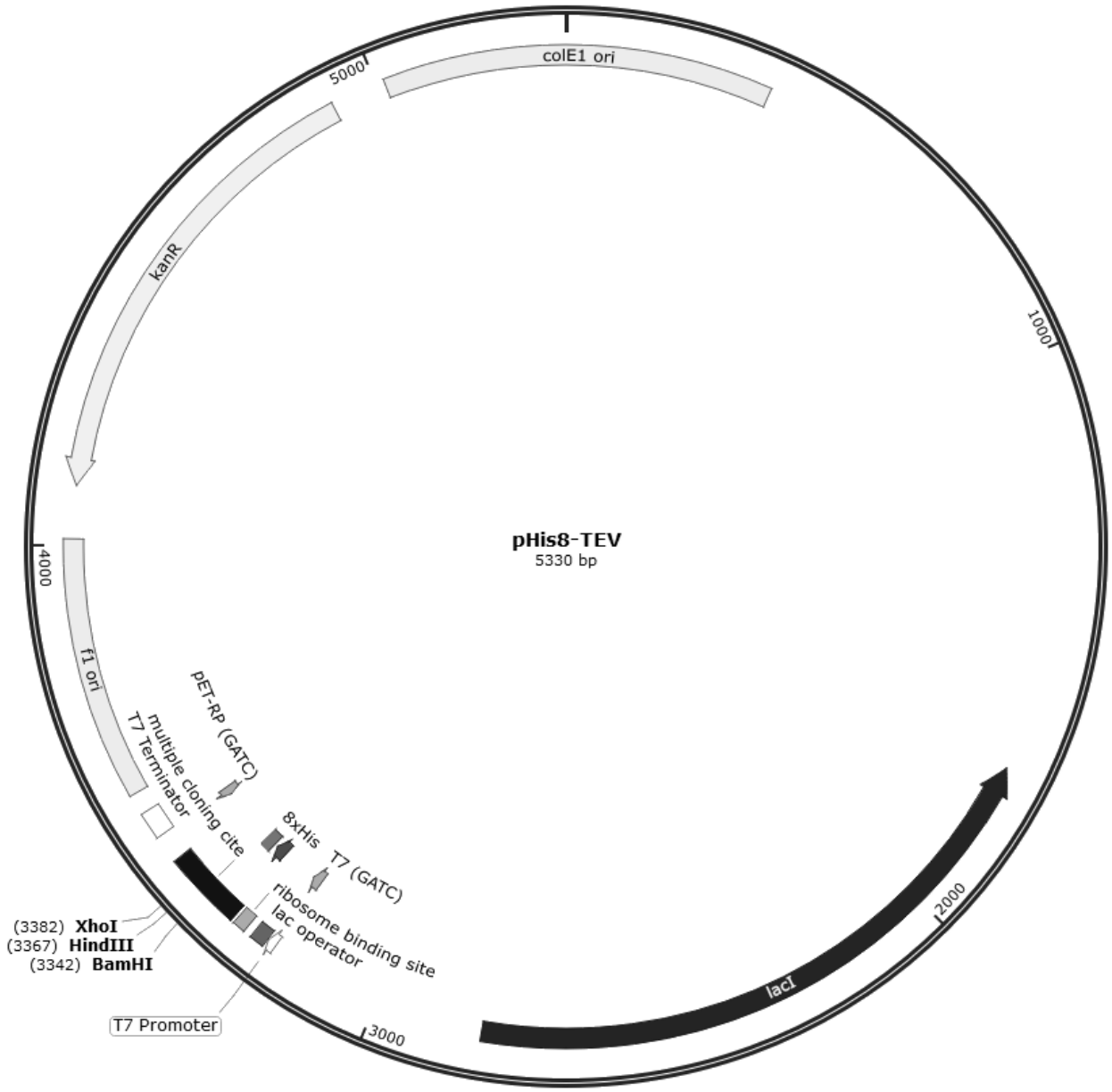


Figure 9.6: Plasmid map of pHis8-TEV.

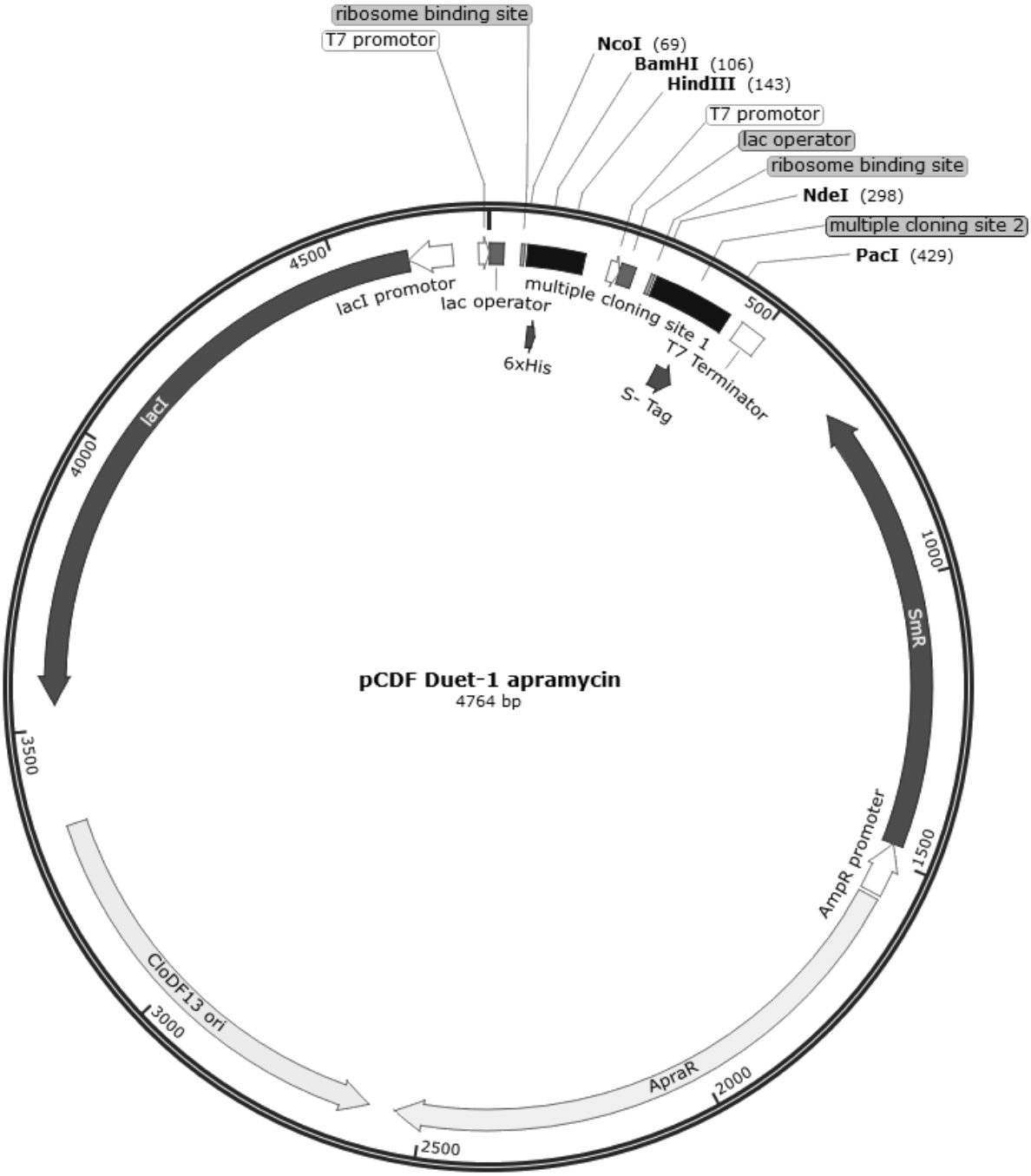


Figure 9.7: Plasmid map of pCDF Duet-1 with apramycin resistance.

9.3 NMR data

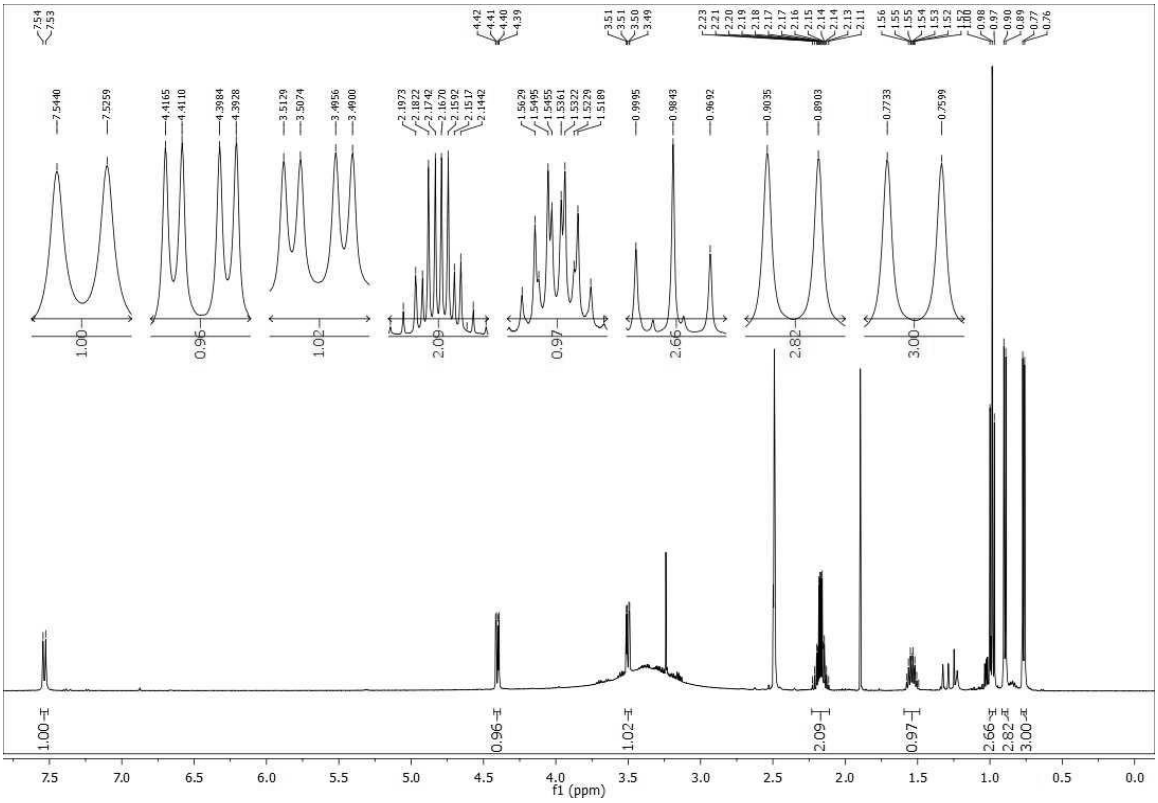


Figure 9.8: ^{1}H NMR spectrum of compound 21 in DMSO-d_6 (500 MHz).

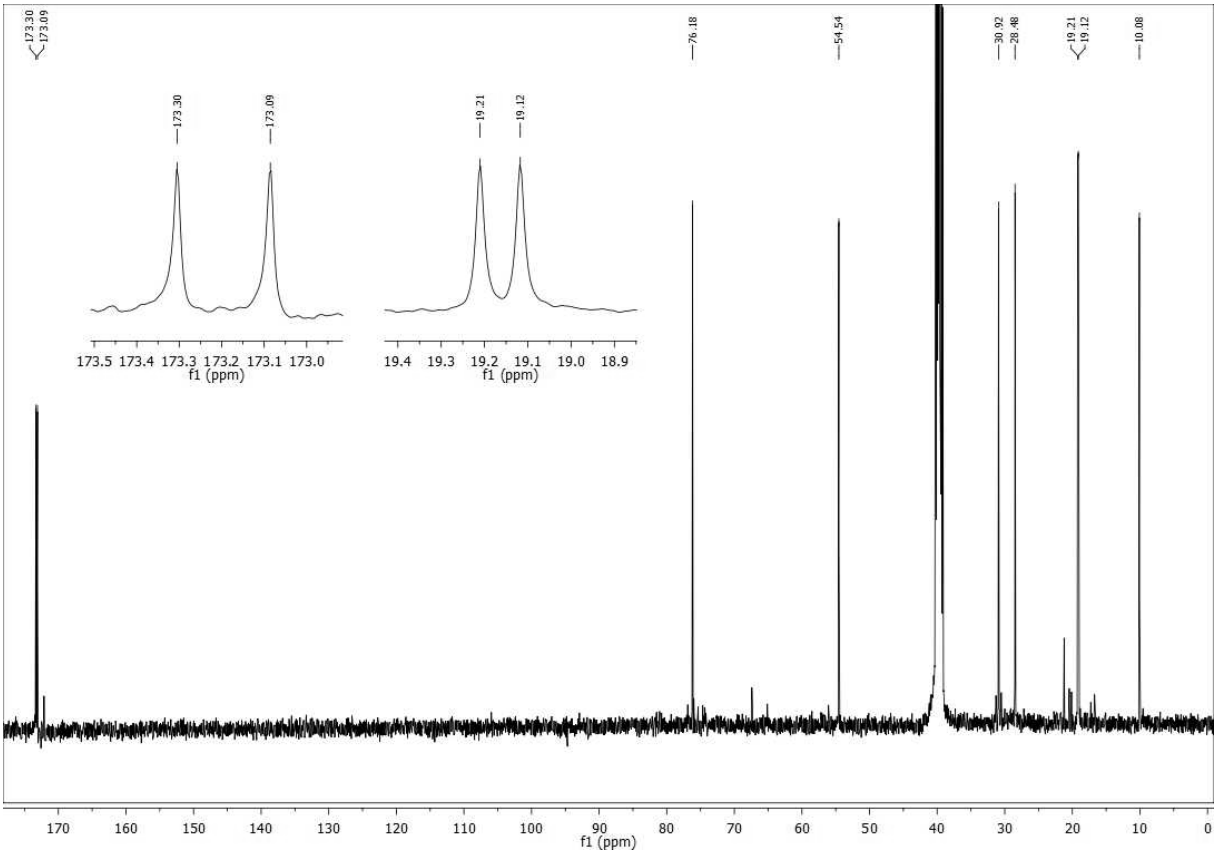


Figure 9.9: ^{13}C NMR spectrum of compound 21 in DMSO-d_6 (125 MHz).

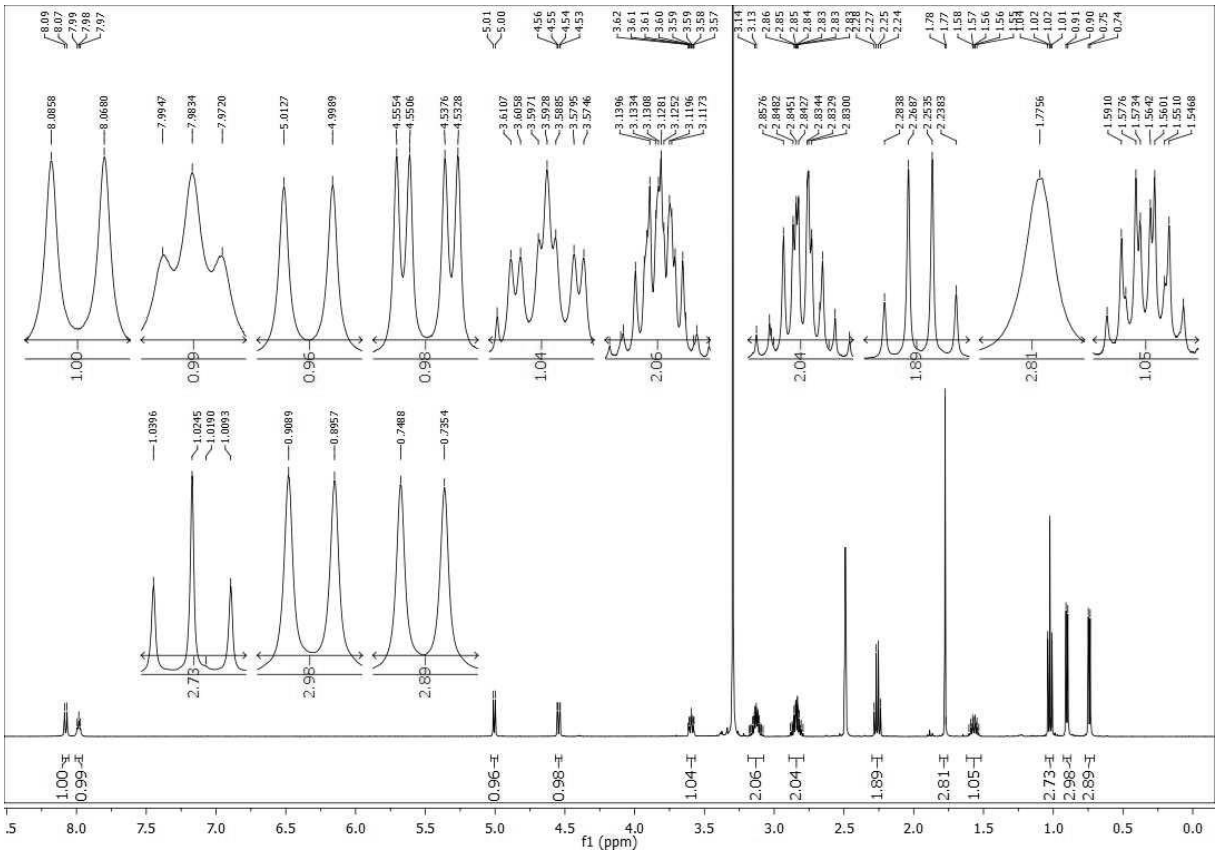


Figure 9.10: ^{1}H NMR spectrum of compound 22 in DMSO-d_6 (500 MHz).

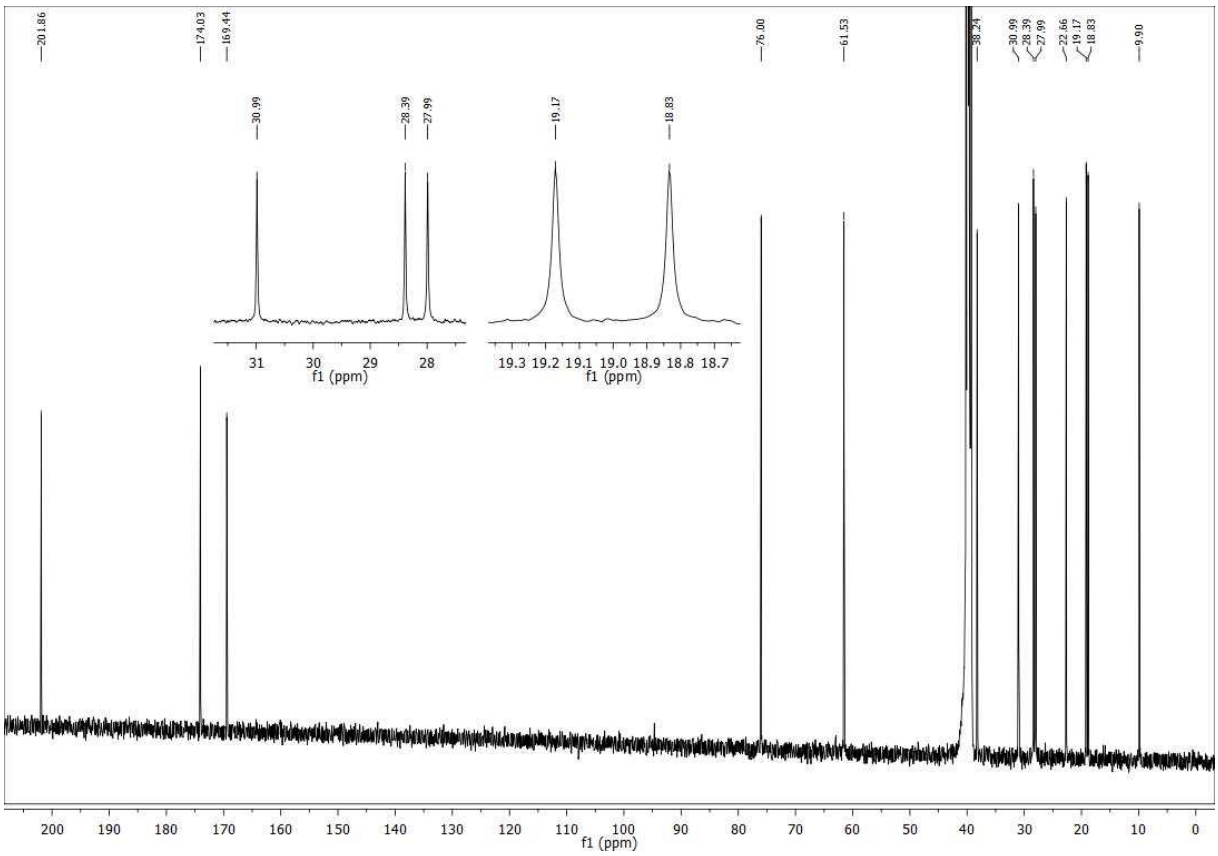


Figure 9.11: ^{13}C NMR spectrum of compound 22 in DMSO-d_6 (125 MHz).

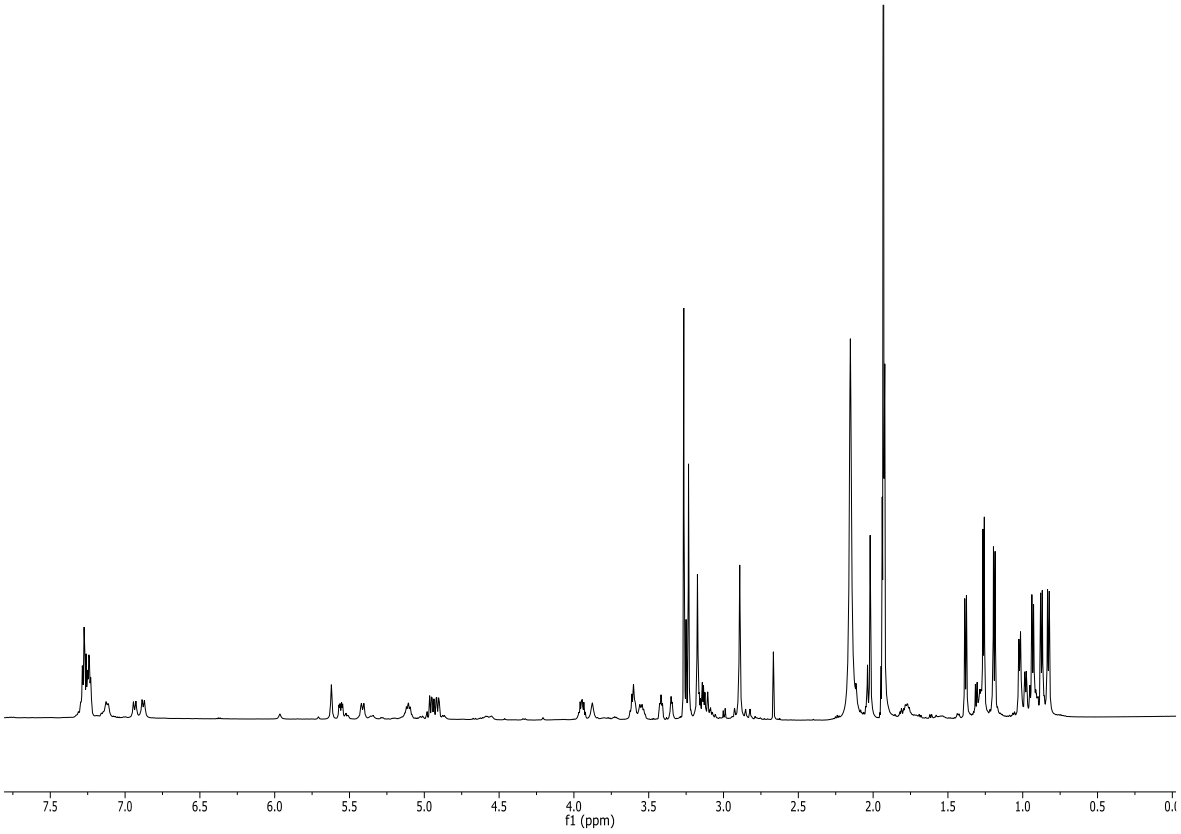


Figure 9.12: ^{1}H NMR spectrum of compound 8 in acetonitrile-d_3 (600 MHz).

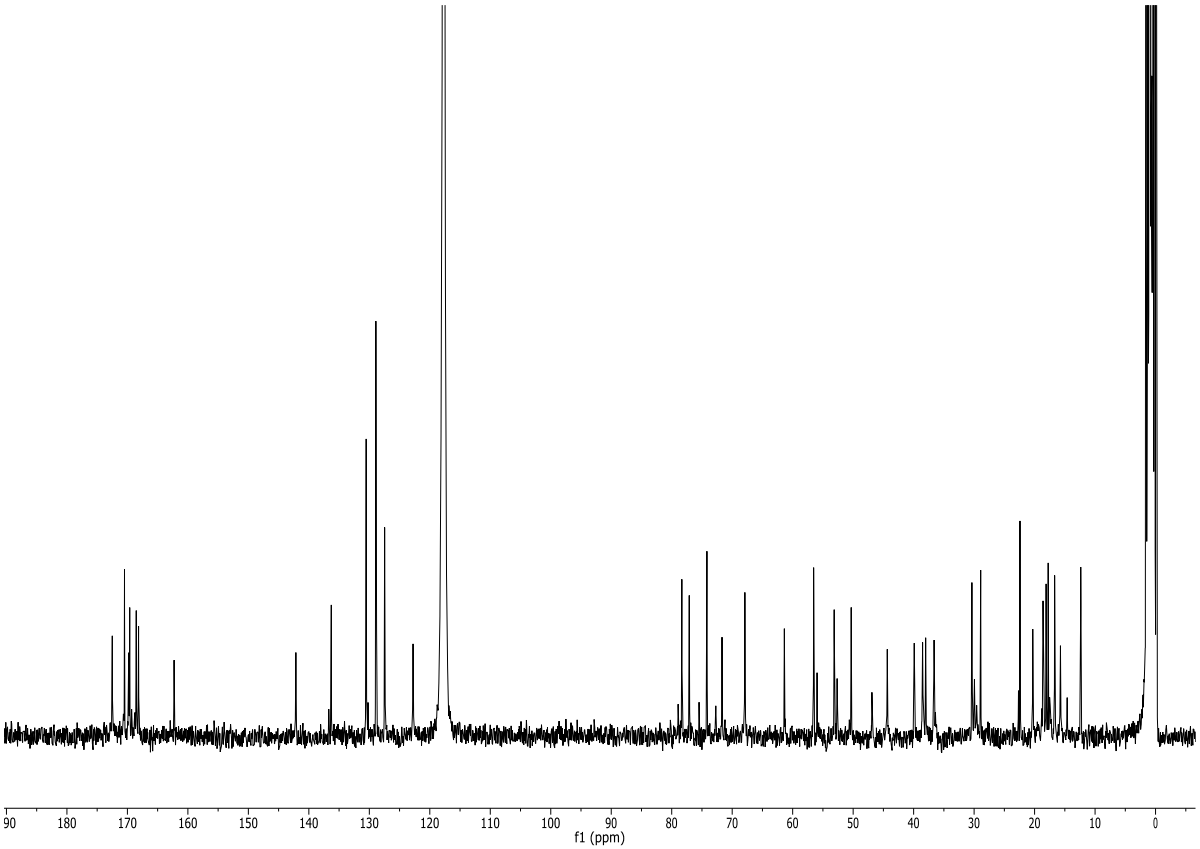


Figure 9.13: ^{13}C NMR spectrum of compound 8 in acetonitrile-*d*$_3$ (150 MHz).

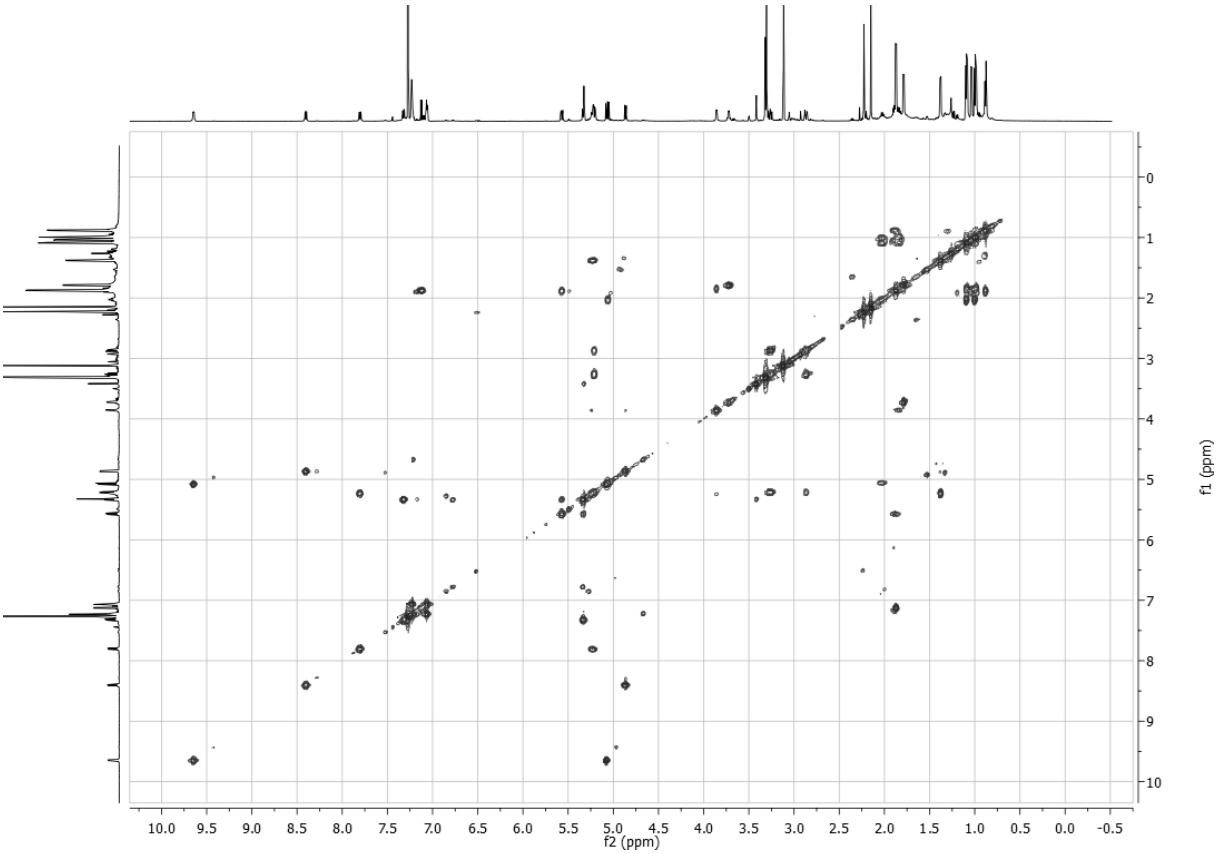


Figure 9.14: ^{1}H-^{1}H COSY NMR spectrum of compound 8 in acetonitrile-*d*$_3$ (600 MHz).

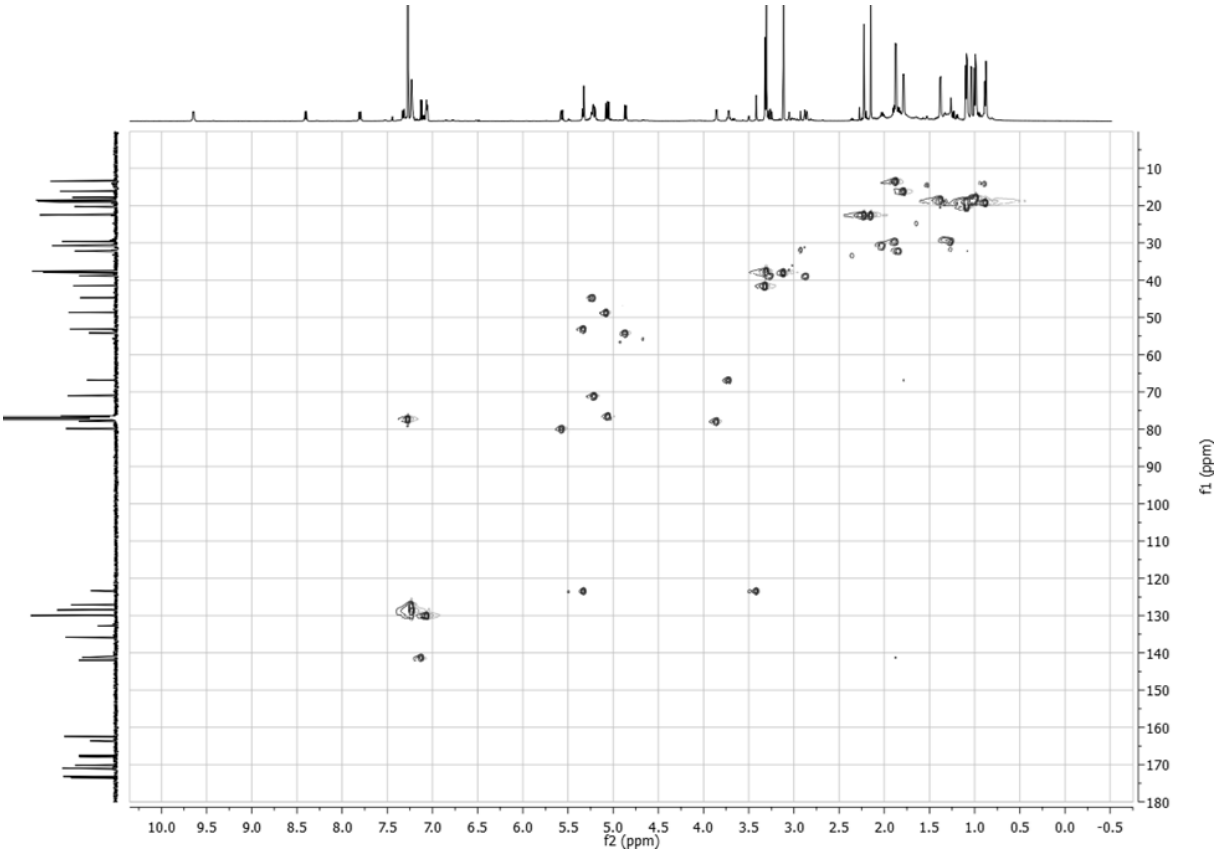


Figure 9.15: ^{1}H-^{13}C HSQC NMR spectrum of compound 8 in acetonitrile-d_3 (600 MHz).

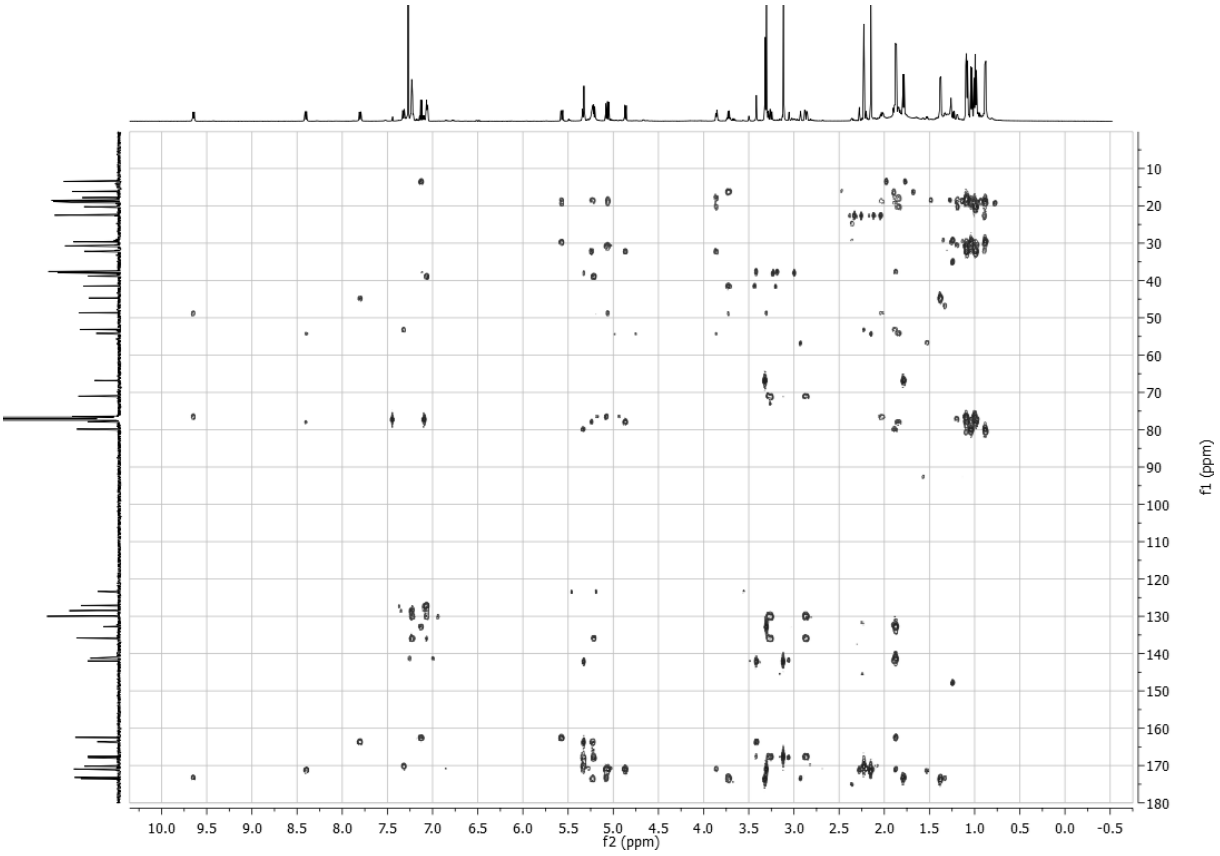


Figure 9.16: ^{1}H-^{13}C HMBC NMR spectrum of compound 8 in acetonitrile-d_3 (600 MHz).

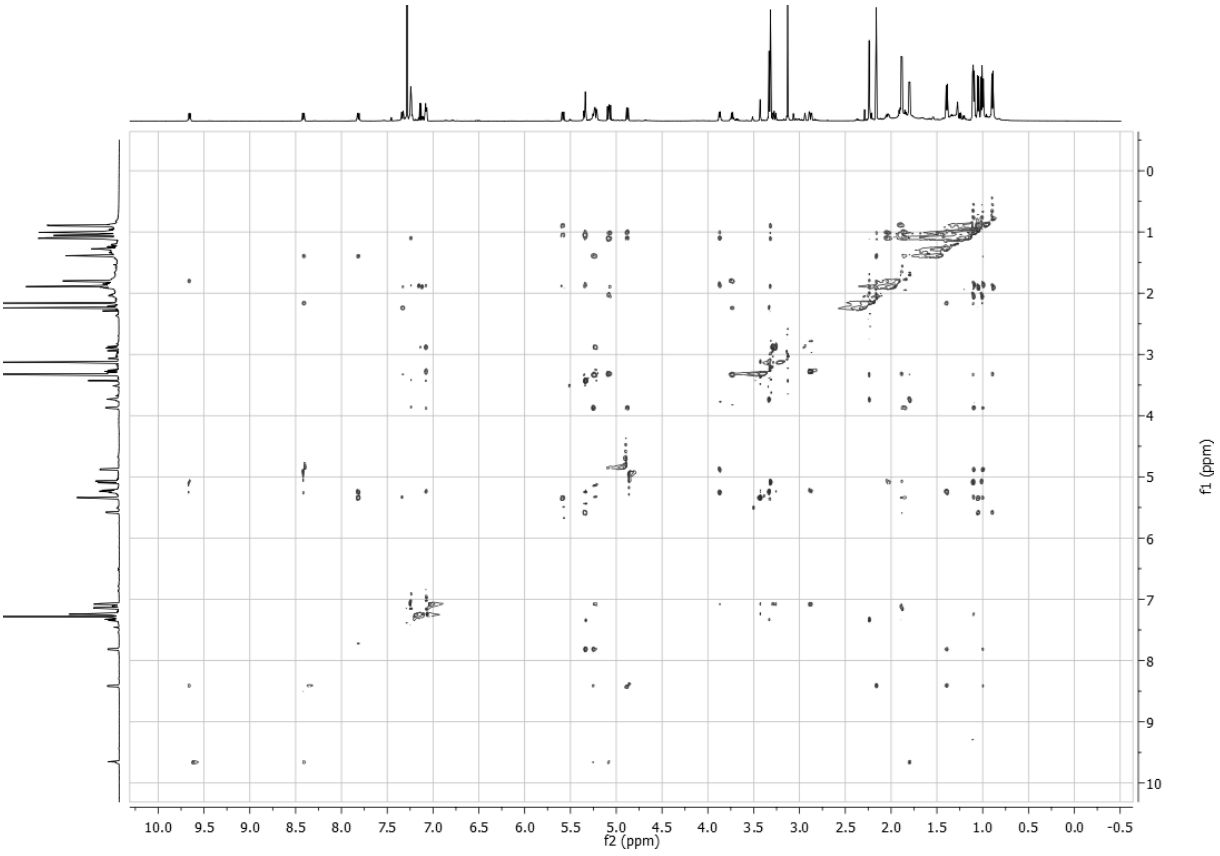


Figure 9.17: ^{1}H-^{1}H ROESY NMR spectrum of compound 8 in acetonitrile-*d*$_3$ (600 MHz).

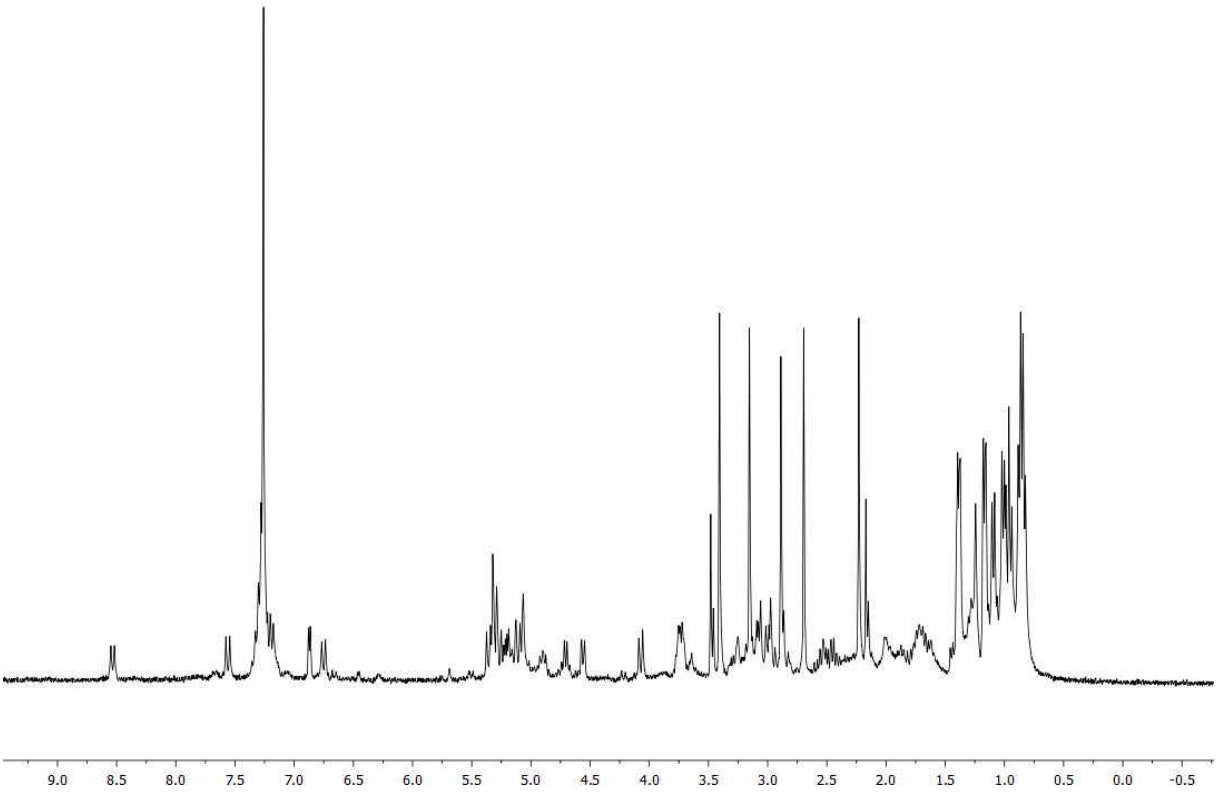


Figure 9.18: ^{1}H NMR spectrum of 19 in $CDCl_3$ (300 MHz).

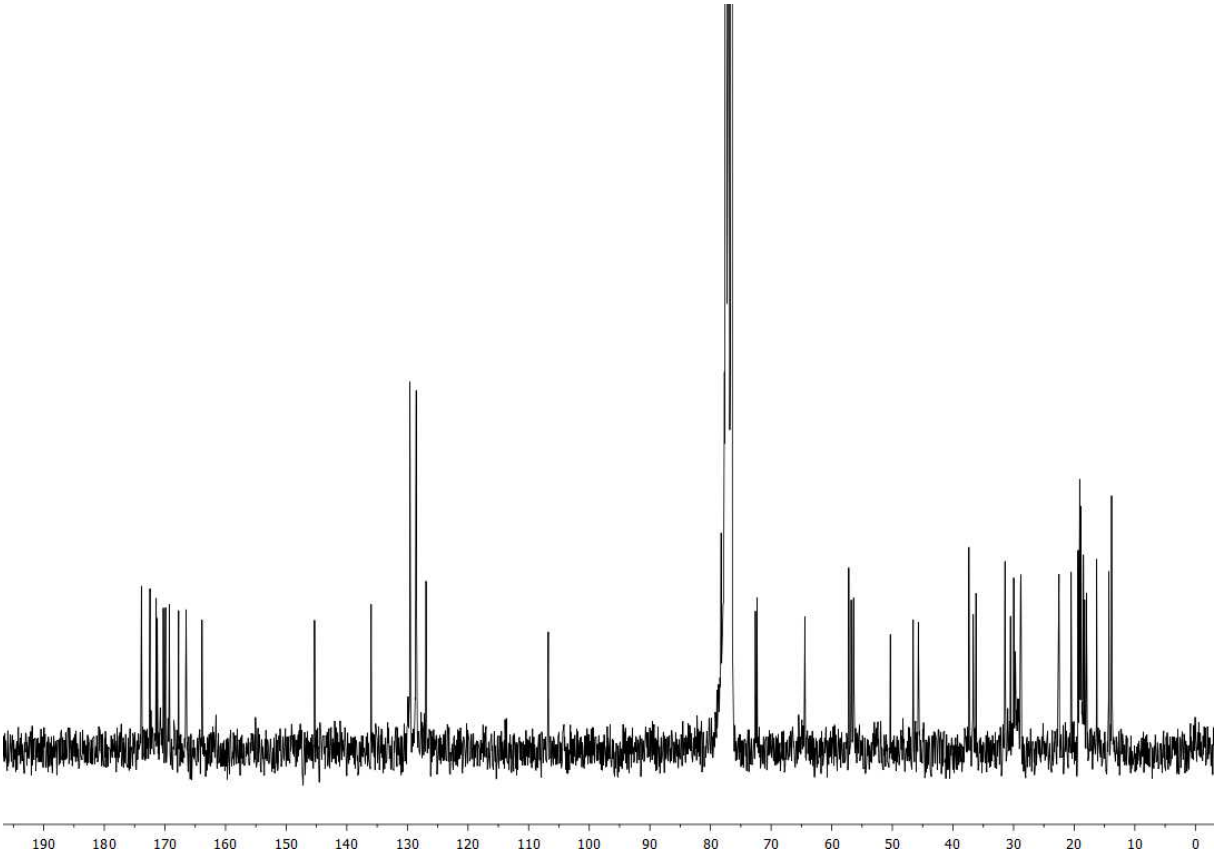


Figure 9.19: ^{13}C NMR spectrum of 19 in $CDCl_3$ (75 MHz).

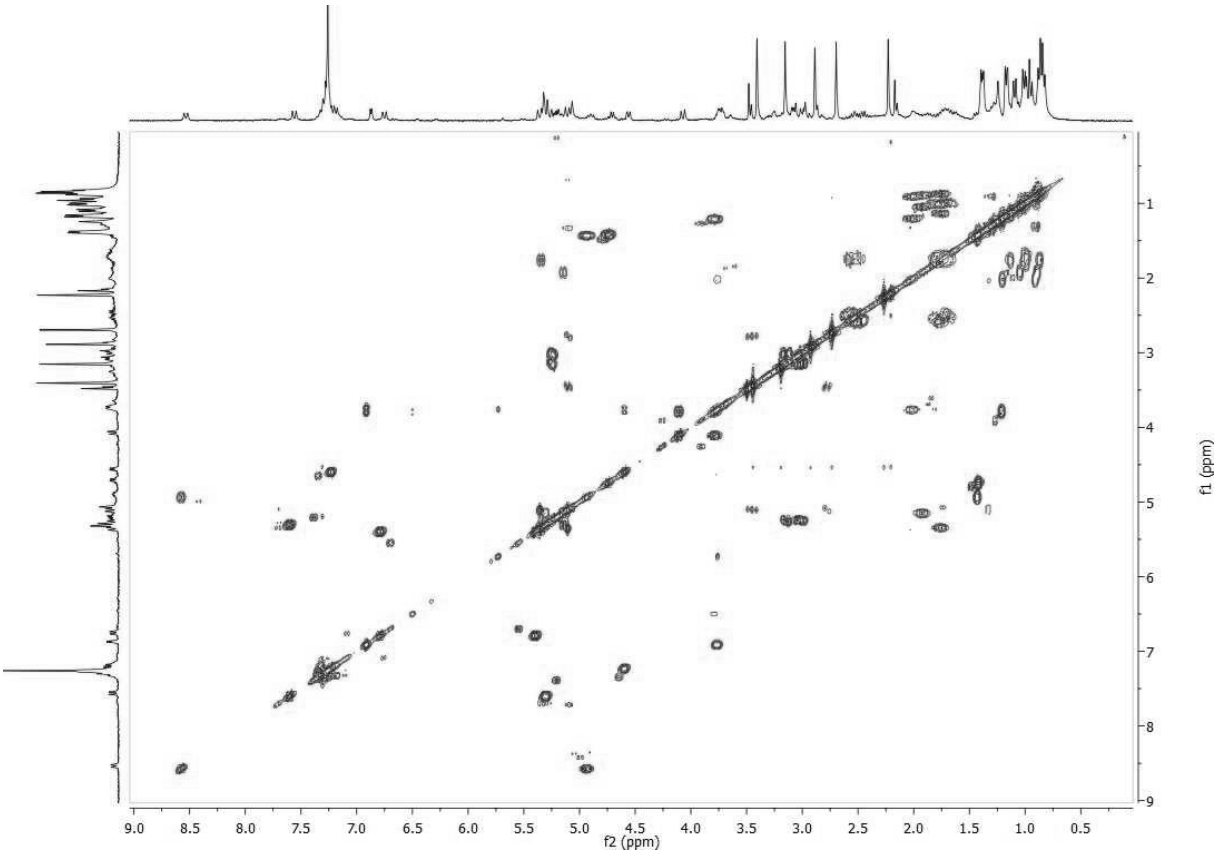


Figure 9.20: ^{1}H-^{1}H COSY NMR spectrum of 19 in $CDCl_3$ (300 MHz).

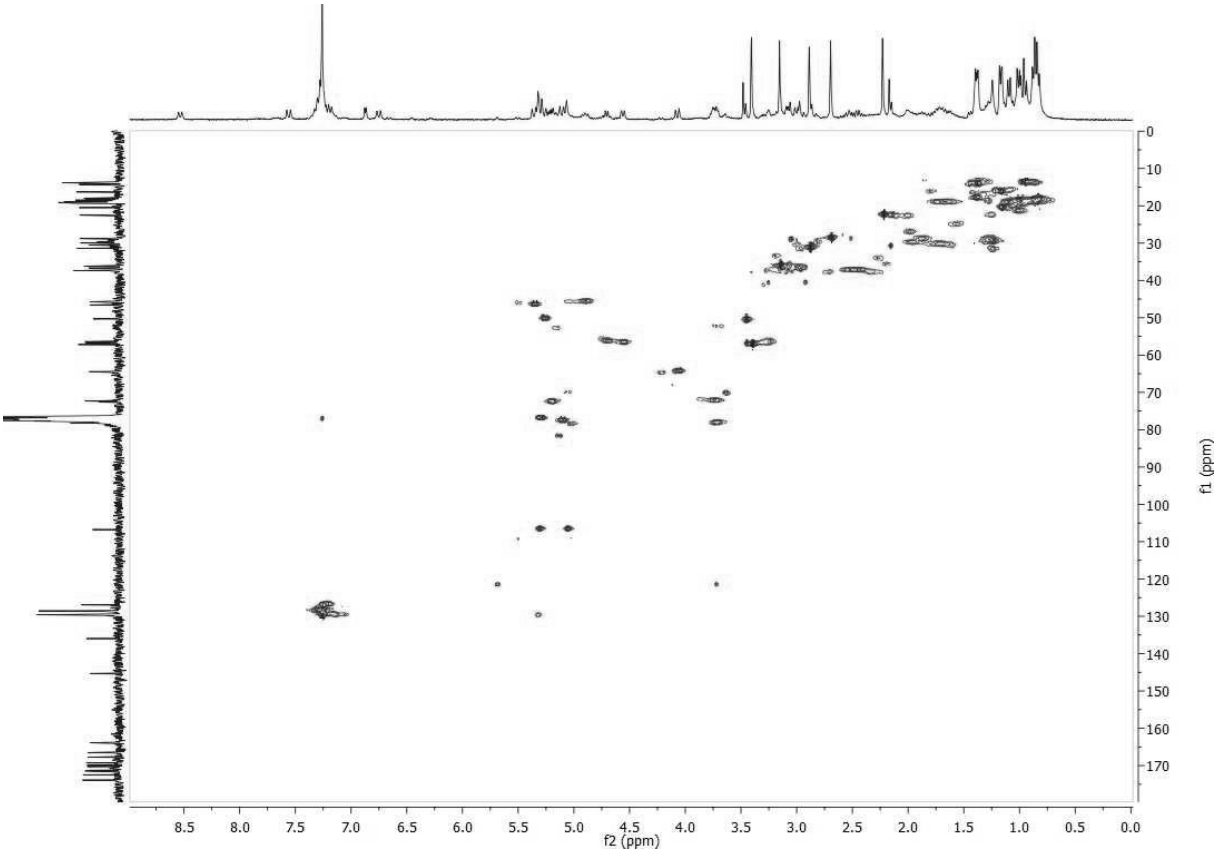


Figure 9.21: ^{1}H-^{13}C HSQC NMR spectrum of 19 in $CDCl_3$ (300 MHz).

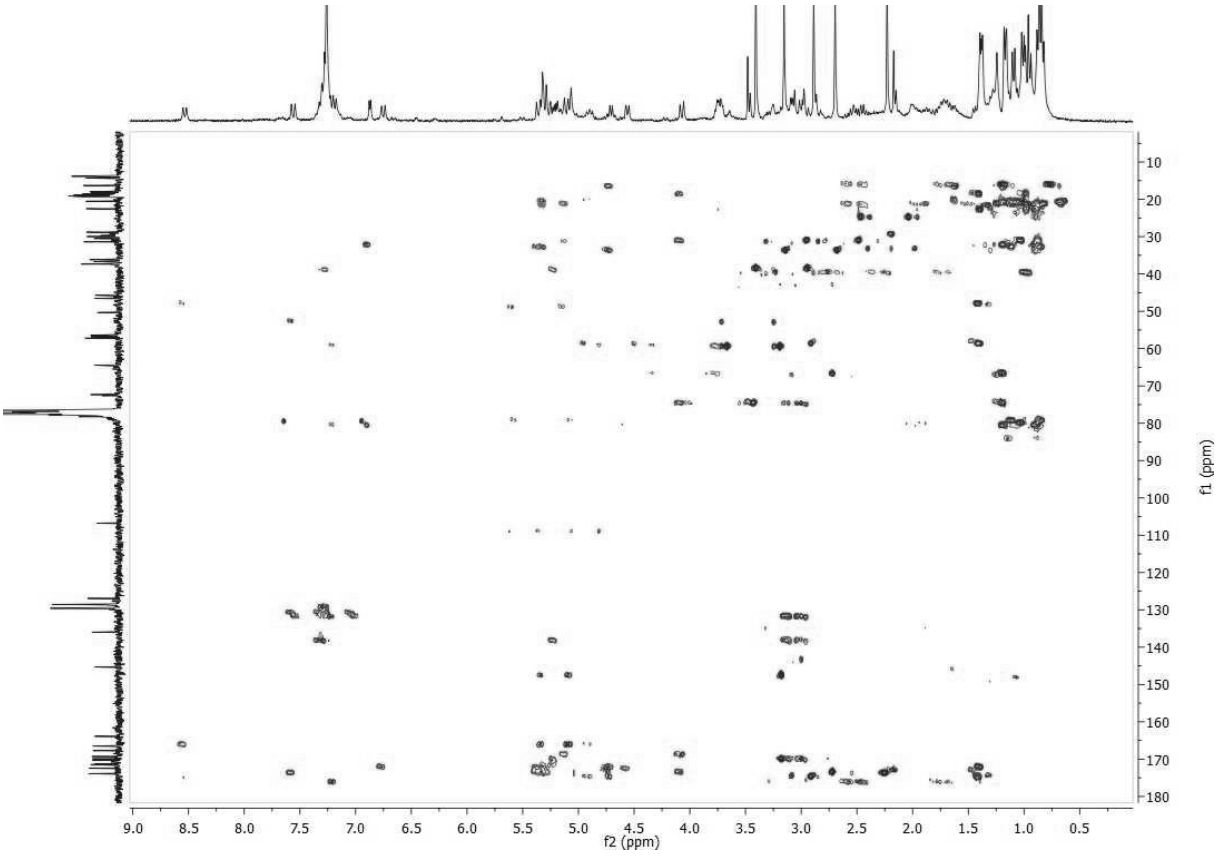


Figure 9.22: ^{1}H-^{13}C HMBC NMR spectrum of 19 in $CDCl_3$ (300 MHz).

9.4 Supplementary data

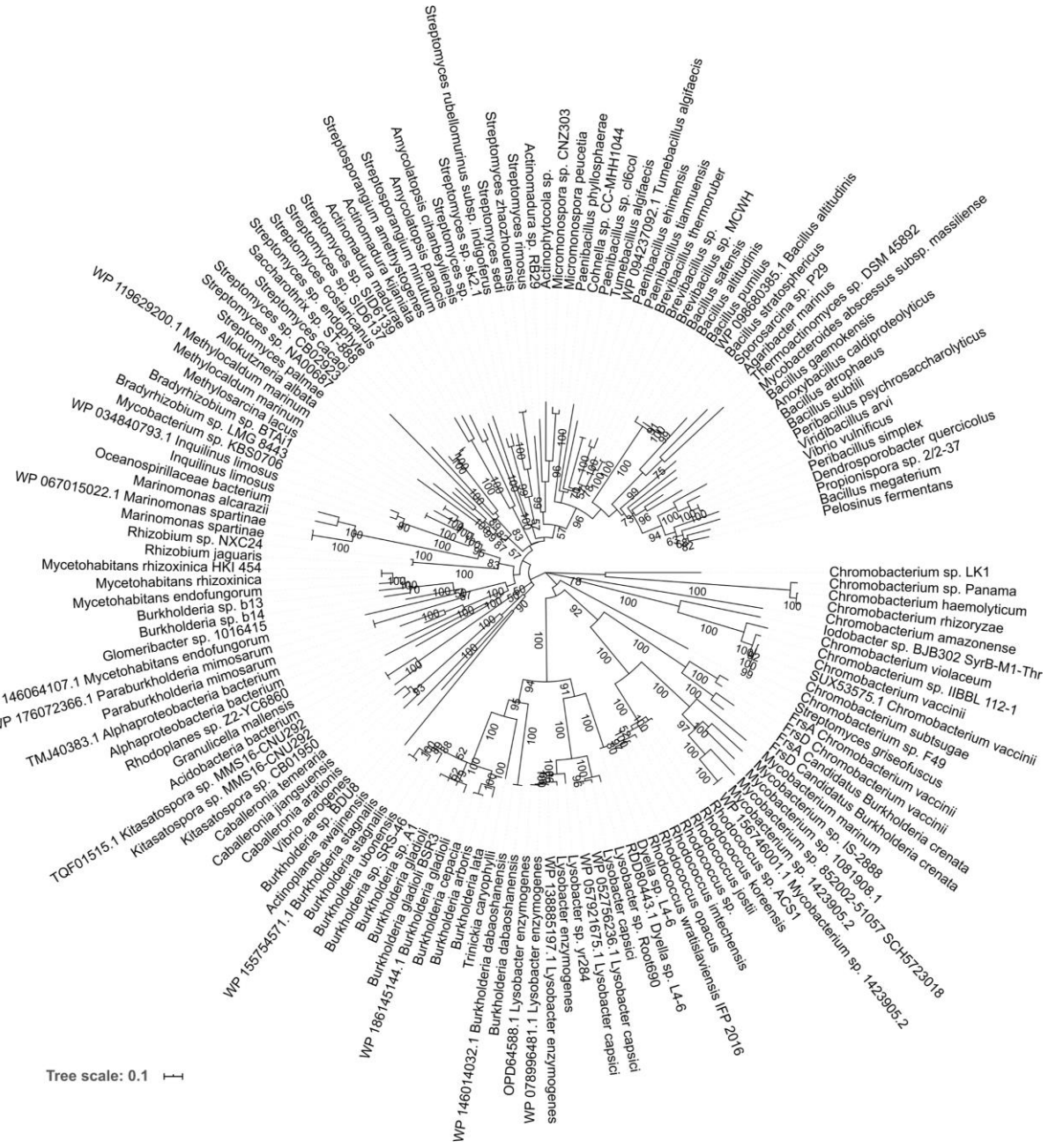


Figure 9.23: Phylogenetic tree of NRPS $C_{starter}$ domains. The scale bar represents 10 substitutions per 100 amino acids. For experimental details, see 6.13.3 and Table 6.23.

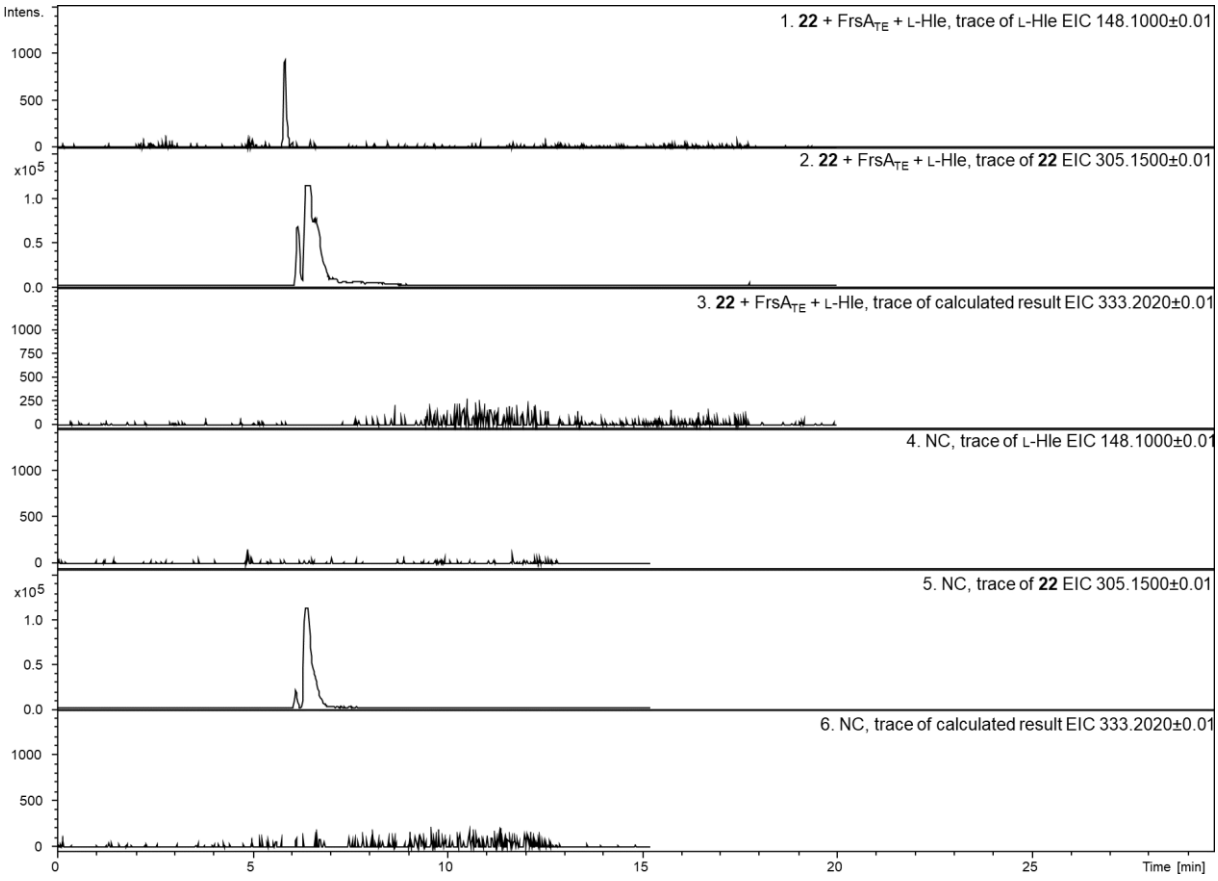


Figure 9.24: *In vitro* assay with L-Hle and FrsA$_{TE}$. LC-MS data of assay of **22** with FrsA$_{TE}$ and L-Hle. **1.** extracted ion traces of L-Hle *m/z* 148.1000; **2.** extracted ion traces of **22** *m/z* 305.1500; **3.** extracted ion traces of calculated result *m/z* 333.2020; **4.** negative control with heat-inactivated protein, extracted ion traces of L-Hle *m/z* 148.1000; **5.** negative control, extracted ion traces of **22** *m/z* 305.1500; **6.** negative control, extracted ion traces of calculated result *m/z* 333.2020.

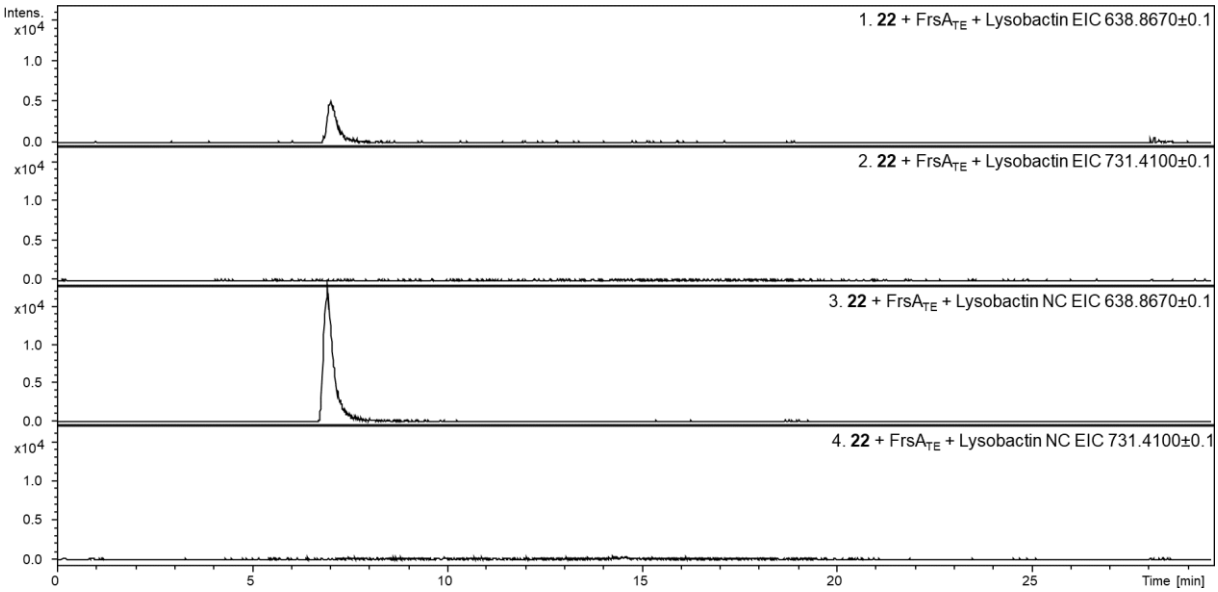


Figure 9.25: *In vitro* assay with Lysobactin and FrsA$_{TE}$. LC-MS data of assay of **22** with FrsA$_{TE}$ and Lysobactin. **1.** extracted ion traces of double charged Lysobactin *m/z* 638.8670; **2.** extracted ion traces of calculated result, double charged *m/z* 731.4100; **3.** negative control with heat-inactivated protein, extracted ion traces of double charged Lysobactin *m/z* 638.8670; **4.** negative control, extracted ion traces of calculated result, double charged *m/z* 731.4100.

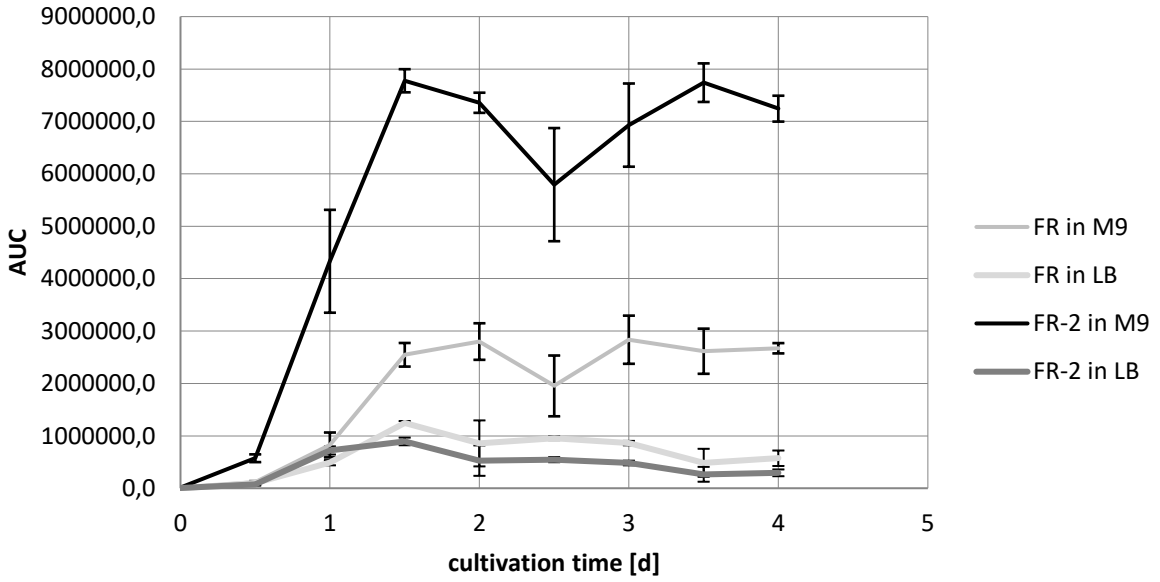


Figure 9.26: FR and FR-2 production of *C. vaccinii* in LB and M9 medium. The areas under the curves of the peaks in the extracted mass traces for FR (*m/z* 1002.54) and FR-2 (*m/z* 988.52) are plotted over the cultivation time in days. The cultivation for each reading point was performed in three biological repeats. The cultures were extracted with n-butanol and the crude extracts measured with the standard LC-MS method for FR. The work was done by Wiebke Hanke.

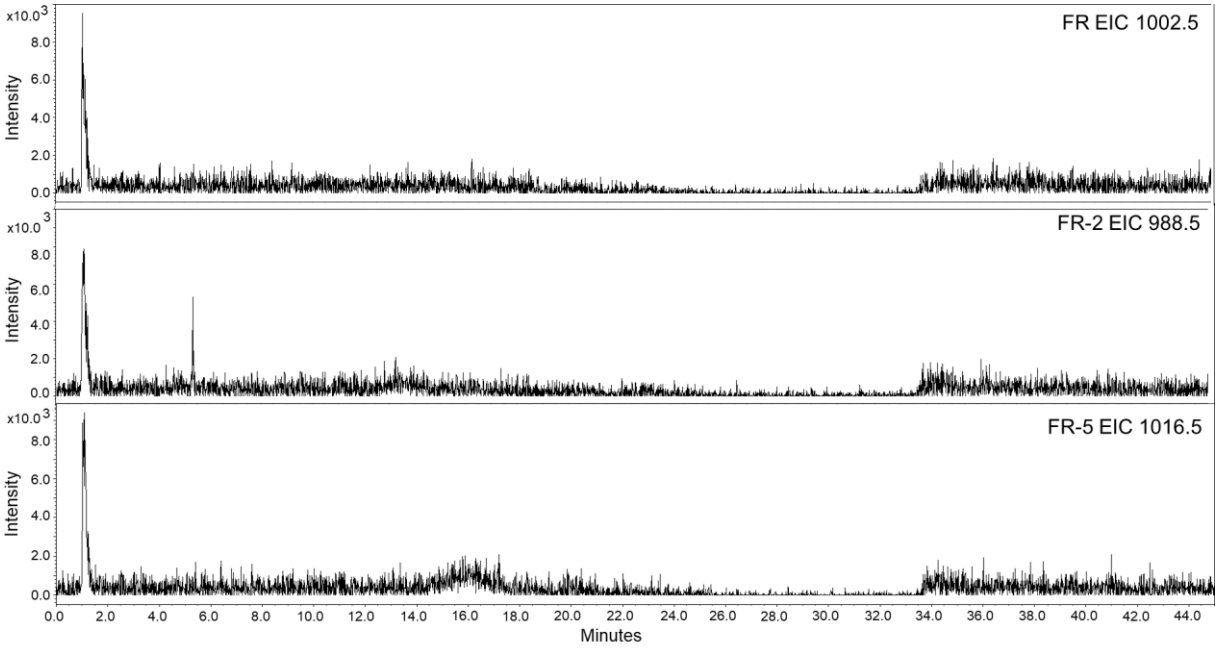


Figure 9.27: Extracted ion chromatograms of the blank M9 medium with 20 mM butyrate feeding solution. Top: EIC of FR *m/z* 1002.5; middle: EIC of FR-2 *m/z* 988.5; bottom: EIC of FR-butyryl *m/z* 1016.5.

10 In Advance Publications of the Dissertation

Research Papers

Hermes, C.; Richarz, R.; Patt, J.; Hanke, W.; Kehraus, S.; Voß, J.H.; Küppers, J.; Ohbayashi, T.; Namasivayam, V.; Alenfelder, J.; Inoue, A.; Mergaert, P.; Gütschow, M.; Müller, C.E.; Kostenis, E.; König, G.M.; Crüsemann, M. "Thioesterase-mediated Side Chain Transesterification Generates Potent Gq Signaling Inhibitor FR900359" *Nature Communications* 2020, accepted for publication.

Oral presentations

Invited talk at the International Association for General and Applied Microbiology (VAAM)-Workshop 2019, Jena, Germany. "Biosynthesis of the depsipeptide FR900359 – Investigation of the first NRPS module FrsA"

Poster presentations

Hermes, C.; Linke, J.; Crüsemann, M. and König, G. M. "Biosynthesis of the Depsipeptide FR900359 – Investigation of the first NRPS Module FrsA" 3rd European Conference on Natural Products 2018, Frankfurt

Hermes, C.; Schamari, I.; Reuter, T; Crüsemann, M. and König, G. M. "Biosynthesis and heterologous expression of the selective Gq protein inhibitor FR900359" International Symposium on GPCRs and G-Proteins 2017, Bonn.

Hermes, C.; Reuter, T.; Crüsemann, M. and König, G. M. "Biosynthesis of the selective Gq protein inhibitor FR900359" Workshop of the International Association for General and Applied Microbiology (VAAM) 2017, Tübingen

11 Acknowledgements

Ich möchte mich zunächst bei Frau Prof. König bedanken, dass sie mich in ihre Arbeitsgruppe aufgenommen und mir ein spannendes Projekt für meine Doktorarbeit gegeben hat. Sie war stets erreichbar, hatte kreative Vorschläge, wenn ich nicht weiterwusste und hat mir geholfen das Projekt in der richtigen Bahn zu halten. Auch hat sie mir den Freiraum gegeben meine eigenen Ideen zu verwirklich und selbst festzustellen was funktioniert und was man vielleicht lieber lassen sollte. Ihre Arbeitsgruppe hat mich freundlich aufgenommen und es war ein wunderbares Arbeitsumfeld, in dem man gerne ins Labor kam.

Genauso danken möchte ich Dr. Max Crüsemann, der für den Großteil der Betreuung meines Projektes verantwortlich war. Max war immer motivierend und unterstützend für mich da und hat mir auch im Laboraltag einiges beigebracht. Gerade wenn mal wieder nichts funktioniert hat oder Geräte Schwierigkeiten machten war er präsent und ansprechbar. Danke für deine Hilfe in den vier Jahren!

Prof. Evi Kostenis, Prof. Michael Gütschow und Prof. Heike Wägele danke ich für die Übernahme der weiteren Gutachten dieser Arbeit. Den beiden Ersteren danke ich auch für der Zusammenarbeit in der Forschergruppe FOR 2372 und die Kooperation mit ihren Arbeitsgruppen für verschiedene Experimente.

Besonders möchte ich mich bei allen Mittarbeitern der AG König/Crüsemann bedanken, die mich all die Zeit begleitet haben! Ihr wart wundervoll und wir hatten viel Spaß bei Weihnachtfeiern, Ausflügen und Mittagspausen. Extra erwähnen möchte ich dabei Daniel Wirtz, Wiebke Hanke, Sophie Klöppel, Cora Herzer und Ben Libor für all die Spieleabende und fröhlichen Momente im Labor. Daniel und Wiebke auch ein besonderer Dank fürs Lesen dieser Arbeit in unfertiger Form und das Finden diverser doppelter Leerzeichen ;-). Stefan Kehraus danke ich für die Hilfe mit den NMR Daten und Katharina und Mila für vielfältige Unterstützung im Labor.

Als nächstes möchte ich mich bei allen Kollegen anderer Arbeitsgruppen für die Zusammenarbeit bedanken:

Jim Küppers für die vielen Stunden der Synthese und dem Beistand beim Beobachten tropfender Säulen;

Julian Patt und Judith Alenfelder für die DMR-Messung meiner isolierten Substanzen;

Jan Hendrik Voß für die Gαq binding studies und das wunderschöne Model des FR-Gαq Komplexes;

Tobias Claff für die Einführung in den thermal shift assasy und dessen Datenauswertung;

Und ganz besonders bei Jonas Mühle, für seine Kristallisierungserfolge und unzählige lustige Skype Unterredungen!

Die Zusammenarbeit mit euch war mir eine Freude!

Ganz besonders möchte ich mich auch bei meiner Familie bedanken, die mich immer unterstützt und begleitet hat. Mama, Papa, ihr seid einfach die Besten und ich kann gar nicht alles aufzählen, wofür ich euch dankbar bin, daher versuche ich es erst gar nicht, euere Liebe und Wertschätzung ist unbezahlbar!

Micha, mein liebster großer Bruder, du hat mir in so manchem den Weg bereitet, bist mit gutem Beispiel voran gegangen und hattest immer ein offenes Ohr für mich. Es gilt weiterhin: For Pony!!!

Tina, Schwesterherz, danke das du immer für mich da bist, wenn ich dich brauche, für Aufheiterung, Rat und gemeinsames „auf andere Gedanken kommen“ live oder in letzter Zeit so häufig digital. Du bist die Beste!

Der Wichtigste kommt zum Schluss: Julian, mein Liebster, Komplize, Kampfgefährte, Lichtblick und unverschämter Troll. Ich weiß nicht was ich die letzten Jahre ohne dich gemacht hätte, aber mein Leben wäre definitiv deutlich langweiliger gewesen. Danke, dass du für mich da bist!